Parts and Wholes in Face Recognition

James W. Tanaka

Oberlin College, Oberlin, Ohio, U.S.A.

Martha J. Farah

University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.

Are faces recognized using more *holistic* representations than other types of stimuli? Taking holistic representation to mean representation without an internal part structure, we interpret the available evidence on this issue and then design new empirical tests. Based on previous research, we reasoned that if a portion of an object corresponds to an explicitly represented part in a hierarchical visual representation, then when that portion is presented in isolation it will be identified relatively more easily than if it did not have the status of an explicitly represented part. The hypothesis that face recognition is holistic therefore predicts that a part of a face will be *disproportionately* more easily recognized in the whole face than as an isolated part, relative to recognition of the parts and wholes of other kinds of stimuli. This prediction was borne out in three experiments: subjects were more accurate at identifying the parts of faces, presented in the whole object, than they were at identifying the same part presented in isolation, even though both parts and wholes were tested in a forced-choice format and the whole faces differed only by one part. In contrast, three other types of stimuli—scrambled faces, inverted faces, and houses—did not show this advantage for part identification in whole object recognition.

Parts and Wholes in Face Recognition

An important issue in face recognition research is whether faces are recognized on the basis of their individual features or more holistically, on the basis of their overall shape. As long ago as the nineteenth century, Galton (1879) proposed that holistic information may be more vital to face recog-

nition than the identification of individual features, and modern re-searchers continue to pursue this hypothesis (see Bruce, 1988, for a detailed review and evaluation). However, the empirical evidence to substantiate such a claim remains equivocal.

One factor that has contributed to the difficulty of resolving this issue is the lack of clear, generally accepted definitions of the concepts *holistic* and *featural*. Without clear definitions of what these terms mean, it is difficult to operationalize them in experimental tests. In this article, we propose an explicit definition of the holistic/featural distinction. We then use that definition to interpret the available evidence and to design new empirical tests.

We take as a starting point the idea that visual object representations are hierarchically organized, such that the whole object is parsed into portions that are explicitly represented as parts (cf. Palmer, 1977). For example, a house might be decomposed by the visual system into a set of doors, windows, a roof, etc. The resulting representation of the house would consist of representations of these parts, somehow linked together. Some objects may be decomposed into many parts, others into relatively few or none at all. In this context, the claim that faces are recognized holistically would mean that the representation of a face used in face recognition is not composed of representations of the face's parts, but more as a whole face. Although visual information from the eyes, nose, etc. would of course be included in the face representation, that information would not be contained in representational *packets* corresponding to the parse of the face into these features. In other words, these parts or features would not be explicitly represented as structural units in their own right in the final face representation. Instead, faces would be recognized "all of a piece"—or, to use a somewhat embattled term, as *templates*. The altern-ative hypothesis, that faces are recognized featurally, implies that faces are represented in terms of representations of their component parts. The holistic/featural distinction need not be a strict dichotomy, as both types of representations may exist and be used to different degrees for different classes of objects. Because of this, we would like to recast the question of whether faces are recognized holistically as the question: does face recog-nition rely on holistic visual representations to a greater degree than do other forms of pattern recognition? Before presenting our experiments, we briefly review what is known about this issue from other studies.

Bradshaw and Wallace (1971) addressed the issue of whether faces are perceived featurally using a matching task in which pairs of simultaneously presented Identikit faces were to be judged "same" or "different". They found that the number of features by which a pair of faces differed pre-dicted the latency of correct "different" responses, with shorter reaction times associated with more featural differences. Based on this finding, they argued that faces were inspected according to a serial self-terminating

search and were therefore perceived in terms of their features. In another study with Identikit faces in a simultaneous matching task, Mathews (1978) found evidence that faces are perceived both featurally and holistically. He found that subjects' reaction times increased linearly for detecting differences in eyebrow, nose, and mouth features, thus indicating a top-to-bottom serial comparison process. However, he also found that reaction times for detecting changes in hair, eyes, and chin were essentially the same across features, which he interpreted as evidence for a holistic or at least a parallel comparison process. Mathews reconciled these results by proposing a dual processing strategy in which features are checked both in serial and parallel.

Other researchers have also arrived at the conclusion that faces are perceived both featurally and holistically using slightly different paradigms and methods of data analysis. For example, Smith and Nielsen (1970) used a matching paradigm with schematic line drawings of faces, but introduced a delay between the two stimuli to be matched. At delays of 1 or 4 sec, they found results similar to those of Bradshaw and Wallace: the more features differed between two faces, the more quickly the faces were judged to be different. In addition, by varying the number of features present in the faces, they were able to examine the effects of number of features on the latencies of "same" judgments. They found that "same" judgments were not affected by the total number of features, conflicting with their findings from the "different" trials and suggesting that subjects were not serially comparing the individual features. However, at the longer delays of 10 sec, both "same" and "different" trials yielded patterns of reaction times consistent with a feature by feature comparison process. Sergent (1984) used a matching task with Photofit faces and reasoned that if features are processed independently, the time to make a "different" response when faces varied by two features should never be faster than the time to make a "different" response when faces varied by the most salient feature. Her data indicated that changes in chin contour led to faster reaction times than changes to either eyes or to a feature that she termed "internal spacing"—the distance between the nose and mouth. She found that the time to decide that two faces differed with respect to their chin contours and internal spacings was faster than the time to decide that two faces differed with respect to their chin contours only. Sergent concluded that whereas some features seemed to be processed independently of each other (e.g. eyes and chin contour), other features (e.g. chin contour and internal spacing) interact and are processed more holistically. However, as Bruce (1988) has pointed out, Sergent's conclusion was weakened by the fact that only one feature type—internal spacing—produced the holistic effect, and it was not a feature in the sense of being a part of the face but was, rather, a relation among parts.

The foregoing studies suggest that faces may be perceived both in terms

of their individual component features and in terms of more holistic ensembles of those features. However, these studies fall short of answering the question posed earlier—does face recognition rely on holistic visual representations to a greater degree than other forms of pattern recognition?—for several reasons.

First, with the exception of Sergent's study, all of the experiments appear to be based on the assumption that the number of features in a face, or the number of features by which two faces differ, would not affect performance if subjects were using a holistic representation. However, this is not necessarily true. The more features that are in a face, the more information there is in the holistic representation, and the longer it could take to use that holistic representation. Similarly, the more features differ between two faces, the more different their two holistic representations will be, and the more easily a discrepancy will be discovered. These studies might better be regarded as testing whether face matching is carried out independent of capacity limitations, regardless of whether the matching is holistic, parallel featural, or serial featural.

Second, none of the studies is designed to distinguish between the possibility that faces are represented by features that can be processed in parallel, and the possibility that faces are represented holistically, that is, without explicit representations of the features.[1] The question of whether facial features, when and if they are explicitly represented, can be compared in parallel or only serially is an interesting one, but not the one to which we are addressing ourselves in this article.

Third, it is not clear how similar the visual processes elicited by these tasks are to those used in normal face recognition. All of the studies described above involved face matching rather than face recognition. Subjects may well use different strategies and, as a result, different types of visual representations when they can consult a percept or short-term memory representation of the face to be matched, rather than having to consult the long-term memories used for face recognition. The generalizability of these studies to real face recognition can also be called into question on the grounds of the stimulus pictures used, some of which were highly artificial and schematic.

Finally, without comparing the results obtained with faces in these paradigms to results obtained with objects other than faces, we cannot assess the extent to which the holistic or featural representation of faces is special

---

[1]Some researchers have couched the question in terms of configurational versus featural processing, but these studies do not address this distinction either. Although parallel processing of multiple features is presumably necessary for configural processing, they are not the same thing. Features could be processed in parallel without representing their spatial configuration. We will return to the issue of configuration in face recognition and how it relates to the idea of holistic representation in the General Discussion.

to faces. A series of experiments by Bruce, Doyle, Dench, and Burton (1991) avoids many of these problems. They presented subjects with sets of computer-generated faces with identical features, but slightly different spatial configurations in an incidental memory task. They found that subjects abstracted the prototypical configuration for each set, and that this tendency to identify the prototype as most familiar was greater for faces than for houses. This finding argues strongly for a special role of non-featural information in face recognition. However, it does not speak directly to the issue of holistic representation.

Our approach to the issue of holistic versus featural representations in face recognition is based on the following logic: if some portion of a stimulus is explicitly represented as a part in the stimulus representation, then it should be relatively more easily recognized as coming from that stimulus, when viewed in isolation, than if the stimulus representation does not contain it as an explicitly represented part. Similar reasoning has been used by Bower and Glass (1976) and Palmer (1977) to distinguish between psychologically real and less plausible parsings of patterns into parts. Bower and Glass showed subjects a set of abstract line drawings and then asked them to reproduce these drawings given fragments of the drawings as memory-retrieval cues. Fragments that corresponded to "good" parts according to Gestalt principles, which were therefore hypothesized to be explicitly represented as parts in a hierarchical representation of the visual pattern, were more effective in cueing memory than were other equally large and complex fragments.

Palmer (1977) gathered converging evidence that certain portions of abstract geometric patterns were explicitly represented as parts, for example by asking subjects to divide the patterns into their natural parts or rate the goodness of parts. He then showed that portions of a pattern that appeared to be explicitly represented as parts according to these criteria were also more easily verified as coming from their whole patterns than portions that were not. A similar finding was obtained by Reed (1974), although the aim of his research was not to elucidate the part structure of patterns but, rather, the information available in mental images. Reed found that subjects were able to verify the presence of pattern fragments in their mental images of the whole pattern only when the fragments corresponded to "good" parts.

The research of Bower and Glass (1976), Palmer (1977), and Reed (1974) suggests that when a portion of a stimulus pattern is explicitly represented as a part in the subject's representation of the pattern, it will be better recognized as having come from that pattern than if it is not explicitly represented as a part. The experiments to be reported here make use of this finding as a way of testing the *parthood* and *objecthood* of visual stimuli. If a portion of a stimulus is represented as a part in the visual

representation of the stimulus that underlies recognition, it should be identified more accurately than if it does not have the status of a part in the stimulus representation.

In each of the three experiments to be reported, subjects learn to recognize a set of normal faces and a set of some contrasting class of stimuli: scrambled faces (Experiment 1), inverted faces (Experiment 2), and houses (Experiment 3). Subjects are trained so that they are at least as accurate at recognizing the normal faces as the contrasting stimuli. We can then compare the identification of isolated features from normal faces with the identification of isolated features from the contrasting classes of stimuli. For face stimuli, the tested parts were the facial features of the eyes, nose, and mouth. These facial features are not only the nameable parts of a face but also correspond to the natural parsings of a face based on the discontinuities of its contours (Biederman, 1987; Hoffman & Richards, 1984). If face recognition is more holistic than the recognition of other kinds of stimuli, then identification of isolated features from the normal faces should be disproportionately less accurate than identification of isolated features from the contrasting stimulus classes, relative to the identification of the part in the whole face and whole contrast object.

EXPERIMENT 1

In this experiment, subjects were asked to memorize intact and scrambled faces. Scrambled faces were chosen as a contrasting stimulus class because their parts are the same as the parts of a normal face, and yet we would not expect special, face-specific recognition abilities to be used in recognizing scrambled faces. After learning the normal and scrambled faces, subjects were given a forced-choice recognition task in which they identified facial features presented in isolation and in whole-face context. The whole-face test items were constructed such that the target and foil faces differed only with respect to the feature being tested. Examples of these two types of test for intact and scrambled faces are shown in Figure 1. In the isolated part test condition, subjects would be asked to identify "Larry's nose". In the full-face test condition, subjects would be asked to identify "Larry". Note that the only difference between the "Larry" target and foil in the whole face test is the nose feature. That is, the information available for making the discrimination was exactly the same: the face outline, hair, eyes, and mouth were held constant.

If the recognition of normal faces involves representing their component parts to the same degree as the recognition of scrambled faces, then we should expect that identification of the features of normal faces will be just as good relative to the identification of the whole face as identification of the features of scrambled faces are relative to the identification of whole scrambled faces. However, if normal faces are recognized more holistically

than scrambled faces, then there should be a disadvantage for identifying isolated features compared to whole faces for normal faces, relative to part and whole test performance for scrambled faces.
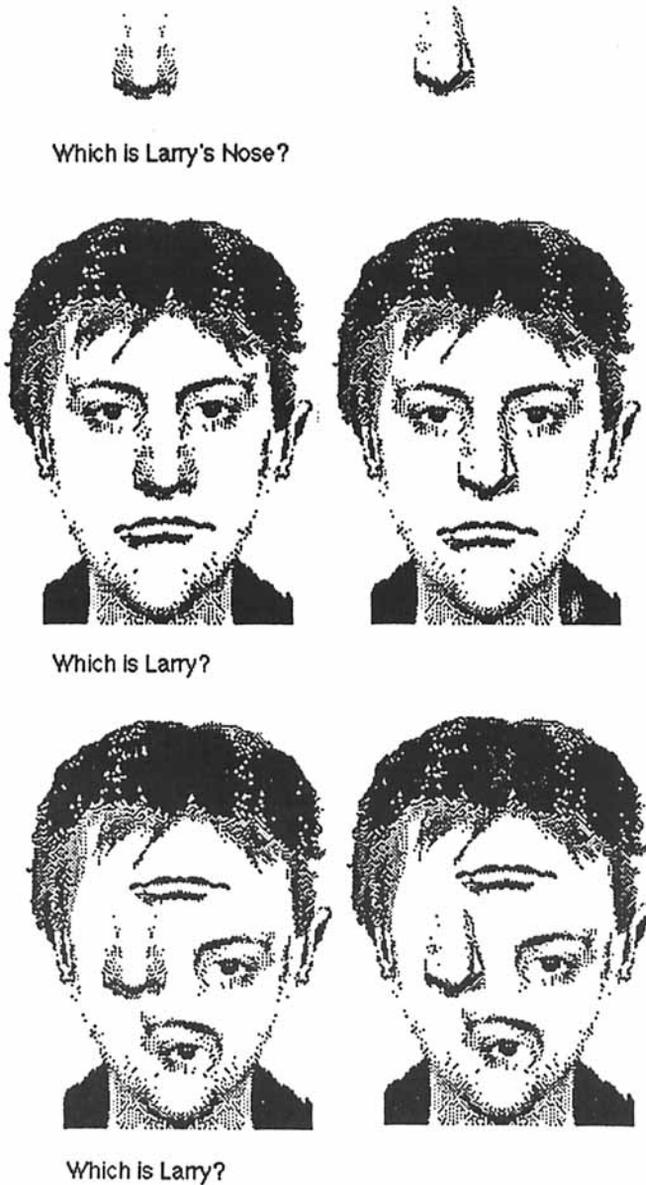

FIG. 1.   Example of isolated part, intact face, and scrambled face test items.

Method

Subjects

Twenty first-year psychology students from Carnegie-Mellon University served as subjects in the experiment. Subjects were tested individually and received course credit for their participation.

Materials

Stimuli consisted of two groups of six male faces that were generated on a Macintosh computer using a Mac-a-Mug program. Faces were constructed by selecting one of the three exemplars for each of the three feature types (e.g. eyes, nose, mouth).[2] The exemplars for one group of faces are shown in Figure 2. For both groups, exemplars were placed within the same face outline. Face stimuli were constructed such that no one exemplar was unique to a particular face, with each exemplar present in two of the six faces in the group. Scrambled and intact versions of each face were generated. For the scrambled faces, the spatial positions of the features were consistent across faces (e.g. the nose was always located below and to the left of the mouth). Half of the subjects saw one group of faces as the scrambled set and the other group as the intact set. For the other half of subjects, the versions of the face groups were reversed. Thus, each face appeared an equal number of times in its scrambled and intact version. Faces were photocopied onto 4" × 5" white card stock.
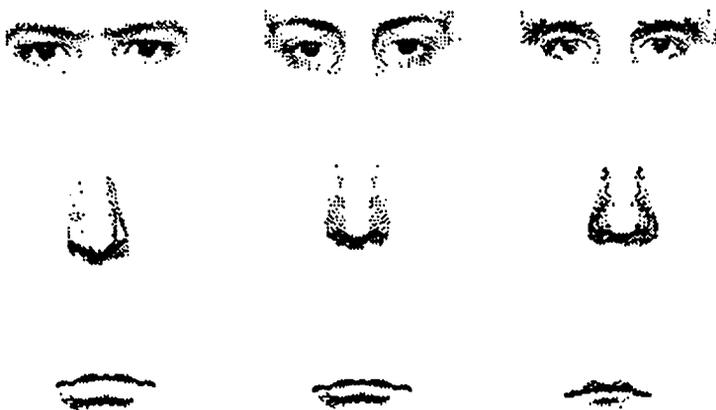
FIG. 2.   Eye, nose, and mouth exemplars used for one group of faces in Experiment 1.

[2]We will reserve the term "feature" to refer to a discrete part of a face (e.g. eyes, nose, and mouth) and the term "exemplar" to refer to a particular instance of a feature (e.g. long nose).

Procedure

Subjects were seated at a table directly facing the experimenter at a viewing distance of approximately 2 m. Subjects were informed that they would be shown pictures of scrambled and intact faces paired with male names and their task was to learn the correct face–name associations.

Learning Phase.   Learning and test trials were blocked according to face version (scrambled and intact). For each learning trial, the face stimulus was randomly presented for 5 sec accompanied by its verbally spoken name. Each learning block contained six learning trials, one trial per face. There were a total of five learning blocks per face version.

Test Phase.   Immediately following learning, a two-choice recognition test was administered. One feature from each of the learned faces was included in the recognition test. An equal number of eyes, nose, and mouth features were tested. In the isolated part test condition, subjects identified isolated features of the learned faces (e.g. which is Bob's nose?). Item foils were taken from one of the other learned faces. In the full-face test condition, subjects were shown the same target features and their foils presented in the full-face configuration (scrambled or intact) and asked to identify the face that matched the given name (e.g. which is Bob?). The target and foil faces differed only with respect to the individual feature that was tested in the isolated test condition; all other feature information was held constant. The full-face foil did not correspond to any previously learned face. Thus, subjects identified each feature twice, once presented in isolation and once presented in the full face (intact or scrambled face). After the learning and test phases for one of the face versions (scrambled or intact) was completed, subjects learned and were tested on the other version with the initial face version counterbalanced across subjects.

Results and Discussion

As shown in Figure 3, subjects were able to identify isolated parts from intact faces correctly on 62% of the trials. When the same parts were tested in the whole face, performance improved to 73%. For scrambled faces, there was a different pattern of results. Subjects were actually better at identifying the parts tested in isolation (71% correct) than tested in the whole face (64% correct). An analysis of variance (ANOVA) with face version (intact and scrambled), test type (isolated part and whole face), and facial feature (eyes, nose, mouth) as within-subjects factors confirmed this interaction between face version and test type, $F(1, 19) = 7.55$, $p < 0.02$. Direct comparisons between part- and whole-face performance showed that the part–whole difference was reliable for normal, intact faces,
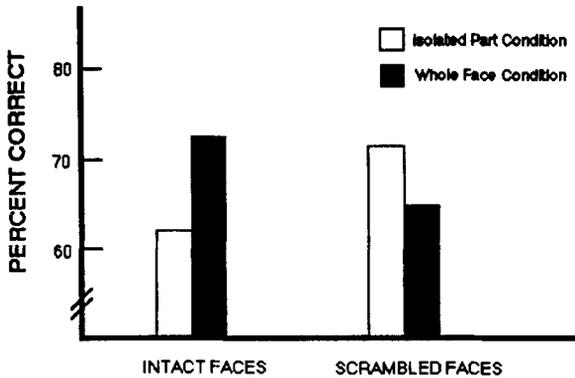
FIG. 3.   Percentage of correctly identified isolated part- and whole-face test items for intact and scrambled faces.

$t(19) = 2.16$, $p < 0.05$, but not for scrambled faces, $t(19) = 1.25$. The advantage of whole-face recognition for normal, intact faces over scrambled faces suggests that the normal face is mentally represented more in terms of a whole object (holistically) as compared to the representation of a scrambled face, which is more in terms of its parts (featurally).

The main effect of facial feature was also reliable, $F(2, 38) = 7.96$, $p < 0.01$. Subjects were more accurate making eye judgments (80% correct) than they were making nose judgments (62% correct) or mouth judgments (63% correct). This finding is consistent with previous results involving simultaneous matching tasks (Sergent, 1984; Walker-Smith, 1978) in which eye features were perceptually more discriminable than nose or mouth features. No other main effects or interactions were reliable, $p > 0.10$.

The results of Experiment 1 indicate that subjects are better at identifying facial features from normal faces when they are presented in the whole face than when they are presented alone, relative to recognition of facial features from scrambled faces when presented as isolated parts and wholes. This is true despite the fact that the whole-face test items had no more discriminating information in them than did the isolated parts: for each choice between isolated parts, the corresponding whole-face test items differed only by those same parts. This outcome is consistent with the hypothesis that normal faces are recognized more holistically than are scrambled faces. Note that this result can be interpreted in either of two ways. It can be argued that part representations are less available for normal faces relative to scrambled faces or that holistic representations are more available for normal faces relative to scrambled faces. Direct comparisons of part- and whole-face recognition performance for intact and scrambled faces suggests that the latter interpretation might be more

accurate. Although difference in part recognition for intact and scrambled faces was not reliable, $t(19) = 1.52$, whole intact face recognition performance was reliably better than whole scrambled face recognition, $t(19) = 2.07$, $p < 0.05$. Thus, the recognition of intact faces differs from the recognition of scrambled faces primarily in engaging holistic representations.

EXPERIMENT 2

One could argue that scrambled faces are too unnatural to provide an appropriate comparison for the processing of normal faces. Perhaps scrambled objects in general would be more likely to be represented featurally than normal objects. If so, one could not conclude from the previous experiment that face recognition is particularly holistic. For this reason, we turned to different contrasting stimulus set, inverted faces. Inversion disproportionately impairs the recognition of faces more than it does the recognition of other types of objects, such as airplanes, buildings, or costumes (Yin, 1969). These effects appear to be fairly robust, and results have been obtained for a variety of face stimuli including famous and novel faces (Scapinello & Yarmey, 1970; Yarmey, 1971), simple line drawn faces (Yin, 1969), photographs of faces (Carey & Diamond, 1977; Diamond & Carey, 1986) in different experimental paradigms, including forced-choice recognition (Yin, 1969) and "old" versus "new" judgments (Valentine & Bruce, 1986). The face inversion effect has been taken to index the operation of specialized face recognition mechanisms not normally used for recognizing other kinds of objects (e.g. Carey & Diamond, 1977; Yin, 1969). Thus, inverted faces provide a contrasting stimulus set that includes the same parts, in the same relative configuration, as normal faces but does not engage the hypothesized face-specific recognition mechanisms. In this experiment, subjects learned the face–name associations for six upright faces and six inverted faces. In test, subjects were asked to identify both the individual features of the learned upright or inverted faces presented in isolation and whole upright and inverted faces. If upright features are recognized using more holistic representations than inverted faces, then subjects should be more accurate at recognizing upright features contained in a whole face than in isolation, relative to whole face and isolated part recognition of inverted features.

Method

Subjects

Twenty first-year psychology students from Carnegie-Mellon University served as subjects in the experiment. Subjects were tested individually and received course credit for their participation.

Materials

The same two groups of face stimuli used in the previous experiment were used in this experiment. Two versions of each set were prepared: one in its normal upright orientation, and one inverted by 180°. Instead of presenting the stimuli on cards, they were presented on a Macintosh computer screen. The test items were presented in the same orientation as the study items.

Procedure

During the learning phase of the experiment, subjects learned the name–face associations for six upright (inverted) faces presented on a Macintosh computer. Faces and their assigned names were blocked according to face orientation. One learning block consisted of six learning trials, one trial per face, and there were five learning blocks in total. Learning was self-paced. Immediately following learning, a two-choice recognition test was administered. In contrast to Experiment 1, isolated part- and whole-face test items were randomly presented with the restriction that features from the same face were separated by at least two test trials, and the same feature type (e.g. nose feature) was not tested on consecutive trials. Also different from Experiment 1, the eyes, nose, and mouth features from each face were tested in the isolated part- and whole-face test conditions. Presentation of the test items was initiated by the subject, and test items were displayed until a response was made. Responses were recorded by computer. After the learning and test phases were completed for faces in one orientation (upright or inverted), the faces in the other orientation were learned and tested. Half of the subjects learned one group of six faces in the upright orientation and the other six faces in the inverted orientation. For the other half of the subjects, the face groups and their orientation was reversed. Learning and test phases were blocked according to face orientation, and presentation order of the face orientation was counterbalanced across subjects.

Results and Discussion

As shown in Figure 4, recognition of inverted, whole faces and inverted parts was roughly equivalent, 65% accuracy for whole face and 64% accuracy for parts, respectively. However, for upright faces, whole face stimuli were better recognized than part face stimuli. That is, subjects correctly identified 74% of the whole-face stimuli as compared to 65% of the part-face stimuli. The recognition advantage found for whole upright faces, but not for whole inverted faces, relative to the recognition for their parts, is consistent with the interpretation that holistic processing is used for upright
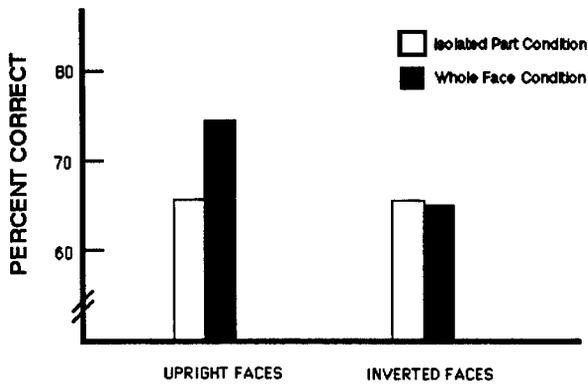
FIG. 4.   Percentage of correctly identified isolated part- and whole-face test items for upright and inverted faces.

faces. An ANOVA with face orientation (upright and inverted), test type (isolated part and whole face), face feature (eyes, nose, mouth) as within-subjects factors and order as a between-subjects factor confirmed the reliable Face Orientation × Test Type interaction, $F(1, 18) = 8.92, p < 0.01$. Consistent with the results of Experiment 1, the direct comparison between isolated part and whole face recognition for upright faces was again reliable, $t(19) = 3.41, p < 0.01$. Further, whereas there was little difference in performance between recognition of isolated parts of upright and inverted faces, whole upright faces were reliably better recognized than whole inverted faces, $t(19) = 2.94, p < 0.01$, again suggesting that normal upright faces are recognized more holistically relative to inverted faces.

The main effect of test type was also reliable, $F(1, 18) = 8.47, p < 0.01$, indicating that overall, whole-face stimuli—either whole upright faces or whole inverted faces—were better recognized than were part-face stimuli. Consistent with the results of Experiment 1, the main effect of face feature was reliable, $F(2, 38) = 8.47, p < 0.001$, such that eye features were better recognized (76% correct) than nose features (64% correct) or mouth features (63% correct). However the relative saliency of the face features was affected by the orientation of the face as indicated by the reliable Orientation × Face Feature interaction, $F(2, 38) = 3.88, p < 0.05$. No other main effects or interactions were reliable ($p > 0.10$).

In Experiment 2, we found that subjects were poorer at recognizing the parts of upright faces when presented in isolation than they were at recognizing the whole face, even though they showed no disadvantage for parts over wholes when the same faces were inverted. As in Experiment 1, the part disadvantage for upright faces was observed despite the fact that the

same discriminating information was available in both the part and whole test items for all types of stimuli: whichever pair of feature exemplars was presented in the forced-choice test of part identification, the corresponding pair of whole stimuli differed only by those features. In a related study, Young, Hellawell, and Hay (1987) found that inversion improved recognition of the top or bottom halves of composite faces, which they attributed to the disruption of configural processes. Taken together, these results suggest that the face representations affected by inversion are relatively holistic representations. Given that inversion is more disruptive to face recognition than to the recognition of other kinds of stimuli, this supports the hypothesis that face recognition involves more holistic representations than the recognition of other stimuli.

EXPERIMENT 3

In Experiments 1 and 2 it was found that intact, upright faces were encoded more holistically than scrambled faces or inverted faces. It is possible that these results do not reflect anything special about face recognition per se, but only demonstrate holistic processing for the recognition of coherent, upright objects. The purpose of the present experiment is to contrast face recognition with the recognition of normal upright stimuli other than faces. Houses have been used as the contrast stimuli to faces in other studies (Bruce et al., 1991; Valentine & Bruce, 1986; Yin, 1969) and seemed particularly suited to goals of our research for several reasons. Like faces, houses have internal features (i.e. doors and windows) that share an overall configuration. Also like faces, the parts of a house can be varied independently of each other without disrupting the house schema. Finally, house stimuli can be constructed such that the number of house features corresponds to the number of face features. As shown in Figure 5, the house stimuli used in Experiment 3 had three features—a door, a large window, and two small windows—analogous to the mouth, nose, and eyes of a face. If holistic processing is not restricted to faces, then a disadvantage should also be evident for house parts relative to whole-house stimuli. On the other hand, if the use of holistic representations is a particular characteristic of face processing, houses should not show the relative part disadvantage.

Method

Subjects

Twenty first-year psychology students from Carnegie-Mellon University served as subjects in the experiment. Subjects were tested individually and received course credit for their participation.

Materials

House stimuli were generated on a Macintosh computer using an architectural design software package. As shown in Figure 5, similar to the faces, houses were constructed by selecting one of the three feature values for each of the three feature types (e.g. door, big window, small window). The six stimulus houses were created according to the exemplars specified by the face stimuli.

Procedure

The procedure was similar to the one used in Experiment 2. Subjects were informed that they would see a house (face) picture accompanied by a name, and their task was to learn the name–picture association. In the case of the houses, subjects were told that the name corresponded to the
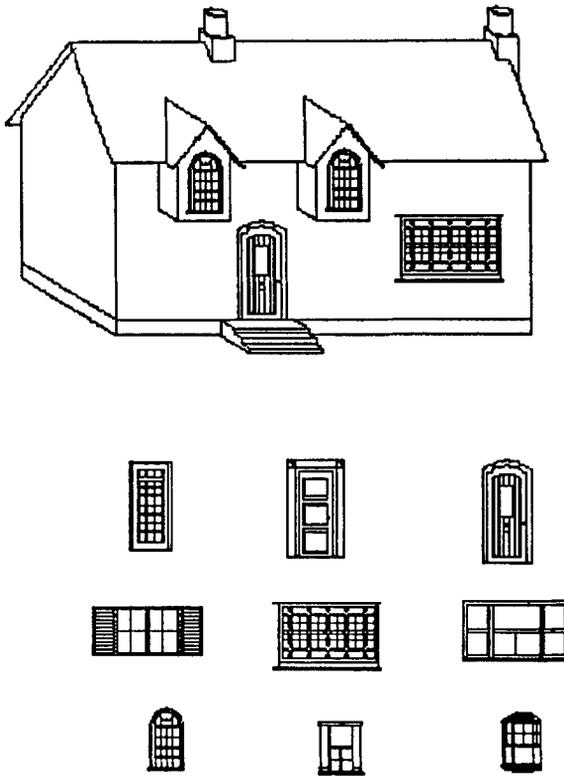

FIG. 5.   Door, large window, small window exemplars and a sample stimulus house used in Experiment 3.

person who lived in the house. A learning block consisted of six learning trials, one trial per house (face) picture, and pictures were randomly presented on a Macintosh computer. There were five learning blocks per object type. After the learning phase was completed, recognition memory for the part shown in isolation and embedded in the whole object was randomly tested in a forced-choice paradigm. The same item order restrictions described in Experiment 2 were used. Whole-object foils (house and face) were constructed such that they were distinct from any previously learned object. Recognition memory was tested for the three house features (i.e. door, small window, big window) and three face features (eyes, nose, mouth) presented in isolation and in the whole-object conditions. Learning and test were blocked according to object type (house and face). The order of the object type presented for learning and test was counterbalanced across subjects.

Results and Discussion

As shown in Figure 6, whereas only 65% of the face features were recognized in isolation, recognition improved to 77% when the same features were shown in the whole-face context. This finding replicates the holistic effect found for faces demonstrated in the previous two experiments. In contrast, recognition of the house features was roughly equivalent in the isolated and whole-house test condition, 81% and 79% correct, respectively. Thus, unlike faces, no advantage was found for identifying house features as part of their whole object. An ANOVA with object type (houses and faces) and test type (isolated part and whole face) as within-subjects factors and
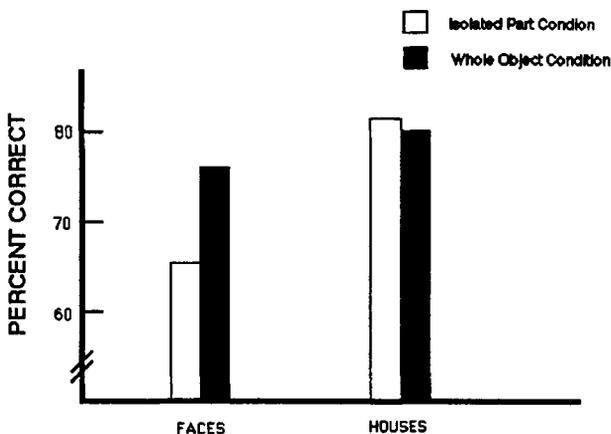

FIG. 6.   Percentage of correctly identified isolated part- and whole-object test items for faces and houses.

order as a between-subjects factor revealed a reliable Object Type × Test Type interaction, $F(1, 18) = 17.47$, $p < 0.001$, as predicted. A direct comparison also showed that facial features were more readily recognized in the whole-face condition than in the isolated part condition, $t(19) = 4.46$, $p < 0.01$. The main factor of object type was also reliable, $F(1, 18) = 9.20$, $p < 0.01$, indicating that houses were recognized more accurately than were faces. A reliable effect was also found for test type, $F(1, 18) = 9.11$, $p < 0.01$; however, this effect should be interpreted as the result of its higher-order interaction with object type. Finally, the effect of order was also reliable, $F(1, 18) = 4.41$, $p < 0.05$, but order did not interact with any other factor. No other interactions were reliable, $p > 0.10$.

In comparing recognition for different types of objects, it is difficult to equate the relative discriminability of features—in this case, face features and house features. However, the focus of the present study was not on comparing part recognition across object types, only in comparing parts and wholes recognition within an object type. In this regard, we found an advantage for the recognition of the wholes of faces relative to the isolated face part, but found no difference between part and whole recognition for houses. Furthermore, the possibility that a difference in part discrimination across object types is, in some indirect way, responsible for a difference in reliance in part versus whole recognition cannot explain the results of Experiments 1 and 2 in which the part features were the same. Thus, the main finding of Experiment 3 is consistent with the claim that face recognition is different from the recognition of other objects, such as houses, in its relatively greater reliance on holistic representations.

GENERAL DISCUSSION

In these experiments we tested the hypothesis that face recognition is relatively more dependent on holistic representations than the recognition of other types of stimuli. By *holistic representation* we mean one without an internal part structure. Following other researchers, we reasoned that if a portion of an object corresponds to an explicitly represented part in a hierarchical visual representation, then when that portion is presented in isolation it will be identified relatively more easily than if it did not have the status of an explicitly represented part. The hypothesis that face recognition is holistic therefore predicts that the isolated parts of a face will be disproportionately more difficult to recognize than the whole face, relative to recognition of the parts and wholes of other kinds of stimuli. This prediction was borne out in three experiments: subjects were less accurate at identifying the parts of faces, presented in isolation, than they were at identifying whole faces, even though both parts and wholes were tested in a forced-choice format and the whole faces differed only by one part. In

contrast, three other types of stimuli—scrambled faces, inverted faces, and houses—did not show this disadvantage for part identification.

At first glance, these results are reminiscent of the *face superiority effect*, according to which the parts of a face are better perceived if presented in the context of a whole face than in the context of a scrambled face (e.g. Homa, Haver, & Schwartz, 1976; Mermelstein, Banks, & Prinzmetal, 1979). The two phenomena are indeed similar in that both reflect the influence of representations of wholes on subjects' performance. However, they are distinct phenomena, differing from each other in several ways. (1) The face superiority effect comes into play only under conditions of threshold vision, suggesting that its locus is in the visual encoding of facial features, not their access to stored memory representations. In contrast, our task did not tax visual encoding, but taxed memory access. (2) As Pomerantz (1981) has noted, in face and object superiority effects the perception of a part in context is as good as, but not better than, recognition of just the isolated part. Performance with the whole face is superior only to performance with a scrambled face. In contrast, we found that recognition of whole faces was better than recognition of isolated parts. (3) The face superiority effect does not appear to be specific to faces but is a more general phenomenon involving the visual encoding of parts in context, alongside the word superiority effect (Reicher, 1969; Wheeler, 1970) and object superiority effects for geometric forms (Enns & Gilani, 1988; Weisstein & Harris, 1974), and chairs (Davidoff & Donnelly, 1990). In contrast, the present results with faces were not found with the other types of tested stimuli.

How do these findings relate to the idea that face recognition is particularly dependent on "configuration"? If by a configurational representation we mean one in which the spatial relations among the parts of a face are as important as the shapes of the individual parts themselves (Haig, 1984; Hosie, Ellis, & Haig, 1988), then we would suggest that the concepts of configurational representation and holistic representation are highly similar, and possibly identical. The shapes of the individual parts are essentially within-part spatial relations. In the limiting case of configurational representation, in which between-part spatial relations are as precisely specified as the within-part relations, parts have lost much, if not all, of their special status. Presumably, for this reason, the terms *holistic* and *configurational* have often been used interchangeably in the face recognition literature.

Recent findings in neurophysiology and neuropsychology seem consistent with our conclusions regarding the relatively holistic representation of faces. It has been demonstrated that a subpopulation of neurons located in the superior temporal sulcus of the monkey responds selectively to the sight of face parts and whole faces (e.g. Desimone, Albright, Gross, &

Bruce, 1984; Perrett, Mistlin, & Chitty, 1987) and display some ability to discriminate among different faces (Baylis, Rolls, & Leonard, 1985). Although the responses of these neurons to a face are not greatly diminished by deleting a feature, they are abolished if all features are present but scrambled (Desimone et al., 1984), consistent with holistic rather than featural representation. Although many interpretations of this fact of anatomy are possible, it is at least consistent with the notion that face representations are relatively holistic.

The fact that the temporal cortex of monkeys also contains cells responsive to individual facial features, especially eyes, has been taken by some to indicate that faces are represented hierarchically, with explicitly represented component parts (Perrett et al., 1987). However, Desimone (1991) has raised the possibility that the "feature" cells may not be representing facial features per se. For example, a cell that responds to an eye in isolation might respond to any dark spot on a white background. Furthermore, it appears that the functional role of many of the "eye" cells may be to represent direction of eye gaze, an important form of social interaction among monkeys (Perrett et al., 1985). In our view, a critical test of the hierarchy hypothesis for interpreting the role of "feature" cells in face processing would be to verify that their latencies of response are, on average, shorter than the latencies of "face" cells. This test has not yet been carried out (Perrett, personal communication).

Human neuropsychology is also consistent with the hypothesis of relatively holistic face recognition. Brain-damaged patients may be impaired at face recognition, object recognition, or printed word recognition. In analysing the patterns of co-occurrence among these impairments, Farah (1991) found that two possible combinations of these impairments did not occur: object recognition impairments without either face or word impairments, and both face and word impairments without some degree of object impairment. This suggested the existence of two, rather than three, underlying representational capacities responsible for the recognition of faces, objects, and words, which are used in complementary ways: one that is essential for face recognition, needed to a lesser extent for the recognition of common objects and not needed at all for printed word recognition, and one that is essential for printed word recognition, needed to a lesser extent for the recognition of common objects, and not needed at all for face recognition. The representational capacity lacking in patients with impairments in printed word recognition appears to be the ability to represent multiple explicitly represented structural units (e.g. letters in a word; see Farah & Wallace, 1991, for a review of the evidence). The ability to represent shape holistically would seem a good candidate for a complementary representational capacity, and the neuropsychological evidence suggests that this capacity is particularly taxed by face recognition.

REFERENCES

Baylis, G.C., Rolls, E.T., & Leonard, C.M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex of the superior temporal sulcus of the monkey. *Brain Research, 342,* 91–102.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Bower, G.H., & Glass, A.L. (1976). Structural units and the redintegrative power of picture fragments. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 456–466.

Bradshaw, J.L., & Wallace, G. (1971). Models for the processing and identification of faces. *Perception and Psychophysics, 9,* 443–448.

Bruce, V. (1988). *Recognizing faces.* Hove: Lawrence Erlbaum Associates Ltd.

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition, 38,* 109–144.

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

Davidoff, J., & Donnelly, N. (1990). Object superiority: A comparison of complete and part probes. *Acta Psychologica, 73,* 225–243.

Desimone, R. (1991). Face selective cells in temporal cortex of monkeys. *Journal of Cognitive Neuroscience, 3,* 1–8.

Desimone, R., Albright, T.D., Gross, C.G., & Bruce, C.J. (1984). Stimulus-selective responses of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4,* 2051–2068.

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

Enns, J.T., & Gilani, A.B. (1988). Three-dimensionality and discriminability in the object-superiority effect. *Perception and Psychophysics, 44,* 243–256.

Farah, M.J. (1991). Patterns of co-occurrence among the associative agnosias: Implications for visual object representation. *Cognitive Neuropsychology, 8,* 1–19.

Farah, M.J., & Wallace, M.A. (1991). Pure alexia as a visual impairment: A reconsideration. *Cognitive Neuropsychology, 8,* 313–334.

Galton, F. (1879). Composite portraits, made by combining those of many different persons into a single, resultant figure. *Journal of the Anthropological Institute, 8,* 132–144.

Haig, N.D. (1984). The effect of feature displacement on face recognition. *Perception, 13,* 104–109.

Hoffman, D.D., & Richards, W.A. (1984). Parts of recognition. In S. Pinker (Ed.), *Visual Cognition.* Cambridge, MA: MIT Press.

Homa, D., Haver, B., & Schwartz, T. (1976). Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Memory and Cognition, 4,* 176–185.

Hosie, J.A., Ellis, H.D., & Haig, N.D. (1988). The effect of feature displacement on the perception of well-known faces. *Perception, 17,* 461–474.

Matthews, M.L. (1978). Discrimination of Identikit construction of faces: Evidence for a dual processing strategy. *Perception and Psychophysics, 23,* 153–161.

Mermelstein, R., Banks, W., & Prinzmetal, W. (1979). Figural goodness effects in perception and memory. *Perception and Psychophysics, 26,* 472–480.

Palmer, S.E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9,* 441–474.

Perrett, D.I., Mistlin, A.J., & Chitty, A.J. (1987). Visual neurones responsive to faces. *Trends in Neuroscience, 10,* 358–364.

Perrett, D.I., Smith, P.A.J., Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., & Jeeves, M.A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London, Series B, 223,* 293–317.

Pomerantz, J.R. (1981). Perceptual organization in information processing. In M. Kubovy & J.R. Pomerantz (Eds.), *Perceptual organization* (pp. 141–180). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Reed, S.K. (1974). Structural descriptions and the limitations of visual images. *Memory & Cognition, 2,* 329–336.

Reicher, G.M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81,* 275–280.

Scapinello, K.F., & Yarmey, A.D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science, 21,* 329–330.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *The British Journal of Psychology, 75,* 221–242.

Smith, E.E., & Nielsen, G.D. (1970). Representations and retrieval processes in short-term memory: Recognition and recall of faces. *Journal of Experimental Psychology, 85,* 397–405.

Valentine, T., & Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica, 61,* 259–273.

Walker-Smith, G.J. (1978). The effects of delay and exposure duration in a face recognition task. *Perception, 6,* 63–70.

Weisstein, N., & Harris, C.S. (1974). Visual detection of line segments: An object-superiority effect. *Science, 186,* 752–755.

Wheeler, D.D. (1970). Process in word identification. *Cognitive Psychology, 1,* 59–85.

Yarmey, A.D. (1971). Recognition memory for familiar "public" faces: Effects of orientation and delay. *Psychonomic Science, 24,* 286–288.

Yim, R.K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141–145.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configuration information in face of perception. *Perception, 16,* 747–759.