

- and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 18, 319-332.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts*. New York: Academic Press.
- Harris, G., Begg, I., & Upfold, D. (1980). On the role of the speaker's expectations in interpersonal communication. *Journal of Verbal Learning and Verbal Behavior*, 19, 597-607.
- Hildyard, A. (1979). Children's production of inferences from oral texts. *Discourse Processes*, 2, 33-56.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.
- Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, 14, 78-106.
- Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W. J. M. Levelt (Eds.), *Crossing the boundaries in linguistics: Studies presented to Manfred Bienisch*. Dordrecht: Reidel.
- Litman, D. J. (1985). *Plan recognition and discourse analysis: An integrated approach for understanding dialogues*. Ph.D. dissertation, University of Rochester, Computer Science Department.
- Markman, E. M. (1981). Comprehension monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills*. New York: Academic Press.
- Marslen-Wilson, W., Levy, E., & Tyler, L. K. (1982). Producing interpretable discourse: The establishment and maintenance of reference. In R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action*. London: Wiley.
- McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrumental inference. *Journal of Verbal Learning and Verbal Behavior*, 20, 671-682.
- Robinson, E. J. (1981). The child's understanding of inadequate messages and communication failure: A problem of ignorance or egocentrism? In W. P. Dickson (Ed.), *Children's oral communication skills*. New York: Academic Press.
- Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language: Explorations in comprehension beyond the sentence*. Chichester: Wiley.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.
- Snow, C. (1972). Mothers' speech to children learning language. *Child Development*, 43, 549-565.
- SPSS (1983). *SPSSX user's guide*. Chicago: McGraw-Hill.
- van Dijk, T. (1977). Semantic macro-structures and knowledge frames in discourse comprehension. In P. Carpenter & M. Just (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Erlbaum.
- Walker, C., & Yekovich, F. (1984). Script-based inferences: Effects of text and knowledge variables on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 23, 357-370.
- Whitehurst, G. J., & Sonnenschein, S. (1981). The development of informative messages in referential communication: Knowing when versus knowing how. In W. P. Dickson (Ed.), *Children's oral communication skills*. New York: Academic Press.
- Winograd, T. (1977). A framework for understanding discourse. In P. Carpenter & M. Just (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Erlbaum.

(Accepted December 15, 1986)

Identification and Ratings of Caricatures: Implications for Mental Representations of Faces

GILLIAN RHODES

Stanford University

SUSAN BRENNAN

Hewlett-Packard

AND

SUSAN CAREY

Massachusetts Institute of Technology

S. Brennan (1985, *Leonardo*, 18, 170-178) has developed a computer-implemented caricature generator based on a holistic theory of caricature. A face is represented by 37 lines, based on a fixed set of 169 points. Caricatures are produced by exaggerating all metric differences between a face and a norm. Anticaricatures can be created by reducing all the differences between a face and a norm. Caricatures of familiar faces were identified more quickly than anticaricatures. There was no difference in identification accuracy for the three types of representation. The best likeness was considered to be a caricature. We discuss the implications of these results for how faces are mentally represented. The results are consistent with a holistic theory of encoding in which distinctive aspects of a face are represented by comparison with a norm. We suggest that this theory may be appropriate for classes of visual stimuli, other than faces, whose members share a configuration definable by a fixed set of points. © 1987 Academic Press, Inc.

When a face is represented in memory, its distinctive aspects are likely to be encoded. This idea is referred to as the distinctiveness hypothesis. Its distinctive aspects noted, a face can be differentiated from other faces, and can be identified or recognized when seen again.¹ There is

This research was supported by NSF Grant BNS 80-05517 to Roger Shepard. It was reported as part of a doctoral dissertation by the first author. We thank Herb Clark, Phoebe Ellsworth, Ann O'Leary, Roger Shepard, and Barbara Tversky for helpful discussions about this work, and Ian McLean and Gilbert Mesa for photographic assistance. Correspondence, including reprint requests, should be addressed to Gillian Rhodes, Department of Psychology, Canterbury University, Private Bag, Christchurch, New Zealand. Inquiries about the caricature generator should be addressed to Susan Brennan, Department of Psychology, Stanford University, Stanford, CA 94305.

¹ A face could in principle be encoded in terms of its typical rather than its distinctive aspects. This is possible because individual faces differ in which aspects are typical.

evidence that subjects use distinctive information when encoding unfamiliar faces. Winograd (1981) has shown that the presence of a distinctive feature, and instructions drawing attention to that feature, facilitate memory for photographs of faces. Similarly, atypical or distinctive faces are remembered better than typical faces (Bartlett, Hurry, & Thorley, 1984; Cohen & Carr, 1975; Going & Read, 1974; Light, Kayra-Stuart, & Hollander, 1979).² No doubt this recognition advantage for atypical faces occurs because such faces are less confusable with other faces than are typical faces. This interpretation is supported by the results of category formation studies, which show that typical category members, never previously seen, are often falsely reported as having been seen before (Franks & Bransford, 1971; Solso & McCarthy, 1981).

The recognition advantage for distinctive faces provides some evidence for the distinctiveness hypothesis. However, in these studies distinctiveness is varied across faces, thereby leaving open the possibility that some other aspects of faces (e.g., attractiveness) might covary with distinctiveness and mediate recognition performance. A better test of the distinctiveness hypothesis would be to vary distinctiveness within faces, which could be done using caricatures.

Caricatures are representations of faces in which distinctive aspects are exaggerated. The caricaturist Low described caricatures as an art of "calculated emphasis," an "evolutionary process of simplification, a kind of survival-of-the-most-significant." The caricaturist wants to know "in what particulars a man is like himself and no-one else" (cited by Gibson, 1969, p. 102). According to the distinctiveness hypothesis, caricatures should be good representations of faces, and should be better than representations in which the distinctive aspects are minimized.

A number of authors have taken the distinctiveness hypothesis a step further, and suggested that memory representations might be schematized so as to emphasize the distinctive properties of what is being represented (Bartlett, 1932; Brennan, 1982; Tversky & Baratz, 1985). The idea that distinctive properties are exaggerated in memory, which we will refer to as the caricature hypothesis, seems especially attractive in the case of faces, because caricatures are such effective facial representations. Sev-

² The idea that distinctive aspects of faces are encoded does not imply that faces are represented in terms of discrete features, as has been assumed by some authors (e.g., Winograd, 1981). Both continuous dimensions and discrete features can take on distinctive values. A continuously varying dimension such as the ratio of eye separation to face height is distinctive to the extent that it takes on an extreme value in the population distribution of values for that dimension. A feature that can take on one of a given set of discrete values (including presence or absence), is distinctive to the extent that it takes on a value that occurs with low frequency in the population.

eral authors have tried to test the prediction, which follows directly from the caricature hypothesis, that caricatures are better representations of individuals than are more veridical representations (Hagen & Perkins, 1983; Tversky & Baratz, 1985).

Both Tversky and Baratz (1985), and Hagen and Perkins (1983) compared old/new recognition performance for caricatures and photographs and failed to find any recognition advantage for caricatures. Tversky and Baratz (1985) also failed to find a caricature advantage in a face/name matching task. However, these results may simply mean that photographs contain so much more spatial information than hand-drawn caricatures that they are more recognizable despite being less distinctive. A better test would be to vary distinctiveness within a single representational medium.

Another problem with one of these studies (Hagen & Perkins, 1983) is the assumption that 3–5 s viewing time for completely unfamiliar faces is sufficient to allow the kind of encoding that would produce a schematized or caricatured memory representation. A more realistic approach would be to use well-known faces, as done by Tversky and Baratz (1985).

Brennan (1982, 1985) has developed a computer-implemented caricature generator that allows the distinctive aspects of line-drawing representations of individual faces to be varied continuously. In the present studies we use the caricature generator to test the distinctiveness and caricature hypotheses, while avoiding the problems of comparability of medium and the use of unfamiliar faces. The caricature generator produces a caricature in three steps. First, a line-drawing representation is created based on a digitized photograph of a person. The lines were chosen with the goal of representing a recognizable face with as few lines as possible. The line-drawing representation is essentially a "traced" version of the digitized image, in which high spatial frequency information (i.e., rapid changes in lightness) is preserved. Second, the line drawing is compared with a norm (average) face by scaling the two faces and aligning the pupils. The metric differences between corresponding points in each face are represented by vectors. Third, all metric differences between the target and norm faces are exaggerated by an amount specified by the user, to produce a caricature. Thus distinctive aspects of a face are made more distinctive, whereas those aspects that are close to the norm (average face) are changed by only a small amount.

Anticaricatures can also be created, in which distinctive aspects are de-emphasized, thereby bringing the faces closer to the average or norm face.

The caricature generator has a number of advantages as a research tool. It allows continuous parametric variation of the metric properties in

faces, a facility that far exceeds the control possible in other currently available methods of facial stimulus construction. The implementation includes a facility for averaging line-drawing representations of faces. The technique sidesteps the problem of knowing explicitly in what ways faces can be distinctive, and it has great flexibility in that the same variations are not forced onto all faces—whatever is distinctive about a face is varied.

Two experiments were carried out to test the distinctiveness and caricature hypotheses. In Experiment 1 we compared identification performance and likeness judgments for caricatures, anticaricatures, and veridical line drawings of familiar faces produced by the caricature generator. For a given degree of exaggeration, the caricature and the anticaricature differ equally in metric terms from the veridical face, but the caricature is less like other faces (i.e., it is more individuated) than the anticaricature, which in turn is more like the average face, and therefore more like other faces. If distinctive properties of a face are especially likely to be encoded and used in identification or recognition (the distinctiveness hypothesis), then the caricatures should be more easily identified and be considered better likenesses than the anticaricatures. The more interesting comparison, however, is between the caricatures and the veridical line drawings. Is there some level of exaggeration that is actually more easily identified or judged to be a better likeness of a person than the veridical drawing? If so, then this would be consistent with the caricature hypothesis, which states that distinctive aspects of faces are exaggerated in memory.

Previous studies (Hagen & Perkins, 1983; Tversky & Baratz, 1985) have used old/new recognition tasks (although Tversky & Baratz also included a name/face matching task). We used an identification task, which requires that semantic information about the face (in particular its name) is accessed in addition to a simple familiarity decision. Identification therefore requires additional processing to a familiarity judgment (required in an old/new recognition task). An identification task was used because it is more like what we normally do when we see a face than is an old/new recognition task. We would not, however, make different predictions for the two types of recognition task.

Typicality ratings were obtained for a large set of faces. These data were used to get norm faces for the caricature generator. The three most typical female and the three most typical male faces were averaged to produce female and male norm faces, respectively.

Experiment 2 was designed to rule out explanations based on biases that do not depend on familiarity with the faces used and to check that the veridical drawings were in fact veridical.

EXPERIMENT 1 Method

Subjects

Twenty graduate students from the Stanford University Psychology Department, who were familiar with the faces, participated in the identification and likeness rating tasks. Forty Stanford undergraduates, who were unfamiliar with the faces used, made the typicality judgments.

Stimuli

Typicality ratings. The stimuli were 65 3.5 × 5 in. black and white full-face photographs of faces. There were 37 male and 28 female faces. All posers displayed a neutral expression and all were Caucasians (except for one Chicano) between 17 and 32 years of age. The posers were volunteers from the Stanford University Psychology and Computer Science Departments, and from a Stanford University undergraduate dorm.

Caricature identification and likeness judgments. The stimuli were veridical line drawings, caricatures and anticaricatures, produced by the caricature generator, based on photographs of members of the Stanford University Psychology Department (a subset of the photographs used in the typicality judgments task). There were 8 male and 10 female faces making a total of 18. All posers were between 22 and 32 years of age. The men were clean-shaven.

The photographs were digitized, and line drawings were produced from these interactively with a Macintosh computer, by an experienced user of the system (SB). Each drawing consisted of 37 lines defined by 169 points (see Fig. 1). The points were found by eye, although this process could in principle be carried out automatically using top-down point-finding algorithms of the sort developed by Kelly (1970) and Kanade (1973).³ The points on each line are either points of inflexion, when present, or else are evenly spaced along the trajectory of the line, and are found in a fixed top-to-bottom, left-to-right order. Some points of inflexion correspond to precise locations on the face (edge of eye, tip of nose, etc.). The points in each line were joined automatically by an interpolating spline (smooth) curve.

Each face in the database was congruently represented in that it contained its own version of every line of the canonical face. If a particular line was partially occluded in the digitized photograph, which was sometimes the case with ears or hair, an estimate of its potential location was made. If a line did not exist at all, a location for it was estimated and it was stored as a virtual (or invisible) line. This strategy enabled us to compare faces that did not have all the same lines visible.

To produce each caricature the line drawing was compared with a norm face of the same sex, and all differences between the two were exaggerated by a specified percentage or proportion. Different norms were used for male and female faces, structurally different subsets of faces, which we assumed would have different norms. Two different norm faces were used for each sex. One was the average of (the veridical line drawings of) the three most typical faces (av3), as determined by subjects' ratings. The other was the average of all

³ A selection process was necessary to define the canonical face and to fit this representation to the veridical faces. We would, of course, prefer to automate this process. While we are able to filter photographs to get rough line drawings (which share many of the lines that we represent schematically), there is still the need to map individual points onto these lines. We have been unable to automate the entire point-finding process, even using goal-directed picture processing techniques.

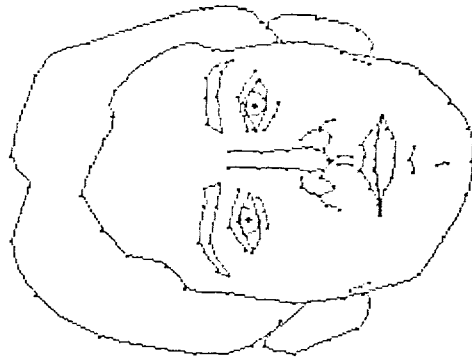


FIG. 1. The points and lines used to depict a face. The face is represented by 169 points (darker dots), joined automatically to form 37 lines. Any of the lines can be made invisible. Invisible lines are stored as virtual lines in the representation, and are used (but not displayed) if the face is compared with one that has these lines visible.

the faces of the same sex used in this experiment (avpsych).⁴ It should be noted that the computational average of any heterogeneous set of faces should produce a reasonably good "average" face. The four norm faces used are shown in Fig. 2.

The norm face and the stimulus face were scaled and translated so that their pupils were aligned. The caricature generator compares the corresponding points in each face and represents the distance and direction between them as a vector. Caricatures are created by adding a proportion of the difference between the two vectors to the points for the line drawing of the stimulus face. An example of the caricatures produced for a famous face (John F. Kennedy) is given in Figure 3. Anticaricatures are produced by reducing these distances by the same amount, i.e., by interpolating between the veridical drawing and the norm face. The drawings were exaggerated and de-exaggerated by the following proportions: .25, .50, and .75. The images were scaled by multiplying the width by 1.06 before printing, in order to correct for the constant horizontal distortion introduced by the printer.

These six images, together with the veridical line drawing, constituted the pictures of each person used as stimuli. An example of such a set is shown in Fig. 4. A photograph of the person depicted is shown in Fig. 5. It should be noted that for each degree of exaggeration, the caricature and anticaricature differ equally in their metric properties from the veridical line drawing. For each face there were two groups of seven pictures, one group for each norm face used. These groups were randomly assigned to two stimulus sets, A and B,

⁴ In our model, a norm provides a comparative basis for the mental encoding of faces. It is possible that the basis of comparison could be composed of a metric average of all the faces ever seen, or of a single typical face, or of an average of typical faces, or even some cultural norm (such as a classical ideal of beauty). The two norms we used differ in the following way: av3 is the average of the three (same sex) faces judged most typical, and avpsych is the average of all the (same sex) faces in the stimulus set. We used both sorts of norm in order to explore the possibility that one sort would more closely resemble our internal norms than the other.

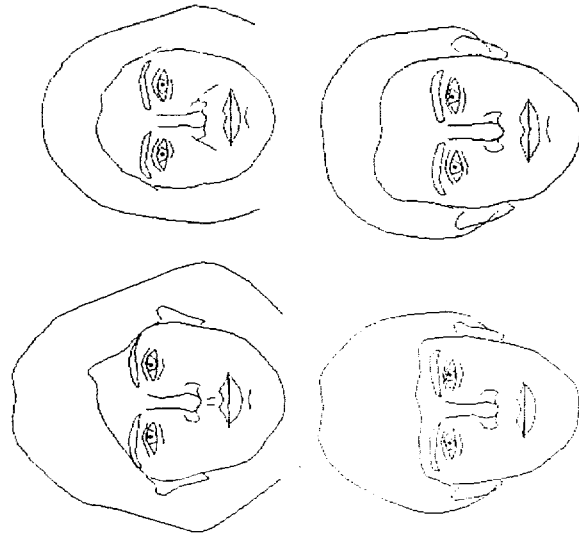


FIG. 2. The female (top) and male (bottom) norm faces used. The two norms on the left were created by averaging the three most typical faces of the appropriate sex (av3 norms). The two norms on the right were created by averaging all the psychology faces of the appropriate sex (10 female and 8 male faces).

so that all faces appeared in each set, but half the drawings in each set were based on av3 norms and half on the avpsych norms.

Procedure

Typicality ratings. Subjects who were unfamiliar with the faces were asked to rate the typicality of each face on a 7-point rating scale. Examples were given of what was meant by typicality, e.g., a robin is a typical bird, whereas a penguin is an atypical or unusual bird. Subjects were instructed to judge typicality in relation to all (Caucasian) faces (of the same sex), not just in relation to the ones shown. They were told that for each face they should think about whether it looked like most other faces, or whether it was unusual or atypical.

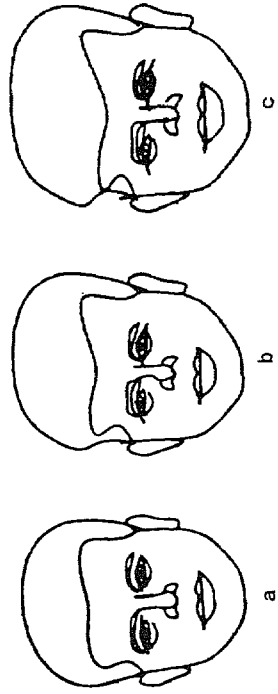


FIG. 3. An example of caricatures produced by the caricature generator: John F. Kennedy: (a) = 0 exaggeration, (b) = .50 (50%) exaggeration, (c) = 1.00 (100%) exaggeration.

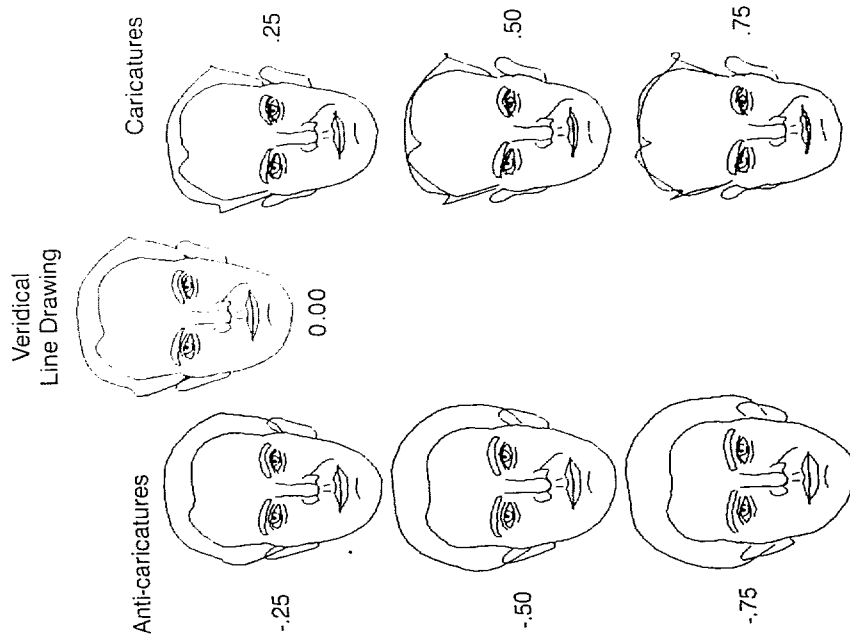


FIG. 4. An example of the stimulus set for a face.

They were also told to base their judgments on permanent aspects of the faces, and not on things like hairstyle that could be easily changed.

Male and female faces were presented in separate sets. Half the subjects rated female faces first and half rated male faces first. The faces in each set were presented in one random order to half the subjects and in the reverse order to the other half of the subjects. This was counterbalanced with the order in which the male and female faces were rated.

After the faces in each set were rated, subjects selected the three most typical and the three least typical faces in the set, which they then ranked for typicality. They were also asked to indicate whether they had seen any of the faces before. Subjects who knew more than seven of the faces were replaced, because it seemed likely that familiar faces would be judged differently than would unfamiliar faces. For example, personality and other information about the person might affect the typicality judgments for familiar faces.

After both sets had been rated, subjects were asked to rate the difficulty of making the typicality judgments on a 7-point scale (1 indicated no difficulty and 7 indicated extreme difficulty).

Identification. Subjects who knew the people depicted were told that they would be shown some pictures of people in the psychology department (mostly of graduate students, with some of staff and faculty), and that they should try to identify them. For a given face,



FIG. 5. A photograph of the person depicted in Fig. 4.

one-third of the subjects saw the veridical line drawing (0 exaggeration), one-third saw the .50 caricature, and the remaining third saw the -.50 anticaricature. Each subject saw six faces at each level of exaggeration. Thus degree of exaggeration was both a within-subject and a within-face factor. This factor was crossed with the type of norm used (av3 or avpsych).

Each picture was placed beneath a sheet of cardboard. At a signal from the experimenter, the subject lifted the cardboard to view the picture. More than one attempt at identification was allowed, but there was a maximum viewing time of 60 s. The dependent variables were accuracy and identification time, which was measured using a stop watch.⁵

Likeness judgments. The likeness judgments were made after the recognition test. The seven drawings of each face were placed in random order in a row in front of the subject. Subjects were instructed to arrange the pictures in order of how good a likeness of the person each was. "Goodness of likeness" was explained as meaning recognizability. Subjects also rated how good a likeness each picture was on a 7-point scale, where 1 was *poor* and 7 was *great*. Half the subjects saw set A faces and half saw set B faces. The order in which each set of faces was seen was reversed for half the subjects.

Subjects also rated how familiar they were with each face on a 7-point scale, so that unknown faces could be excluded from the analysis. It also seemed likely that a caricature advantage would be stronger for highly familiar than for less familiar faces in the likeness judgments (there were too few faces at each familiarity level to include familiarity as a factor in the identification analyses). After the likeness judgments had been made for the drawings, subjects made likeness judgments for the photographs on which the drawings were based. They were instructed to rate how good a likeness of the person each photograph was, as far as photographs go. These judgments were obtained so that the efficacy of the computer-generated images could be assessed in relation to the quality of the initial photographs. At the end of the session the subjects were asked whether they knew how the pictures had been produced, and what the hypothesis of the study was.

⁵ This measurement technique is not sensitive to small time differences and probably does not accurately measure times less than 1 s. This level of accuracy was, however, appropriate for the identification latencies in the present experiment.

Results and Discussion

Typicality Ratings

Four subjects were replaced because they recognized more than seven of the faces. Kendall's coefficient of concordance, W , which assesses the degree of consistency among a set of judges, was used to assess intersubject reliability.⁶ There was significant intersubject agreement on typicality ratings. For male faces, $W = .26$, $\chi^2(36) = 376.99$, $p \leq .001$. For female faces, $W = .35$, $\chi^2(27) = 378.49$, $p \leq .001$. Selection of the three most typical and the three least typical male and female faces was consistent with the rating data.

Typicality ratings for 36 of the 37 male faces in this study were also obtained (from different subjects) in Experiment 2. Therefore a measure of intersubject reliability, the correlation coefficient for the two sets of ratings, could be examined. This correlation was moderate and significant, $r(33) = .54$, $p \leq .01$, indicating that the typicality ratings could be made reliably. Subjects' ratings indicated that the task was difficult, $M = 4.7$, $SD = 1.3$.

Identification

Accuracy. A two-way repeated measures analysis of variance was carried out on the mean proportion of correctly identified, known faces, with norm (av3, avpsych) and exaggeration level ($-.50$, 0 , $.50$) as within-subject factors. The data were first scaled using the transformation, $y = x^{0.61}$, to remove a linear relationship between $\log(M)$ and $\log(SD)$.⁷ In this and all subsequent analyses missing data were replaced by the mean for the appropriate cell of the design. There was no main effect of norm, $F(1,19) < 1$, or exaggeration, $F(2,38) = 2.06$, ns., nor any interaction between norm and exaggeration level, $F(2,38) = 2.15$, ns. Because the mean in each cell of the design was based on only three faces, and because there was no effect of norm, or any interaction of this factor, the data were collapsed across the two norm conditions, and scaled using the transformation $y = x^{-.43}$, to remove a linear relationship between $\log(M)$ and $\log(SD)$. These data were analyzed using a one-way repeated measures analysis of variance. The mean proportion (untransformed) of faces correctly identified was .38 for the veridical drawings, .33 for the .50 cari-

⁶ Kendall's coefficient of concordance, W , tests the null hypothesis that there is no agreement among subjects in the rank order of the rated items. W is distributed as χ^2 with $N - 1$ degrees of freedom, where N is the number of items to be judged. W , which ranges from 0 to 1, is closely related to the average Spearman rank correlation among the judges' ranks (Hays, 1981).

⁷ If $\log(SD) = a + b \log(M)$, then Dixon (1985) recommends the transformation $y = x^{(1-b)}$. This formula was used whenever transformations were required.

catures, and .27 for the $-.50$ anticaricatures. However, there was no effect of exaggeration level on the proportion of faces recognized, $F(2,38) < 1$. Therefore, the hypotheses that the caricatures would be at least as identifiable as the veridical drawings, and that both would be more identifiable than the anticaricatures, was not supported by the accuracy data.

Subjects found many of the drawings difficult to recognize. This was not due to lack of familiarity with the faces used (see later). Rather, the drawings themselves, even the veridical ones which were essentially tracings of the photographs, were not easily identified. This may have been due to the schematic nature of the drawings. Information about hair color and texture, both of which are normally important cues to identity, was absent. Nor did the drawings contain shading, an important depth cue, although facial lines provided some depth information. One subject did extremely well on the recognition test, identifying 100% of the caricatures, 83% of the veridical drawings, and 50% of the anticaricatures. This subject was an amateur artist who draws faces frequently. He may therefore have been particularly familiar with how lines can be used to depict faces, and so have had less difficulty interpreting the drawings than the other subjects. Alternatively, he may have had especially good visual memory for faces.

Reaction times. A two-way repeated measures analysis of variance was carried out on the median reaction times (seconds) for correctly identified faces. Norm (av3, avpsych) and exaggeration level ($-.50$, 0 , $.50$) were within-subject factors. The data were first scaled using the transformation $y = x^{0.17}$ to remove a linear relationship between $\log(M)$ and $\log(SD)$. There was a significant main effect of exaggeration level, $F(2,38) = 28.87$, $p < .0001$. There was no effect of norm, $F(1,19) < 1$, nor any interaction between norm and exaggeration level, $F(2,38) < 1$. The means of the median reaction times (untransformed) for each exaggeration level are shown in Table 1. The .50 caricatures were identified most

TABLE 1
Mean Median Reaction Times (Seconds) to Identify Drawings at Different Levels of Exaggeration

	Norm	Exaggeration			
		Mean	SE	Mean	SE
		.00			
		Mean	SE	Mean	SE
Av3		8.9	1.2	6.3	1.8
Avpsych		12.5	2.5	6.0	1.5
Pooled across norms		12.3	2.3	6.4	1.3
		+.50			
		Mean	SE	Mean	SE
Av3				3.2	0.7
Avpsych				3.4	0.7
Pooled across norms				3.2	0.4

Note. SE = standard error. The nonlinear relationship between the individual and pooled norms occurs because these data are the means of median reaction times.

quickly, followed by the veridical drawings, and the $-.50$ anticaricatures, which were identified very slowly.

Planned comparisons were carried out to test the hypotheses that the caricatures would be identified more quickly than the veridical drawings, and that both would be identified more quickly than the anticaricatures. These tests were carried out on the pooled data from the two norm conditions, which gave mean reaction times of 3.2 s ($SD = 1.7$ s), 6.4 s ($SD = 6.0$ s), and 12.3 s ($SD = 10.4$ s), for the caricatures, veridical drawings, and anticaricatures, respectively. The data were first scaled using the transformation $y = x^{-.36}$ to remove a linear relationship between $\log(M)$ and $\log(SD)$. Caricatures were recognized more quickly than veridical line drawings, $t(19) = 2.18$, $p < .025$, one tailed, and veridical drawings were recognized more quickly than anticaricatures, $t(19) = 3.48$, $p < .005$, one tailed.

In summary, although caricatures were not identified more accurately than veridical drawings, they were identified more quickly. Veridical drawings in turn were identified more quickly, but not more accurately than anticaricatures. These results support the caricature and distinctiveness hypotheses, respectively. The large difference in reaction times between the caricatures ($M = 3.2$ s) and the anticaricatures ($M = 12.3$ s) suggests that different identification strategies were used. Long reaction times for the anticaricatures may indicate that a serial search in memory of all possible faces in the set was used, whereas the shorter reaction times for the caricatures may indicate that these drawings were identified spontaneously. However, there was no difference in the number of identification attempts for the three types of drawing ($M = 1.2$, anticaricatures; $M = 1.1$, veridical drawings; $M = 1.2$, caricatures), $F(2,38) < 1$. [The data were first scaled using the transformation $y = x^{-.1}$ to remove a linear relationship between $\log(M)$ and $\log(SD)$]. The relatively long reaction times for the veridical drawings ($M = 6.4$ s) confirm the identification difficulty indicated by the accuracy data. There was no evidence that the type of norm affected identifiability of the caricatures.

Likeness Judgments

Best drawing. The mean exaggeration level of the drawing chosen as the best likeness was $+.16$ for the av3-norm-based faces, and $+.15$ for the avpsych-norm-based faces. Both values were significantly greater than zero, $t(19) = 6.48$, $p \ll .001$ (av3) and $t(19) = 4.23$, $p < .001$ (avpsych). Thus, the best likeness was not a veridical drawing, but rather was a drawing in which the distinctive aspects had been exaggerated by approximately 16%. It should be noted, however, that this is an interpolated value, and that such a drawing was never shown to the subjects.

In this and all subsequent analyses faces rated as 1 on the familiarity

scale (i.e., never seen) were excluded. By this criterion 0.9 faces per subject were excluded ($SD = 1.5$). Faces for which all drawings were rated 1 (very poor) on the goodness of likeness scale were also excluded. This led to the exclusion of 0.7 faces per subject ($SD = 1.3$).

Ratings. A three-way repeated measures analysis of variance was carried out on the mean ratings assigned to the drawings. An identical analysis on mean ranks gave the same pattern of results and these are not reported here. Familiarity (low or high), norm (av3 or avpsych), and degree of exaggeration ($-.75$, $-.50$, $-.25$, 0 , $.25$, $.50$, $.75$) were within-subject factors. There was no significant main effect of familiarity, nor were there any significant interactions involving this factor. Therefore, mean ratings were recalculated from the raw data collapsing across familiarity levels, and a two-way repeated measures analysis of variance was carried out with norm and exaggeration as within-subject factors. There was a significant main effect of exaggeration, $F(6,114) = 80.20$, $p \ll .0001$, and a significant interaction of norm and exaggeration, $F(6,114) = 3.98$, $p < .001$. There was no main effect of norm, $F(1,19) < 1$. Figure 6 shows the mean ratings for each exaggeration level for each norm separately. Examination of Fig. 6 shows that the interaction resulted from higher ratings for the av3 than the avpsych norm faces at the 0 and .25 levels of exaggeration, and vice versa for all other exaggeration levels. This has little substantive significance, especially in view of the fact that the 0 exaggeration (veridical) drawings are identical in the av3 and avpsych norm conditions.

From Fig. 6 it can be seen that the 0 and .25 exaggeration levels obtained the highest ratings. Thus the veridical and .25 caricatures were considered the best likenesses, although both received ratings that were only moderately good. A planned comparison was carried out to compare the performance of the best caricature (.25) with the equivalent anticaricature ($-.25$) and the veridical line drawing. It was hypothesized that the caricature would be rated as high or higher than the veridical line drawing, and be rated higher than the anticaricature. The contrasts 0, 0, $-.2$, $+1$, $+1$, 0, 0 (applied to the $-.75$, $-.50$, $-.25$, 0 , $.25$, $.50$, and $.75$ drawings, respectively) were used to test the hypothesis that the .25 caricature and the veridical line drawing were considered better likenesses than the $-.25$ anticaricature.⁸ These comparisons were carried out separately for the av3 and avpsych norm faces, because of the interaction between norm and exaggeration level. The .25 caricature and the veri-

⁸ The contrasts were used to obtain a single number for each subject, according to the formula $X = \sum c_i x_i$, where c_i are the contrasts, and x_i are the mean reaction times for each exaggeration level. A repeated measures t test was then carried out on these numbers. Other comparisons using contrasts, discussed below, were carried out in the same manner.

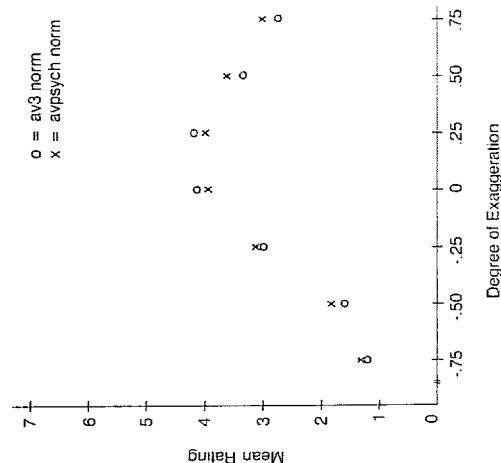


FIG. 6. Likeness judgments: Mean rating versus exaggeration level for each type of norm face.

dical line drawing were rated significantly higher than the $-.25$ anticaricature, both when the av3 norm was used, $t(19) = 7.60, p \ll .001$, and when the avpsych norm was used, $t(19) = 5.98, p \ll .001$. After the $.25$ caricature and the veridical drawing, the next most highly rated drawings were the $.50$ caricatures. For the avpsych faces there was no significant difference between the ratings for the $.50$ caricatures and the veridical line drawings, $t(19) = 1.91$, ns. However, for the av3 norm condition the $.50$ caricatures were rated significantly lower than the veridical line drawings (and hence the $.25$ caricatures, which were rated slightly higher than veridical drawings), $t(19) = 5.03, p \ll .001$. In the avpsych, but not the av3 norm conditions, the $.50$ caricatures were rated higher than the $-.25$ anticaricatures, $t(19) = 2.63, p < .02$. In sum, for one of the norms used the $.50$ (as well as the $.25$) caricatures were considered as good likenesses as the veridical line drawings. These results provide no evidence for a clear superiority of one type of norm over the other.

Only a third of the faces were actually identified, and this was independent of the caricature level at which they were shown. If one considers likeness ratings for only those faces that were correctly identified, then $.50$ caricatures were rated as good likenesses as veridical drawings. Mean likeness ratings were 5.1 ($SD = 1.3$), 5.2 ($SD = 1.1$), and 2.1 ($SD = 1.2$) for the caricatures, veridical drawings, and anticaricatures, respectively.

It is interesting to note that the ratings given to the veridical drawings and the $.25$ caricatures are higher than might be expected given the initial

identification difficulties. This result partly reflects the fact that faces for which all the representations were rated 1 (*poor*) were excluded from this analysis. More significantly, once subjects were told who a drawing represented, they almost always saw the resemblance. This knowledge, and perhaps also the greater familiarity with the medium by this stage in the study, may account for the moderately good likeness ratings.

It seemed possible that the initial distinctiveness of a face might influence the effect of caricaturing it. However, there was only a limited range of distinctiveness in these faces. The mean typicality rating on a 7-point scale for the five most distinctive faces was 3.3 ($SD = 0.4$), and for the five least distinctive faces was 4.8 ($SD = 0.3$). We found no difference in the effect of caricature level for these two sets of faces.

The mean goodness of likeness rating for the photographs was 5.1 ($SD = 0.6$), indicating that the photographs were quite good likenesses of the people. However, some subjects commented that the posers (all of whom had neutral expressions) did not display characteristic expressions. More identifiable drawings would be expected to result if the original photographs displayed characteristic expressions.

Ten of the 20 subjects knew or guessed that the stimuli were varied with respect to distinctiveness. However, only two subjects knew or guessed the hypotheses of the experiment, so the subjects may be regarded as substantially naive. Exclusion of the data from these two subjects did not affect any of the major results.

To summarize, both the distinctiveness and the caricature hypotheses were supported by the identification response time results. Caricatures ($.50$) were identified more quickly than veridical drawings, which were in turn identified more quickly than anticaricatures ($-.50$). Moreover, the mean exaggeration level of the drawing chosen as the best likeness was $.16$, which was a significant exaggeration relative to the veridical drawing. Although the best likeness was significantly exaggerated relative to the veridical drawing, this optimal level of exaggeration ($.16$) was not in fact included in the set of drawings shown to subjects. The closest drawings presented were the $.25$ caricatures and the veridical drawings. The veridical drawings and the $.25$ caricatures (and the $.50$ caricatures based on the avpsych norms) were considered to be the best, and equally good, likenesses of the faces of familiar people. These were followed by the $.50$ caricatures (based on av3 norms) and the $-.25$ anticaricatures.

In likeness judgments veridicality or realism constraints may be operating in opposition to the recognizability criterion. Caricatured drawings look less like real faces than do veridical drawings and anticaricatures. This is because the exaggeration used to produce the caricatures is unconstrained, so that the relationship between lines may be reversed (parts

of the bottom lip may rise above the top lip, an eye may rise above an eyebrow, etc.). Even when lines do not cross each other in impossible ways, parts of the face may be exaggerated beyond normal limits, particularly the lines affecting hairstyle. Nevertheless, for all exaggeration levels, the caricatures were considered better likenesses than the corresponding anticaricatures. The likeness results clearly support the distinctiveness hypothesis and offer some support for the caricature hypothesis.

EXPERIMENT 2

It could be argued that the composition of the stimulus sets in Experiment 1 (a progression of exaggeration levels from $-.75$ to $+.75$) biased judgments against the anticaricatures. Anticaricatures, which are similar to the norm face, look similar for the different faces, and this may be expected to bias choices against them. However, there are also factors that might be expected to bias choices against the caricatures. For example, as noted above, the caricatures may look less like real faces than the anticaricatures.

Experiment 2 was designed to determine whether biases, such as those described above, were operating in Experiment 1. These biases should be independent of familiarity with the faces. Therefore subjects who were unfamiliar with the faces used in Experiment 1 were asked to make likeness judgments in the presence of photographs of the people.

An immediate memory condition was also included to obtain preliminary data about whether a caricature advantage would occur for a newly encoded face or whether, as seemed more likely, such an advantage would require greater familiarity with a face. This experiment also provided a check that the veridical drawings were in fact veridical, and would be judged as the best likenesses in the presence of the photographs.

The caricatures and anticaricatures are equally similar to the veridical drawing metrically, but it is important to ensure that they would be perceived as being equally similar to it when the face is unfamiliar. This was tested by having subjects judge the similarity of the .25 caricatures and $-.25$ anticaricatures to the veridical line drawings.

Experiment 2 should also provide preliminary data about how much information about a face is necessary to obtain a caricature advantage. Likeness judgments were compared after experience with a single full-face photograph and with five views (left and right profiles, left and right three-quarter views, and the full-face view), which provide more information about the face. We predicted that any caricature advantage would be more likely to occur in the multiple view condition, which provided more information about the three-dimensional structure of the faces.

Method

Subjects

Twenty Stanford undergraduates (different from those who participated in Experiment 1), who were unfamiliar with the faces, participated in this experiment.

Stimuli

The .25 caricatures, the $-.25$ anticaricatures, and veridical line drawings from Experiment 1 were used. For 8 of these faces, 1×1 in. photographs showing the left and right profile and left and right three-quarter views, and a different (but highly similar) full-face view than the one used to make the drawings, were available. This set of photographs was intended to provide more information about the person's appearance than a single full-face view. For the remaining 10 faces only the full-face photographs, from which the line drawings were made, were available. These photographs were shown for those faces. The number of views shown (one or five) was confounded with the identity of the face, and with the size of the photograph(s) (the multiple views were 1×1 in., whereas the single, full-face photographs were 3.5×5 in.).

Procedure

Likeness judgments from direct comparison and immediate memory. Undergraduates, who were unfamiliar with the faces, ranked and rated the veridical line drawing, the .25 caricature, and the $-.25$ anticaricature for each face in terms of how good a likeness each was. The judgments were made by comparing the line drawings with the photograph(s) of the person depicted. Half the subjects made the likeness judgments in the presence of the photograph(s) (direct comparison condition), and half viewed the photograph(s) for 60 s, and then made the judgments from immediate memory (memory condition). Half the subjects judged set A faces and half judged set B faces (see Experiment 1). This was counterbalanced with judgment condition (perception/memory). The order in which the faces in each set were viewed was reversed for half the subjects.

Similarity judgments. For each triple of drawings, subjects judged whether the $-.25$ anticaricature and the veridical line drawing, or the .25 caricature and the veridical line drawing, were more similar. The three drawings were arranged from left to right. The veridical drawing always appeared in the center, with the .25 caricature on the left half of the time and on the right half of the time. The left-right arrangement of the caricature and the anticaricature was counterbalanced with the norm face used.

Half the subjects did the similarity judgments before the likeness judgments, and half did the likeness judgments first. Two subjects were tested at the same time. They worked independently. One did the likeness judgments first, and the other did the similarity judgments first. One had the perception condition and the other had the memory condition. Half the subjects had set A faces for the likeness judgments and set B faces for the similarity judgments. Half the subjects had the reverse arrangement. Thus identical pictures were not used for the two tasks. The assignment of face sets to the two tasks was counterbalanced with the order of the tasks. (Assignment of face sets could not be counterbalanced with both order of tasks and perception/memory condition, because N was not a multiple of 8. However, there were equal numbers of each assignment condition in the memory and perception conditions.) The order in which the faces in each set were viewed was reversed for half the subjects. This was counterbalanced with the order of the likeness and similarity judgments, and the assignment of face sets to these tasks.

Results and Discussion

Likeness Judgments from Direct Comparison and Immediate Memory

Ratings. A four-way analysis of variance was carried out on the mean ratings, with judgment condition (direct comparison or immediate memory) as a between-subjects factor, and norm, number of photographs given (one or five), and exaggeration level ($-.25, 0, .25$) as within-subject factors. An identical analysis on the rank data gave the same pattern of results, and these are not reported here. There was a significant main effect of exaggeration, as predicted, $F(2,36) = 16.59, p \ll .0001$. The veridical drawings received the highest ratings ($M = 4.12$), followed by the $-.25$ anticaricatures ($M = 3.75$), and $.25$ caricatures ($M = 3.29$). Planned comparisons show that the veridical drawings were rated significantly higher than both the caricatures, $t(38) = 2.69, p < .02$, and the anticaricatures, $t(38) = 5.76, p \ll .001$. This result validates the assumption that the veridical drawings were in fact veridical. The veridical drawings, which received the highest ratings, were rated as only moderately good. There was no significant difference between the caricatures and the anticaricatures, according to a Scheffé post hoc comparison, $t(38) = 3.07, ns$. These comparisons were based on the mean ratings obtained by collapsing across all levels of the other factors, none of which interacted with exaggeration level. There was a significant main effect of judgment condition, $F(1,18) = 5.17, p < .04$, with the drawings rated as better likenesses when compared directly with the photographs ($M = 4.08$) than when judged from immediate memory ($M = 3.36$). There was also a significant main effect of number of photographs, $F(1,18) = 6.63, p < .02$. The mean rating when a single, large, full-face photograph was seen was higher ($M = 4.19$) than when multiple small views, with a different full-face view than used to generate the drawings, were seen ($M = 3.97$). All but three subjects reported that they looked mostly or entirely at the full-face photograph when multiple views were given, because this was the view depicted in the drawings. Therefore, the attempt to manipulate the degree of information subjects had about the three-dimensional structure of the face was unsuccessful.

The rating results rule out the possibility that the likeness judgment results in Experiment 1 were due to a bias against the anticaricatures. There was no significant difference in the ratings given to the $.25$ caricatures and $-.25$ anticaricatures, according to the post hoc test used. Had we used a less conservative test (e.g., the usual significance level), the bias would have favored the anticaricatures, not the caricatures.

The absence of a caricature advantage in the immediate memory condition (with either single or multiple views of faces) suggests that for rela-

tively unfamiliar faces there is no caricatured information in memory competing against veridicality considerations when judging goodness of likeness. For familiar faces, however, caricatured information appears to compete with veridicality considerations resulting in the $.25$ caricatures and the veridical drawings being judged equally good likenesses (Experiment 1). The effectiveness of caricatures as representations appears to require some minimal level of familiarity with the face, not attained with 60-s exposure to photographs. Unfamiliar faces may be inappropriate for testing the caricature hypothesis, because the advantage of caricatures over anticaricatures and veridical line drawings appears to depend on properties of the long-term memory representation not shared by a perceptual or short-term memory representation.

Similarity Judgments

Subjects judged which of two pairs of drawings were most similar, the $-.25$ anticaricature and the veridical drawing, or the $.25$ caricature and the veridical drawing. The mean number of choices in favor of the latter pair was 8.4 ($SD = 3.6$), out of a total of 18. This did not differ significantly from 9.0, which was the expected mean if the veridical drawings were equally similar to both the anticaricature and caricatures, $t(19) < 1$. Thus the caricatures and anticaricatures did not differ in perceived similarity to the veridical drawings. This result shows that the metric equivalence of caricatures and anticaricatures to the veridical drawings results in perceived equivalence of similarity of the caricatures and anticaricatures to the veridical drawings, at least when the faces are unfamiliar.

GENERAL DISCUSSION

Identification of the veridical drawings was much less accurate and slower than would be expected for photographs. Photographs contain more information, especially three-dimensional information provided by shading, than do line drawings. Even line drawings differ in detail. Davies, Ellis, and Shepherd (1978) showed that photographs of famous people were better recognized (90% accuracy) than were detailed line drawings (47%), which in turn supported recognition better than did line outlines (23%). Our veridical line drawings contained about as much detail as their line outlines. In addition, the definition of each face in terms of the same 169 points and 37 lines led to a sameness of the drawings. Nevertheless, once subjects knew who a drawing represented, they generally saw the resemblance, and made meaningful ratings of goodness of likeness.

Distinctiveness Hypothesis

When the faces were familiar, caricatures of every degree were judged better likenesses of the person depicted than were anticaricatures of the same degree. The .50 caricatures were identified four times more quickly than the -.50 anticaricatures. These results show that exaggerating distinctive aspects of a face produces a better likeness than decreasing distinctiveness by an equal amount. The distinctiveness hypothesis has received previous support when distinctiveness was varied across faces (e.g., Light et al., 1979), but has not before been demonstrated for distinctiveness varied within individual faces.

Caricature Hypothesis

When the faces were familiar, .50 caricatures were identified twice as quickly as the veridical line drawings. Further, .25 caricatures (and .50 caricatures for one norm) were considered to be as good likenesses as the veridical drawings, with the interpolated *best* representation a .16 caricature. Others have been unable to demonstrate such an effect (Hagen & Perkins, 1983; Tversky & Baratz, 1985). As we argued above, our success was most likely due to the use of familiar faces and comparable representational media for the veridical and caricatured images. It also appears that a reaction time measure is more sensitive than an accuracy measure, at least in an identification task. Considerations of veridicality or accuracy of depiction are likely to influence likeness judgments. Therefore, equivalent ratings for caricatures and veridical drawings suggests that there is some caricature advantage competing with the veridicality considerations.

Interpretation of a Caricature Advantage

There are two possible models that could account for the caricature advantage. One is that long-term memory representations are schematized and exaggerated like caricatures (Model 1). The other is that the memory representations are veridical, but that exaggerated representations are more effective at accessing those representations (Model 2). Perhaps the distinctive aspects of a face are simply made more salient in a caricature. In this case caricatures would be recognized better than anticaricatures, but it is not clear that they would be recognized better than veridical drawings, because there would be a *mismatch* on the distinctive aspects. However, if identification occurs when the *relative* level of activation of a face within a set exceeds some criterion, then a caricature might be identified better than a veridical image. This is because a caricature, which is more individuated, and therefore less similar to other faces, might produce a higher *relative* activation level for the target face

than would a veridical image, even when the memory representation is veridical. This advantage for the caricature would occur if there was a monotonically increasing function relating activation to similarity, so that the drop-off in activation for nontarget faces would be greater than that for the target face. In this case caricatures could be more easily identified than veridical images, even though the memory representations themselves were "veridical" (although it is not clear why identification time and accuracy should behave differently). Similarly, atypical faces, for which there are few similar faces in memory, would be more recognizable than typical faces, because activation of the memory representations of such faces will be high relative to that of other (similar) faces in memory.

Is it possible to distinguish between these two models? The fact that caricatures were not considered as good likenesses as veridical drawings for unfamiliar faces seems more consistent with Model 1 than Model 2.⁹ According to Model 2, caricatures should access the veridical representations more effectively than veridical images, irrespective of the familiarity of the face, and therefore be considered better likenesses (a good likeness was defined as being recognizable). However, the veridical drawings were considered the best likenesses. This means that either the process by which memory representations are accessed must change as faces become more familiar, or the memory representations themselves must change. There is no evidence that the access process changes with familiarity of the stimulus. However, there is evidence that the memory representations of faces change with increasing familiarity. Ellis, Shepherd, and Davies (1979) found that famous people were identified more accurately from internal features than from the external face and hair outline, whereas relatively unfamiliar faces were recognized equally well from either internal or external features. People can also discriminate between normal and mirror-reversed images of faces better for more familiar faces than for less familiar faces, suggesting that subtle configural information about the left-right organization of a face is represented better for more familiar faces (Rhodes, 1986). Thus there appears to be more evidence in favor of the hypothesis that memory representations change as faces become more familiar than that the representations are accessed differently.

A further test of the hypothesis that memory representations are caricatured would be to get artists to draw faces from memory, as veridically as possible. Their drawings could then be compared with veridical and caricatured drawings produced by the caricature generator (based on

⁹ A recognition test would provide stronger evidence. However, Experiment 2 was not designed to differentiate these two interpretations of the caricature advantage, and did not include a recognition test.

photographs). The comparison could be made by computing the average distance between corresponding fiducial points in the artists' drawings and in the veridical drawings and caricatures.

Ratings Versus Reaction Times for Identification

Taking reaction times for identification as an indicator of "goodness of likeness," the .50 caricatures are better likenesses than are the veridical line drawings. A different pattern emerged when ratings of goodness of likeness were analyzed. Here, caricatures merely equaled veridical line drawings, and for one norm it was only the .25 caricatures (not the .50 caricatures) that did so. These differences suggest that different processes underlie identification and ratings. This suggestion is supported by another aspect of our data. Neither the distinctiveness nor the caricature hypothesis was supported when subjects made *likeness* judgments of unfamiliar faces (Experiment 2). However, *recognition* studies with unfamiliar faces provide ample support for the distinctiveness hypothesis (see introduction). Subjects apparently have veridical information about faces that they can use to judge how realistic a face is (i.e., whether it conforms to the usual metric proportions), but that they do not use in on-line identification or recognition.

Norm-Based Representations of Distinctive Information

The idea that distinctive aspects of a face are noted or even exaggerated in the memory representation requires knowledge about what is typical in a population. This knowledge may be represented as a norm face, like the norm face in the caricature generator, against which distinctiveness is assessed. One might think that an explicit measure of variability would be necessary for assessing distinctiveness. However, the computational theory on which the caricature generator is based indicates that an explicit representation of population variability information is unnecessary. Distinctiveness may simply be assessed in relation to the norm face, as it is in the caricature generator. Those aspects that are most different from the norm face are most distinctive, and may receive differential emphasis in the representation, or be most easily accessed in recognition.

Knowledge about typicality appears to be population based. People recognize faces of their own racial group better than faces in other racial groups (see Shepherd, 1981, for a review), presumably because for an unfamiliar population they have inadequate knowledge about what is typical and atypical, or even about what aspects of faces show variation across individuals. With more experience or training, recognition of unfa-

miliar race faces improves (Elliott, Wills, & Goldstein, 1973). It therefore seems likely that people have several norms, one for each category of faces that has unique individuating properties. Obvious examples are different sex, age, and racial groups. We plan to use the caricature generator to test the hypothesis that there are different norms for different categories of faces. If the hypothesis is true, then the use of an inappropriate norm should produce poorer caricatures than use of the correct norm.

A Holistic Model of Encoding and/or Recognition

A major question about how faces are mentally represented is whether they are represented holistically or as sets of piecemeal features or dimension values. A related question is whether faces are recognized or identified by holistic comparisons with memory representations, or by feature analytic comparisons. These questions have not proved easy to deal with, at least in part because of lack of clear definitions of terms like *holistic*. The caricature generator provides an explicit statement of what is meant by a holistic comparison. It is one in which corresponding points can be found and simultaneously compared on two representations.¹⁰ The caricature generator works by holistically comparing a face with a norm face, and exaggerating all metric differences between the two without any explicit analysis of distinctive features. We call this model "holistic" rather than "featural" because it does not constrain and manipulate a face in a piecemeal fashion. The encoding preserves the continuously varying, analog terrain of the face. All points on a veridical face are treated alike with respect to a norm; the whole representation is exaggerated in parallel, within a many-dimensional space. An individual face can be considered to occupy a unique location in this many-dimensional space, where the dimensionality is equal to the total number of points in the representation $\times 2$ (because x and y coordinates are given for each point).

The points in the representation could be called features, but their exact nature is somewhat arbitrary. In principle, any set of lines or points sufficient to be identified as a face could be used to define the initial representation. What is critical is that this set be congruent across all the faces in the data base. The fact that the caricatures were more identifiable than the veridical drawings suggests that an analysis of a face into discrete "features" may not, in principle, be necessary for representation. Faces could be encoded by holistic comparison with a norm face, just as

¹⁰ It should be noted that this description is at the level of a computational theory. The comparisons need not be carried out simultaneously in the nervous system, just as they are not carried out simultaneously in the computer implementation.

in the caricature generation algorithm. In this sense, they may be considered to be encoded holistically. Recognition might also be based on a holistic comparison of a face to faces in memory, similar to that used by the caricature generator.

Norm-Based Object Categorization

The caricature generator may provide a theory of categorization for objects other than faces that can also be defined in terms of a fixed set of points that are common to all objects in the class to which they belong. For example, different types of dogs or birds could be encoded by shape profiles, which represent the distinctive aspects of the kind of dog or bird relative to a dog or bird norm (based on the average of all the dog or bird profiles). This hypothesis is consistent with the important role of shape information in basic level categorization. Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) have shown that basic level objects can be identified from averaged shapes of members of that class. Here, we are suggesting that subordinate level classification could be made on the basis of deviations from the basic level norm (averaged) shape. Although discrete features are clearly used in object recognition, our theory relates to the way in which information provided by configural shape may be used.

REFERENCES

- Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge Univ. Press.
- Bartlett, J. C. Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition*, 12, 219-228.
- Brennan, S. E. (1982). *Caricature generator*. Unpublished master's thesis, MIT, Boston.
- Brennan, S. E. (1985). The caricature generator. *Leonardo*, 18, 170-178.
- Bruce, V. (1982). Changing faces: Visual and nonvisual coding processes in face recognition. *British Journal of Psychology*, 73, 105-116.
- Cohen, M. E., & Carr, W. J. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society*, 6, 383-384.
- Davies, G., Ellis, H., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, 63, 180-187.
- Dixon, W. J. (1985). *BMDP statistical software manual*. Berkeley: Univ. of California Press.
- Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of white and oriental faces. *Bulletin of the Psychonomic Society*, 2, 71-73.
- 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.
- Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology*, 90, 65-74.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
- 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.

- Hagen, M., & Perkins, D. (1983). A refutation of the hypothesis of the superfluidity of caricatures relative to photographs. *Perception*, 12, 55-61.
- Hays, W. L. (1981). *Statistics*. New York: Holt, Rinehart & Winston.
- Kanade, T. (1973). *Picture processing system by computer complex and recognition of human faces*. Unpublished doctoral dissertation, Kyoto University, Japan.
- Kelly, M. D. (1970). *Visual identification of people by computer*. Stanford Artificial Intelligence Project Memo AI-130. Stanford University, Computer Science Department, Stanford, CA.
- Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 212-228.
- Perkins, D., & Hagen, M. (1980). Convention, context, and caricature. In M. Hagen (Ed.), *The perception of pictures* (Vol. 1). New York: Academic Press.
- Rhodes, G. (1986). Memory for lateral asymmetries in well-known faces: Evidence for configural information in memory representations of faces. *Memory & Cognition*, 14, 209-219.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Shepherd, J. (1981). Social factors in face recognition. In G. Davies, H. Ellis, & J. Shepherd, (Eds.), *Perceiving and remembering faces*. New York: Academic Press.
- Shepherd, J., Davies, G., & Ellis, H. (1981). Studies of cue saliency. In G. Davies, H. Ellis, & J. Shepherd. (Eds.), *Perceiving and remembering faces*. New York: Academic Press.
- Solso, R. L., & McCarthy, J. E. (1981). Prototype formation of faces: A case study of pseudo-memory. *British Journal of Psychology*, 72, 499-503.
- Tversky, B., & Baratz, D. (1985). Memory for faces: Are caricatures better than photographs? *Memory & Cognition*, 13, 45-49.
- Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 181-190.

(Accepted February 19, 1987)