Journal of Experimental Psychology:
Human Perception and Performance
2001, Vol. 27, No. 3, 573–599

0096-1523/01/$5.00 DOI: 10.1037//0096-1523.27.3.573

# Categorical Perception of Face Identity in Noise Isolates Configural Processing

Elinor McKone, Paolo Martini, and Ken Nakayama
Harvard University

Neuropsychological evidence suggests that face recognition based on configural (holistic) information can occur in isolation from recognition based on local feature cues. The present study shows that configural processing can be isolated experimentally in normal subjects. A phenomenon is reported that exists only for upright whole faces, namely categorical perception (CP) of face identity in noise. Three discrimination tasks (ABX, better likeness, and similarity ratings) were used to test for perceptual distortion across the category boundary predicted from binary classification of face morphs. Noise was added such that any single local region provided unreliable cues to identity. Under these conditions, CP was found for upright faces but not for inverted faces or single features, even with more than 10,000 trials. The CP-in-noise signature phenomenon was then used to show that configural processing survives image plane rotations of 45°–90°.

It has been argued that identification of a known face can occur through either one of two mechanisms (Diamond & Carey, 1986; Rhodes, 1988; Sergent, 1984; Tanaka & Farah, 1993) or through some combination of both (Bartlett & Searcy, 1993; Moscovitch, Winocur, & Behrmann, 1997; Rhodes, Brake, & Atkinson, 1993). In what has been referred to as *feature-based* or *part-based identification,* the recognition of a known face occurs through information from local regions of the face, such as nose shape, eye color, or hairstyle. *Configuration-based* or *holistic identification,* in contrast, relies on information integrated over the entire internal face region, which is matched to a stored representation coding the expected form of an upright face.

There is evidence that each of these mechanisms can contribute to face recognition in both everyday and experimental settings. Anecdotally, it is clear that changes in hairstyle or the shaving off of a beard can lead to initial failures to identify even highly familiar people. A more formal demonstration of a role for feature-based identification can be found in the illusion that U.S. President Bill Clinton's face, when shown with Vice-President Al Gore's hair, can produce an initial percept of Gore (Sinha & Poggio, 1996). In experimental situations, feature-based identification can also contribute to performance; for example, old–new recognition memory judgments can potentially be based on recognizing an unusual hairstyle or nose shape from a studied set of face pictures. That this does indeed occur is suggested by the fact that prosopagnosics can show surprisingly good memory for pictures of whole heads, despite very poor memory for pictures showing only the internal region of the face (Postma, 1998).

Although these cases argue for at least some identification (or misidentification) based on single features, it can be noted that single features provide very unreliable information when an individual must be distinguished from among the full set of possible faces to which people are exposed. Faces form a very homogeneous class of stimuli in which a single feature, or even set of features, is far from unique (e.g., there are many men with beards and people with blue eyes). Furthermore, the appearance of an individual changes substantially with viewing angle, lighting conditions, expression, cosmetics, hairstyle, facial hair, age, and so on. These two facts argue that local features provide insufficient information to ensure the accurate discrimination of identity and, thus, that configural processing of the face is also necessary (Bradshaw & Wallace, 1971; Diamond & Carey, 1986; Rhodes, 1988; Sergent, 1984; Tanaka & Farah, 1993).

Empirical evidence for configural processing comes from the detrimental effects of manipulations that disrupt the holistic structure of the face but leave individual features intact. As examples, such manipulations include scrambling the face (Tanaka & Farah, 1993), "exploding" the face into separated face parts (Farah, Tanaka, & Drain, 1995), misaligning the lower and upper halves of the face (Moscovitch et al., 1997), and splitting the face into strips that do not form a cohesive surface when presented stereoscopically (Nakayama, Shimojo, & Silverman, 1989).

Further evidence demonstrates a role not just for holistic processing but for a particular normative global structure (Rhodes, Brennan, & Carey, 1987; Valentine, 1991), namely that corresponding to an upright face. When faces are inverted, detrimental effects on performance are observed with both memory and perceptual tasks, as illustrated in the following examples. Old–new recognition is worse for faces presented upside down than for upright faces (Yin, 1969). Reaction times in a famous–nonfamous decision task are slower for inverted faces than for upright faces (Valentine & Bruce, 1988). In visual search, in comparison with

---

Elinor McKone, Paolo Martini, and Ken Nakayama, Department of Psychology, Harvard University.

This research was supported in part by grants from the U.S. Air Force Office of Scientific Research, the McKnight Foundation, and the Human Frontiers Science Program. We thank Mary Peterson, Jim Tanaka, and two anonymous reviewers for valuable comments on previous versions of this article.

Correspondence concerning this article should be addressed to Elinor McKone, who is now at the Division of Psychology, Australian National University, Canberra ACT 0200, Australia. Electronic mail may be sent to Elinor.McKone@anu.edu.au.

upright presentation, it is harder both to locate a specific inverted target face and to reject an inverted distractor (Tong & Nakayama, 1999). In naming chimeras comprising the top half of one famous face and the bottom half of another, performance is reduced by alignment of the halves with upright presentation (suggesting the two halves integrate to form a single face) but is less affected with inverted presentation (Young, Hellawell, & Hay, 1987; see also Carey & Diamond, 1994). Finally, the Thatcher illusion, in which flipping the eyes and mouth relative to the rest of the face produces a grotesque appearance, occurs most strongly when the face itself is upright (Bartlett & Searcy, 1993; Thompson, 1980). In contrast to the ubiquitous effect of inversion on whole face processing, a number of studies have shown no effects of inversion on the processing of isolated features (Bruyer & Coget, 1987; Farah, Tanaka, & Drain, 1995; Rhodes et al., 1993). For example, Farah, Tanaka, and Drain (1995) showed that the inversion effect on recognition memory does not occur when the faces are presented as "exploded" parts in the study phase, supporting the view that inversion selectively disrupts configural processing of whole faces.

All of these results indicate that face processing proceeds at least partly through a configural mechanism. Furthermore, this mechanism appears to have two basic properties: First, it extracts information from the whole face, and, second, it codes the expected structure of an upright face and is unavailable for inverted faces. Given evidence such as that just reviewed, almost all researchers agree on these two general aspects of configural face processing. Views differ, however, regarding the more specific properties of configural representations (for a review, see Farah, Wilson, Drain, & Tanaka, 1998). For example, different researchers have proposed that configural processing might rely on information about the spatial relations between features (Rhodes, 1988; Sergent, 1984) or the spatial relations of the features relative to the prototypical arrangement (Carey & Diamond, 1994; Diamond & Carey, 1986; Rhodes et al., 1987), might result from better feature processing in the context of a learned face (Tanaka & Sengco, 1997), or might reflect a shape coding of the whole face that, although including all details of the face, is not constructed by reference to feature-based component parts (Hancock, Burton, & Bruce, 1996; O'Toole, Deffenbacher, Valetin, & Abdi, 1994; Tanaka & Farah, 1993).

## Isolation of Configural Processing

We have argued that it is configural information that is essential to solving the full range of face identification problems encountered in the real world. Farah and coworkers (e.g., Farah et al., 1998) and Moscovitch et al. (1997) went further, arguing that there exists a face-specific recognition system whose purpose is to process holistic information from the internal regions of an upright face, whereas part-based processing, such as that seen for inverted faces, external features (e.g., hairstyle), or features presented alone, arises from a more general purpose object recognition system. Whether or not the more extreme of these views is accepted, both imply that the most important aspect of face processing is the configural aspect. This makes it of considerable interest to be able to investigate the properties of configuration-based identification *in isolation from any contribution of a feature-based* mechanism.

The focus of most studies to date, however, has been on the opposite situation, namely the isolation of feature-based processing from any influences of configuration. In experiments with normal subjects, the standard procedure is to disrupt configural processing (e.g., by inverting or exploding the stimulus) while leaving feature-based processing intact. Moscovitch et al. (1997) pointed out that this standard approach provides only an indirect method of investigating the nature of configural processing. This is because it relies on interpreting what *cannot* be done *without* a configural mechanism rather than directly studying what is done by such a mechanism. These authors also noted that the focus on prosopagnosia as the relevant deficit in neuropsychological studies follows the same logic, in that prosopagnosics appear to have damaged processing of configural information but intact processing of features (e.g., no advantage for upright whole faces over either inverted faces [Farah, Wilson, Drain, & Tanaka, 1995] or face parts [Farah, 1996]).

That it might be possible for configural face processing to operate in isolation from feature processing was first suggested by Moscovitch et al. (1997). These authors described a brain-injured patient, C.K., with exactly the constellation of symptoms expected from intact configural processing of faces but damaged part-based processing of features and objects. When tested with upright whole faces, C.K. performed at normal or above-normal levels on a wide variety of identification tasks. This held whatever the form of presentation, including real people, photographs showing the entire head, photographs showing internal face regions only, caricatures, cartoon faces, high-contrast "Mooney" faces, faces drawn in overlapping spatial locations, faces hidden as contours of other objects in a scene, and "Arcimboldo" faces made up of nonface object parts. At the same time, C.K. showed extremely poor performance when face configuration was disrupted by inversion or by showing the face as separated strips. His object recognition was also extremely poor (see also Behrman, Winocur, & Moscovitch, 1994), even for the overt components of scenes in which he was able to identify the hidden faces and for the component objects of the Arcimboldo faces. Given this pattern of preserved and damaged abilities, patient C.K. would appear to demonstrate an organic isolation of configural face processing.[1]

Having identified an isolation of configural face processing in patient C.K., Moscovitch et al. (1997) went on to demonstrate how this could be used to directly explore some of the properties of configural face representations. They argued, for example, that only a vertical half face is necessary to activate configural face processing, because C.K. could identify famous faces split down the middle vertically but not faces split horizontally. They also argued that configural representations contain full details of the face that can be used to fill in missing parts of an image, because C.K. could identify famous faces with a single part missing (i.e., eyes, nose, or mouth region) as accurately as could controls, and he could select the correct missing part from two alternatives (i.e., the

---

[1] At least one other patient has been reported with intact recognition of upright whole faces but poor recognition of objects (Rumiati, Humphreys, Riddoch, & Bateman, 1994). It is possible that this case also represents an isolation of configural face processing. However, the patient was not tested in nearly as much detail as C.K.; in particular, he was not tested with disrupted face configurations (inverted, strips, etc.).

correct mouth), but only for faces he correctly identified (controls showed the same pattern).

To date, there have been no studies reporting isolation of configural processing in normal subjects. In attempting to demonstrate such isolation, the obvious starting point is a comparison of performance between upright and inverted faces, given (a) the evidence that inverted faces do not access configural representations and (b) the fact that inversion leaves unchanged all low-level visual properties of the stimulus (e.g., contrast and spatial frequency). The types of inversion effects reported in the previous literature, however, are insufficient to demonstrate an isolation of configural processing, because these effects are partial rather than complete. In most studies, performance with inverted faces remains well above floor levels, indicating a residual ability to perceive and remember inverted faces (e.g., inverted faces are named less rapidly than upright faces, but most can still be named; visual search for inverted faces is more difficult than that for upright faces, but inverted faces can still be found; and memory for inverted faces is worse than that for upright faces, but it is still above chance levels).

Findings such as those of Farah, Tanaka, and Drain (1995); Young et al. (1987); and Moscovitch et al. (1997) argue that this residual ability to process inverted faces relies on featural information. Because featural information is also present in *upright* faces, we would argue that experimental performance with upright faces commonly relies on some mixture of feature-based and configuration-based processing (see also Bartlett & Searcy, 1993; Moscovitch et al., 1997; Rhodes et al., 1993). Furthermore, it is not known how featural processing and configural processing might interact to affect performance for upright faces. Therefore, simply subtracting inverted performance from upright performance does not necessarily leave a component of processing that can be attributed to configural information alone. Ideally, isolation of configural processing requires a qualitative difference between upright and inverted performance, whereby some signature phenomenon exists *only* for upright faces. The nonexistence of this phenomenon for inverted faces, as well as for single features presented alone, would argue that no contribution of feature-based processing to the signature phenomenon has occurred. It is only under these circumstances that any presence of the phenomenon (e.g., in upright faces) could then be attributed purely to configural processing.

In the present article, we demonstrate a phenomenon that exists for upright whole faces but not for inverted faces or single isolated features. This argues that configural processing can operate in isolation from feature-based identification, not only in patients with brain injuries, but in normal subjects as well. Furthermore, it provides an experimental technique that, for the first time in normal subjects, allows configural face processing to be studied directly rather than indirectly. Thus, our work has the potential to lead to a clearer specification of the nature and content of configural face representations beyond the general properties of such representations as "holistic" and "tuned to the form of an upright face."

There were two aspects of our approach to isolating configural processing, which are described in detail in the sections to follow. These were (a) selecting a task that required making fine discriminations of identity between similar faces and (b) choosing stimuli in which the information from any single local region of the image provided an unreliable indicator of the identity of the face. Under these conditions, we hoped that integration of information across large regions of the face (i.e., configural information) would be necessary to produce reliable performance on the fine discrimination task.

## A Fine Discrimination Task: Categorical Perception of Face Identity

The phenomenon through which we were able to isolate configural processing was *categorical perception* (CP) of face identity. CP refers to the perceptual distortion of a continuous physical stimulus (e.g., wavelengths of visible light) into discrete categories with sharp boundaries between them (e.g., color bands of the rainbow). It has been demonstrated for categories that many researchers assume to be innate or to emerge early in development, including color (Bornstein, 1987), phoneme boundaries (Eimas & Corbit, 1973; Liberman, Harris, Hoffman, & Griffith, 1957), and facial expression (Etcoff & Magee, 1992). Importantly for the present work, CP has also been demonstrated for a wide range of categories that are clearly learned. These include musical intervals (Burns & Ward, 1978), size–brightness and brightness–saturation combinations (Goldstone, 1994), textures (Pevtzow & Harnad, 1997), and complex visual shapes (e.g., line drawings of microorganismlike objects and photographs of chick cloaca as male or female; Livingston, Andrews, & Harnad, 1998). Indeed, we would presume (see Harnad, 1987b) that the perceptual distortion of the physical world seen in CP reflects a general mechanism essential to discriminating similar stimuli with different identities.

Recently, several studies have investigated learned CP for face identity. In all of these studies, clear photographs showing the full head (including hair) were used. Beale and Keil (1995) demonstrated CP for famous faces using a morphing procedure to produce intermediate images between a pair of endpoint faces (e.g., U.S. presidents Kennedy and Clinton). Subjects first performed a binary classification task (Kennedy vs. Clinton) to determine the predicted category boundary (i.e., the morph producing Clinton responses on 50% of occasions). CP was then demonstrated by showing better discrimination between pairs of stimuli that crossed the category boundary (e.g., the 35% Clinton morph vs. the 55% Clinton morph, for a category boundary at the 45% morph) than for equidistant pairs falling on the same side of the boundary (e.g., the 20% morph vs. the 40% morph). Beale and Keil used two different tasks to assess pair discrimination. The first was the *ABX task:* Two morphs differing by 20% of the continuum were presented as A and B, a third stimulus X was then presented, and the subject reported whether X matched A or B. The second was a *better likeness task:* Here, subjects reported which of two morphs differing by 20% of the continuum was more like a specified endpoint face (e.g., which was more like Clinton). In both cases, results showed better discrimination accuracy for pairs that crossed the predicted category boundary than for equidistant pairs drawn from the same side of the boundary, thus demonstrating CP.

In a second study, Stevenage (1998) showed CP for highly similar, initially novel faces. She trained subjects to distinguish between identical twins using different photographs of each twin. Before training, subjects rated pairs of photographs of different twins and pairs of photographs of the same twin as equally similar.

After training, however, different-twin pairs were rated as less similar than same-twin pairs, again demonstrating CP.

Neither Beale and Keil (1995) nor Stevenage (1998) investigated CP for inverted faces. Levin and Beale (2000) have recently done so, using morphs between initially novel faces and the better likeness task. They obtained CP for inverted faces differing in identity (see also Angeli & Gerbino, 1998). However, given that CP occurs for many visual stimuli that have nothing to do with faces (as described earlier), note that it is not necessary to assign the source of this CP for inverted faces to any face-like structure of the images. Instead, it is quite possible that CP for inverted faces relied on CP for nonconfigural aspects of the stimulus, including extraface information (e.g., hairstyle or presence vs. absence of moustache), single-feature information (e.g., eye shape or color), and information from very localized regions of the image (e.g., angle of the junction between eyelid and iris or presence or absence of a wrinkle). Note that Levin and Beale's stimuli almost certainly contained enough nonconfigural information to support CP, because the endpoint faces differed substantially in hairstyle, age, and, sometimes presence of a mustache.

Although Levin and Beale (2000) obtained CP for inverted faces, the effect was significantly weaker than that for upright faces. That is, CP for face identity demonstrated the standard face inversion effect. As with the other face inversion effects in the literature, however, this effect was partial rather than complete. Thus, we presume that CP for upright whole-head images could be based on some combination of CP for local features and CP for configuration and so does not, in itself, provide a suitable method for isolating configural processing. Isolation of configural processing would require that no CP effect occurs for inverted faces. In the present work, we show that this pattern does in fact emerge when highly similar faces are used and the stimuli are degraded by noise.

## Choice of Stimuli: Unreliable Local Cues to Identity

Figure 1 shows the three pairs of endpoint faces we used to test for categorical perception of identity. Our stimuli were chosen to exclude from the image, to the extent possible, any nonconfigural cues that might be used to support CP. First, we used faces that had no obviously distinguishing features (no facial hair, no spectacles, and no distinguishing marks) and presented them without hair. Second, the endpoint faces in each pair were very similar to each other (same sex and age, with facial features falling in overlapping spatial locations).

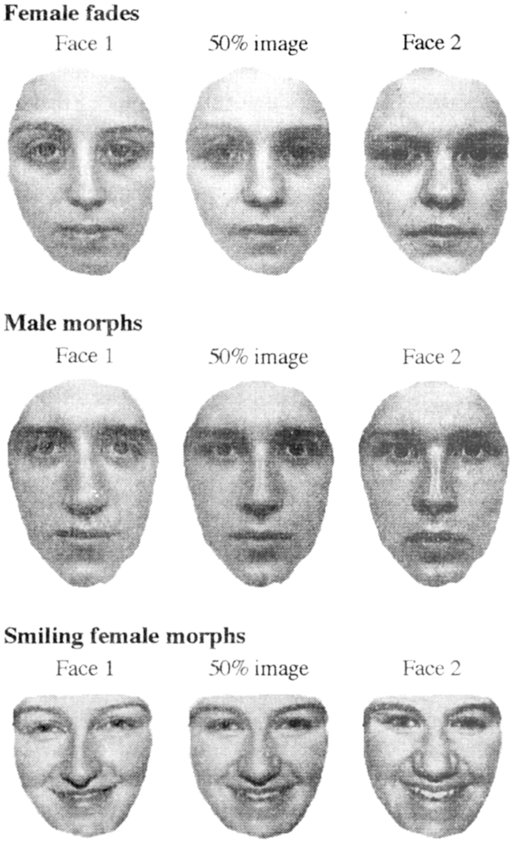


*Figure 1.* Endpoint stimuli for each continuum, along with the 50% image created by the fading or morphing procedures. Smiling female morphs Face 1 and Face 2 from *The Psychological Image Collection at Stirling (PICS)* [Electronic database], by P. Hancock, available on the World Wide Web: http://pics.psych.stir.ac.uk/. Copyright 1995 by Roger Watt. Reprinted with permission. All subjects gave consent for their photographs to be used for scientific studies.

Finally, we added noise to the stimuli. We have argued that the purpose of configural processing is to allow an individual to be identified despite substantial changes in the image due to lighting conditions, viewpoint, expression, makeup, and so on. Our rationale for adding noise was to mimic this unreliability in the image. The types and levels of noise were therefore selected to meet a criterion that different random noise assignments should create noticeable changes in the apparent shape of local contours. Figure 2 demonstrates the effect of such noise on a single isolated feature, specifically the nose region of the "male morph" continuum. As can be seen, different random noise assignments altered the apparent shape of the nostrils and potentially led to an unreliable positioning of intermediate nose morphs (e.g., the 40% and 60% morphs) relative to a category boundary (e.g., the 50% morph). Thus, with the addition of noise to morphs between similar endpoint stimuli, any one local region of the image provided only unreliable information about the identity of the face. Under these circumstances, our idea was that the fine discrimination necessary to allow CP (e.g., reliable differentiation of identity between the 40% and 60% morphs) would require information integrated from across large areas of the face. Thus, we hoped that CP might rely entirely on configural face processing. (We discuss the exact role of noise in more detail in the General Discussion section.)

## The Present Study

In this article, we present three experiments that demonstrate and explore the isolation of configural face processing in normal subjects. We achieve the isolation of configural processing by minimizing local cues to identity in the stimuli and by using categorical perception of identity as the performance measure. Experiment 1 showed CP for upright but not inverted faces at high levels of noise; three different discrimination tasks were used to assess CP. To demonstrate that the lack of CP for inverted faces represents a qualitative difference between upright and inverted processing and not simply slower learning of CP for inverted faces,

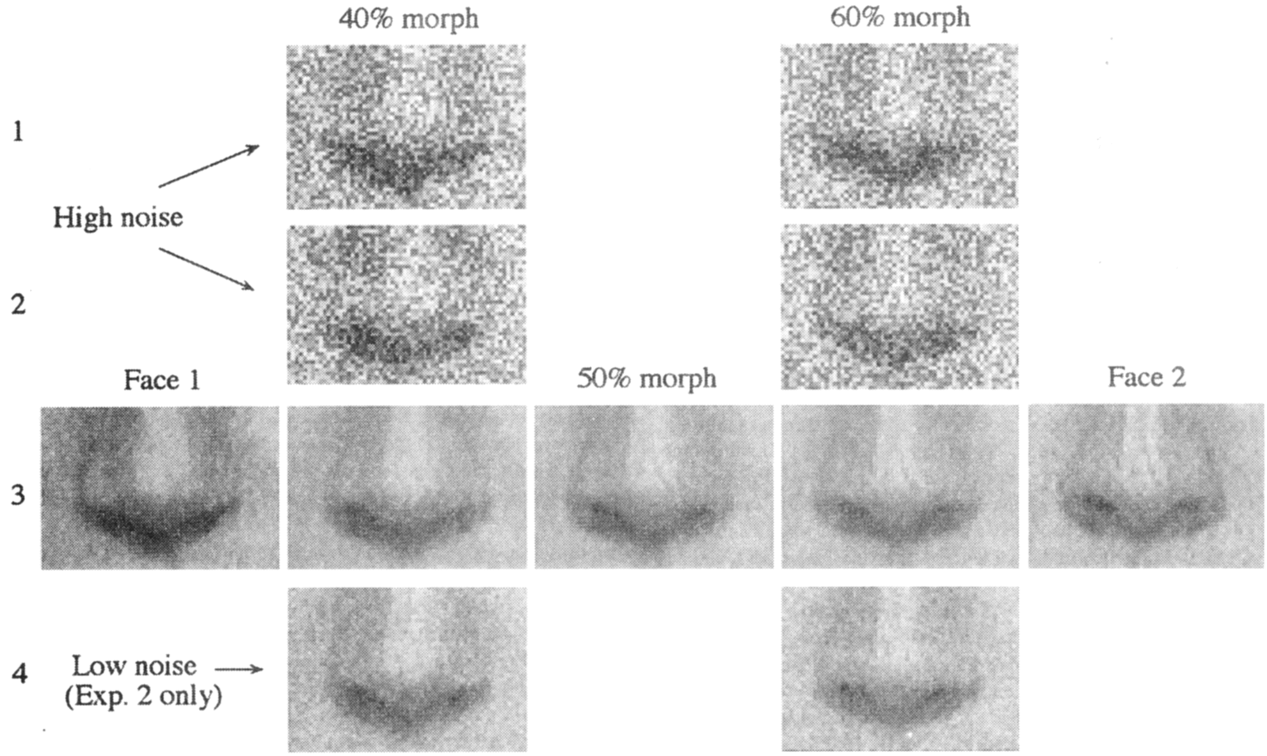


*Figure 2.* Nose region of the male morph continuum, without noise, for the 0%, 40%, 50%, 60%, and 100% morphs (row 3), and the 40% and 60% morphs, each with two different random "high" noise assignments (rows 1 and 2). In row 1, the noise assignments happen to produce nose contours similar to those in the corresponding endpoint images: It is clear that the 40% morph lies on the "Face 1" side of a putative boundary at the 50% morph and that the 60% morph lies on the "Face 2" side. However, most noise assignments produce images more like those in row 2, in which it is difficult to determine where the two images fall with respect to a 50% morph boundary. Thus, the high noise level results in unreliability of the position of the intermediate morphs relative to the identity boundary. Row 4 shows that the low noise level (Experiment 2 only) did not have this effect. Figure should be viewed at arm's length. Exp. = experiment.

we provided subjects with many thousands of exposures to stimuli from the target face continuum. By the end of testing, the level of practice had reached 10,000–30,000 trials with the target continuum spread over several months, for 4 of the 5 subjects tested. In Experiment 2, we examined CP for a single feature versus the whole face at high and low levels of noise. Results confirmed that adding the noise is essential to isolate configural processing: When the noise level is reduced, CP is demonstrated for inverted whole faces as well as for a single feature presented alone. Experiment 3 demonstrates how CP in noise can be used to extend our knowledge of the exact nature of configural processing, in this case by examining its sensitivity to degree of misorientation in the image plane.

## Experiment 1: Inversion Effects on CP for Faces in Noise

In Experiment 1, we examined categorical perception for whole faces presented in heavy noise. The primary aim was to demonstrate that, whereas CP survives noise for upright faces, no CP occurs for inverted faces, even with up to 10,000 trials of practice.

We also investigated CP across a range of tasks. A binary classification task (*Face 1* or *Face 2* response) was used to determine the predicted category boundary for each subject. We then examined CP using three different measures of discriminability, namely the ABX task, the better likeness task, and similarity ratings. Previous studies of CP for face identity have shown clearer CP effects with better likeness than with ABX discrimination (Beale & Keil, 1995). Similarity ratings have also been used to examine CP (e.g., Livingston et al., 1998, for nonface objects, and Stevenage, 1998, for faces). Unlike the better likeness task, similarity ratings can be made without knowledge of the endpoint faces. Thus, they can be used to assess baseline (pretraining) similarity of equidistant pairs across the face continua to show that any CP following training cannot be attributed simply to nonlinearities in the methods used to produce the intermediate images.

The three tasks we used to examine CP differed from each other in a number of respects, including response requirements, the extent to which subjects' perception of categorization was assessed directly (similarity ratings) versus indirectly (ABX and better likeness), and many experimental details such as presentation duration of the stimuli and the provision of feedback (see *Method* section). Our results are strengthened substantially by the agreement we report across two forms of noise and across these very different tasks.

## Method

### Subjects

Five trained subjects saw approximately 1,000 to 10,000 trials with images from at least one face continuum. Some subjects completed training on two face continua. This involved 2 hr to 14 hr of testing per subject per face pair, generally spread over weeks to months. Three subjects (A.D., A.E.S., and P.G.) were naive as to the purposes of the experiment, and 2 (E.M. and P.M.) were not. In addition, 7 naive control subjects were tested on the similarity rating task without prior exposure to the endpoint faces. All subjects were volunteers. The subjects were of the same race (Caucasian) as the face stimuli on which they were tested.

### Face Stimuli

The primary stimuli used in Experiment 1 were continua between two pairs of faces, referred to as *female fades* and *male morphs.* The endpoint faces (Face 1 and Face 2) for each of these pairs are shown in Figure 1. Pairs of endpoint faces were selected to be highly similar to each other: They were the same age, sex, and race; they had the same expression (neutral); they had no hair, no facial hair, no distinguishing marks, and no spectacles; and their internal features fell in overlapping positions. Hair was removed from the images through a hand-drawn window showing only the internal features of the face. This window was the same shape for all faces on a given continuum. All images were gray scale. Female faces were presented on a white background, and male faces were presented on a gray background. Adobe Photoshop (Version 4.0, Adobe Systems Incorporated, 1997), was used in matching the endpoint faces for each pair on brightness and contrast.

Intermediate images between the two endpoints were produced in the following way. For the female fades, the intensity of pixels in each endpoint parent image was weighted by the desired proportion of that image in the intermediate compound; that is, one face was "faded in" and the other was "faded out," keeping the overall brightness constant. For the male morphs, a full morphing procedure was used (Morph, Version 1.0, Gryphon Software, 1992). This procedure combined intensity fading with warping of the intermediate images to ensure exact lineup of multiple points specified by the experimenters (e.g., center of pupils and corners of eyes), thus producing photographic quality intermediate images. Figure 1 shows the image halfway between Face 1 and Face 2 for each pair. Intermediate stimuli were labeled by the percentage of Face 2 in the image (i.e., 0%, entirely Face 1; 50%, halfway between Face 1 and Face 2; 100%, entirely Face 2). Both fading and morphing procedures were expected to produce (for untrained subjects) equidistant perceptual changes in face identity for equidistant changes in the objective proportion of Face 2 in the image (this was confirmed through similarity ratings made by untrained control subjects; see *Results and Discussion* section).

Finally, noise was added to the images. Two methods of producing noise were used, shown in Figure 3. For most face stimuli and tasks (classification, ABX discrimination, and similarity rating tasks for both face pairs and threshold better likeness task for female fades), the noise was "additive uniform noise" of strength 70 in Photoshop 4.0; that is, to each pixel intensity (range 0–256), a value drawn from a uniform distribution with a mean of zero and a range of −70 to 70 was added. For the male morphs in the better likeness task, the noise involved randomly moving 25% of the pixels within the image; specifically, on each trial a noise frame was created by an exhaustive swap of random pixel pairs in the original image, and then 25% of the pixels in the original image were randomly replaced with those of the same position in the noise frame.

A pair of *smiling female morphs* was also used for one subject in one task of Experiment 1. The endpoint images, shown in Figure 1, satisfied the same criteria as the two primary face pairs, although the endpoint faces were now smiling. Intermediate images were created with a full morphing procedure.

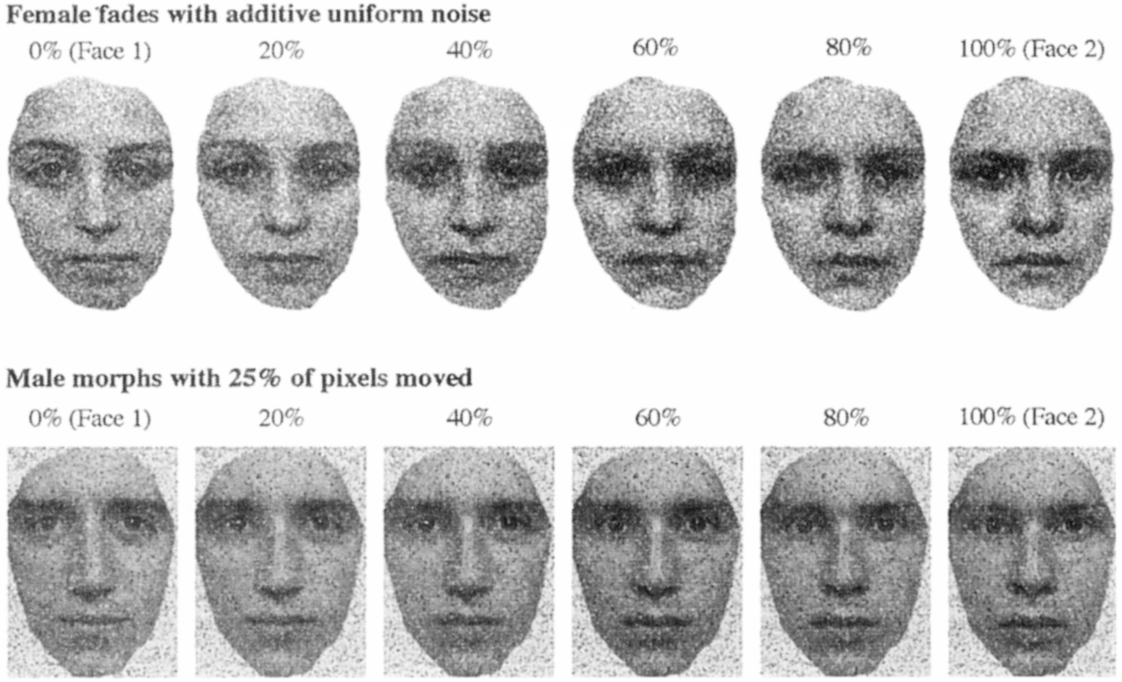


*Figure 3.* The two types of noise (high level) added to the face continua in Experiment 1. All stimuli are labeled in terms of the percentage of Face 2 in the image. Figure should be viewed at a distance of 25 cm. All subjects gave consent for their photographs to be used for scientific studies.

### Procedure

Trained subjects completed the classification task and at least one of the three discrimination tasks. The general order in which tasks was performed was as follows: classification training, ABX discrimination, additional classification training, threshold better likeness task, and, finally, similarity ratings. In some cases, trained subjects received additional exposure to the stimulus faces in later experiments or in tasks not discussed in the present article, interspersed with those reported. To clarify the extent of prior practice with the faces on each task, the figures showing results include running totals of each subject's number of trials performed with the target continuum at the beginning and end of data collection on each task. Control subjects completed only the similarity rating task, with no prior exposure to the endpoint faces or any of the intermediate images.

All stimuli were presented on Power Macintosh computers with high-resolution screens, running in gray scale mode. PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) was used to present stimuli in the classification, ABX discrimination, and similarity rating tasks, whereas in-house software developed within the Vision Shell environment was used for the better likeness task. Viewing distance was not strictly controlled but was approximately 55 cm. At this distance, face stimuli subtended visual angles of 5.1° horizontally and 6.9° vertically.

Each of the four tasks is described in detail subsequently. Figure 4

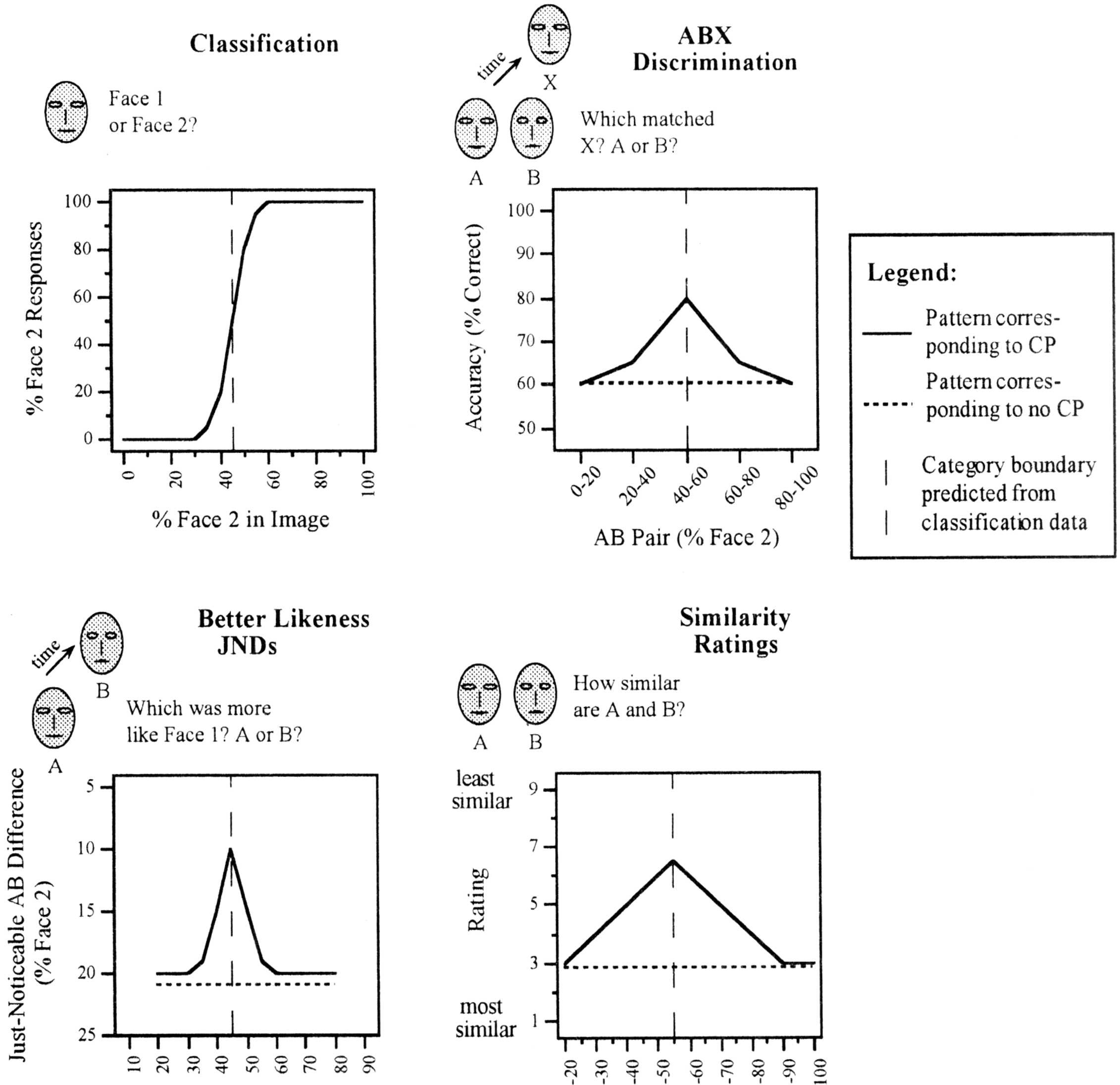


*Figure 4.* The four tasks used in Experiment 1. For all plots, the *x*-axis shows the continuum between Face 1 (left end) and Face 2 (right end). Note the reverse axis used for the better likeness just-noticeable differences (JNDs) and the coding of higher ratings as less similar for the similarity rating task; these methods resulted in the pattern corresponding to categorical perception (CP) being a peak rather than a trough in all tasks. There is no prediction of the width of the peak, and thus a range of possible forms are shown across the three discriminability tasks.

summarizes the procedure for each task, the conditions tested, and the measure obtained.

*Classification task.* Before beginning each block of the classification task, subjects examined the two endpoint faces, labeled Face 1 and Face 2, for as long as they wished. Endpoint faces were examined either upright or inverted, as appropriate for the upcoming block of trials. Subjects then made binary classifications of a series of images from the continuum between the endpoints as either Face 1 or Face 2. No feedback was given.

In each block, 150 trials were presented upright, in random order, followed after a break by 150 inverted trials. The 150 trials comprised 10 trials with each of the 15 images containing 0, 10, 20, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90, and 100% Face 2. On each trial, the face image was shown for 500 ms, after which a further 2,000 ms was allowed for a response before the next trial was presented. Different random noise was used for each of the 15 tested points on the continuum; once assigned, however, this noise remained the same for all trials presenting each of these images. (After several blocks, completely new sets of random noise were assigned for some subjects.) Each subject completed 1 to 10 blocks of classification (300–3,000 trials). The purposes of the classification task were (a) to give subjects a large amount of practice with faces from the target continuum, (b) to ensure thorough learning of the 0% and 100% endpoint faces, and (c) to determine the predicted category boundary.

*ABX discrimination task.* In the ABX discrimination task, two images differing by 20% of the continuum (A and B) were presented side by side for 500 ms. After a 400-ms blank screen, the target image (X) was then presented alone in the center of the screen, and the subject had 2,000 ms to indicate whether X matched A or B. All possible combinations of the following conditions were used equally often: the target on the left, the target on the right, the target being the image more like Face 2, and the target being the image less like Face 2. No feedback regarding accuracy was given. As in the classification task, different noise was used for each distinct point on the continuum, but all trials involving that point had the same noise assignment. This meant that X matched the image A or B exactly (i.e., it was not the same face with different random noise).

The AB image pairs tested were 0%–20%, 20%–40%, 40%–60%, 60%–80%, and 80%–100%. Each block presented 40 trials of each of these pairs in random order, first upright (200 trials) and then inverted (200 trials). Data are reported averaged over one to three blocks per subject.

The purpose of the ABX discrimination task was to examine any CP for upright and inverted faces. This would be demonstrated as more accurate discrimination for pairs that straddled the predicted category boundary than for equidistant pairs away from the category boundary.

*Better likeness task.* In the better likeness task, we measured the minimum difference between the stimuli necessary to produce a fixed accuracy (the *just-noticeable difference,* or *JND*) rather than measuring the accuracy for fixed differences between the stimuli, as did Beale and Keil (1995). This was done to obtain the highest possible sensitivity in a single-subject design.

Subjects indicated which of two sequential images (A then B) was more like Face 1. Each image was presented for 200 ms with an intervening 600-ms blank screen. There was no time limit on response. Feedback regarding accuracy was given on every trial. A staircase procedure (QUEST; Watson & Pelli, 1983) was used to determine the JND in percentage of Face 2 between the A and B images, that is, the minimum image difference around a given midpoint needed to produce a predetermined discrimination accuracy of 82% correct. The QUEST procedure increased or decreased the difference between A and B in a quasi-random fashion depending on the accuracy of the preceding response. Sequences of approximately 35 to 80 trials around a given midpoint were required to converge on the just-noticeable AB difference.

Midpoints tested were 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, and 80% Face 2. Data were obtained from at least two QUEST runs at each of these points (separately for upright and inverted), with three or more runs averaged when the first two values were very different from each other or from those for neighboring points. JNDs were determined to the nearest 1% difference on the face continuum. For the female fades, this was done by blending different proportions of Face 1 and Face 2 (with their preassigned Photoshop noise) on-line for every trial. For the morphed images, it required creating and saving in advance all 101 morphs between Face 1 and Face 2 in 1% steps. For these images, noise was then added by randomly moving a different 25% of the pixels in the image on every trial.

The primary purpose of the better likeness task was to examine any CP for upright and inverted faces. This would be demonstrated by smaller JNDs across the predicted category boundary than away from the category boundary. The use of the better likeness task also clearly focused the subject on face identity rather than on image match as in the ABX discrimination task. Finally, the better likeness task provided additional thousands of training trials with the faces, this time with explicit feedback as to which of the two faces was more like Face 1 on each trial.

*Similarity rating task.* Subjects rated pairs of faces differing by fixed amounts of Face 2 for similarity using a 9-point scale ranging from *most similar* (1) to *least similar* (9). The members of each pair were presented simultaneously. There were no limits on either viewing time or response time. In this task, new random noise was added to each target image on each trial. Thus, subjects were instructed to rate the similarity of face identity and to ignore any differences in local image details (e.g., dark blobs in a certain region) arising from different noise assignments.

Anchor points for the use of the scale were provided and shown simultaneously with the target stimuli on each trial. Anchors were taken from a morph continuum between a pair of new faces. To indicate the appropriate use of 1 on the scale, the 50% morph from this new continuum was shown with two different random noise assignments. To indicate the use of 9, two morphs further apart were chosen. Thus, subjects awarded a 1 if they thought the members of the target pair were as similar as the first anchor pair (i.e., "looks like the same face with different random noise") and a 9 if they thought the members of the target pair were as dissimilar as the second anchor pair.

Target pairs tested included 17 pairs differing by 20% of the continuum (0%–20% through 80%–100% Face 2 in 5% steps). In addition, 4 pairs were included that differed by 0% (0–0, 40–40, 60–60, and 100–100), and 3 pairs were included that differed by 40% (0–40, 30–70, and 60–100). The 24 trials in each block were presented in random order, first upright and then inverted. Trained subjects rated four blocks of trials, and means were taken of the four ratings for each pair. Control subjects rated only two blocks of trials. Thus, their data were more variable than those of the trained subjects, but the chance of learning the endpoint faces was reduced.

The primary purpose of the similarity ratings was to examine CP for upright and inverted faces in trained subjects. CP would be demonstrated by ratings of less similarity (i.e., higher rating scores) for those 20% difference pairs that crossed the predicted category boundary than for equidistant pairs away from the category boundary. The similarity rating task also involved unlimited exposure durations; this is important because it means that any lack of CP cannot be attributed simply to insufficient time to process the stimuli. The inclusion of the 0% and 40% difference pairs also allowed us to confirm that subjects were able to detect objective differences in similarity (i.e., rating scores for 40% pairs should be higher than those for 20% pairs, which in turn should be higher than those for 0% pairs) and to assess the sensitivity of the scale. Finally, the similarity rating task allowed untrained control subjects to be tested.

## Data Analysis and Statistical Criteria

Before presenting the results, it is necessary to consider at some length the basis for determining the presence or absence of categorical perception. The theoretical perceptual distortion function we presume to underlie CP is shown in the top left panel of Figure 5 (for similar assumptions, see Ehret, 1987, p. 305, and Harnad, 1987a, p. 555). The function indicated is a

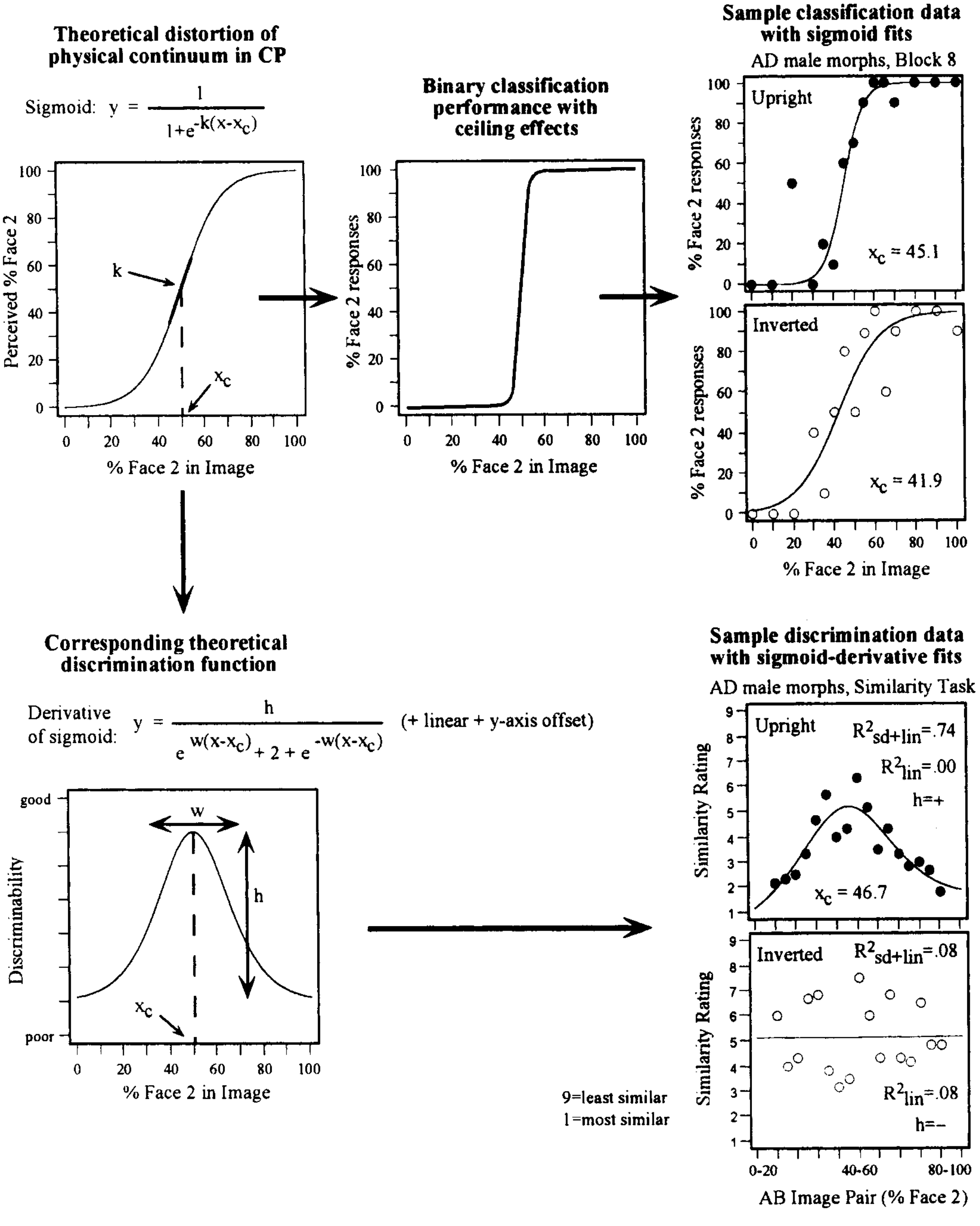


*Figure 5.* Theoretical basis for fitting classification and discrimination data, along with sample fits. If the perceptual distortion of the physical continuum underlying categorical perception (CP) is assumed to be described by the sigmoid function shown at top left ($x_c$ = predicted category boundary, $k$ = slope of tangent at $x_c$), then the corresponding discrimination function should be described by the derivative of this sigmoid as shown at bottom left ($x_c$ = center of peak, $h$ = height of peak, $w$ = width of peak). Fits to sample observed data (from A.D.), as shown at right, demonstrate a case of CP (upright) and a case of no CP (inverted). Proportions of variance explained by the sigmoid-derivative fit ($R^2_{\mathrm{sd+lin}}$) and by a linear-only fit ($R^2_{\mathrm{lin}}$) are provided.

sigmoid (formally, it is the logistic function), and the general form of its equation is given ignoring scaling parameters. Here, the perceived proportion of Face 2 in the image does not reflect the actual proportion of Face 2 (this would give a linear function with a slope of 1) but instead reflects the expansion of differences around a category boundary ($x_c$), and/or the compression of differences away from the boundary, to produce a higher slope at the boundary ($k > 1$).

The first point to be made is that high slopes obtained on a classification task do not necessarily reflect underlying perceptual distortion and can arise in the complete absence of CP. High binary classification slopes merely demonstrate that subjects have successfully learned to split the continuum in two when asked to do so. The top middle panel of Figure 5 shows the actual type of classification performance that can arise given ceiling effects on accuracy of classification. In this example, the fact that the 60% Face 2 image is classified as Face 2 on 100% of occasions does not necessarily indicate that the subject perceives that image as indistinguishable from Face 2 itself, nor does it necessarily show that it is perceived as more like the 80% image than like the 40% image. Thus, slopes determined from classification data do not bear any direct relationship to the slope of the presumed underlying distortion function (in the presence of ceiling effects, they are overestimates of the amount of any distortion).

Binary classification results are limited to prediction of the position of any category boundary. To determine this, we fitted sigmoid functions to the observed classification data, as shown for one subject in one block in the upper right portion of Figure 5. The center point of the fitted function ($x_c$) is indicated; this represents the predicted category boundary. Note that sigmoid fits generally explained a high proportion of variance in the classification data ($R^2 > .7$) and produced random patterns of fit residuals, arguing that the sigmoid provides a suitable description of the data.

Data from the discrimination tasks (ABX, better likeness, or similarity ratings) are necessary to determine the presence or absence of CP. To demonstrate that CP has occurred, there must be a peak in discriminability around the category boundary predicted from the classification task. A flat discrimination function, in contrast, indicates no distortion of the physical continuum. In instances in which multiple points across each face continuum were tested (i.e., 13 for better likeness and 17 for similarity ratings), we determined the statistical reliability of the presence or absence of peaks by fitting what we term a *sigmoid-derivative* function to each subject's discrimination data. The sigmoid derivative was selected because discriminability must reflect the slope (i.e., derivative) of the underlying perceptual distortion function.

The lower left panel of Figure 5 shows the general form of the sigmoid-derivative equation. The three parameters describing the peak are its height ($h$), its width ($w$), and the position of its center ($x_c$), as indicated. To the function plotted, we also added a linear component ($bx$) to allow for the overall left-to-right trends in discriminability that we observed in a number of cases (as did Levin & Beale, 2000). The theoretical origin of such trends remains obscure, but the inclusion of a linear term in our fits allowed for the possibility that a peak could be superimposed on an overall trend. Note that fitting the sigmoid-derivative-plus-linear function to our discrimination data produced random patterns of fit residuals; that is, the function did not consistently overestimate or underestimate the observed scores in any region of the curve.

In evaluating the fit results for statistical evidence of the presence or absence of CP, we considered three pieces of information: (a) the sign of the parameter $h$, (b) the value of the parameter $x_c$ and its standard error, and (c) the proportion of variance explained by the sigmoid-derivative fit including a linear component ($R^2_{\text{sd+lin}}$) relative to the proportion of variance explained by a simple linear fit to the data ($R^2_{\text{lin}}$). The following three results would indicate that CP is present. First, $h$ must be positive; that is, any deviation from a straight line must be in the direction of a peak rather than a dip. Second, any peak must be in the correct place; that is, $x_c$ for the discrimination data should agree (within error) with $x_c$ predicted from the classification data. Third, the proportion of variance explained by the sigmoid-derivative fit should exceed that explained by the linear-only fit; that is, $R^2_{\text{sd+lin}}$ should be substantially and significantly greater than $R^2_{\text{lin}}$.

If, in contrast, CP is absent, there should be no evidence for any peak in the data. First, conditions in which CP is absent should show $h$ scattered on both sides of zero across different data sets.[2] Second, the difference between the proportion of variance explained by the sigmoid-derivative fit and the linear-only fit should be nonsignificant, and, more strongly, the additional amount of variance explained should be very small in absolute terms (i.e., $R^2_{\text{sd+lin}}$ should be equal to $R^2_{\text{lin}}$).

The lower right section of Figure 5 shows sample fits to data for upright and inverted faces, for 1 subject in one task, along with corresponding values of $h$, $x_c$, $R^2_{\text{sd+lin}}$, and $R^2_{\text{lin}}$. The upright data demonstrate a case of reliable CP. Here, $h$ is positive, $x_c$ from the classification data (45.1) agrees well with $x_c$ from the discrimination data (46.7), and the difference between $R^2_{\text{sd+lin}}$ and $R^2_{\text{lin}}$ (.74) is large and statistically significant, $F(3, 12) = 11.38$, $p < .01$. The inverted data, in contrast, demonstrate a case of no CP. In this case, $h$ is negative, and no additional variance whatsoever is explained by the sigmoid-derivative fit over the linear-only fit ($R^2_{\text{sd+lin}} - R^2_{\text{lin}} = .00$).

Fit results for all data sets described in this article can be found in the Appendix. In the next section, relevant aspects of these fits are summarized.

## *Results and Discussion*

### *Classification Task*

Averaged across the last three blocks of classification trials, all subjects were 97%–100% accurate in classifying the endpoint faces as either Face 1 or Face 2. This was true both for upright faces and for inverted faces.[3]

After sigmoid functions had been fitted to the classification data, the predicted category boundary was averaged over all blocks for each subject, excluding any early blocks in which knowledge of the endpoint faces was unreliable. The predicted boundary was generally close to the middle of the continuum, presumably because subjects were trained with a set of stimuli containing no objective bias toward Face 1 or Face 2. Small subjective biases were apparent (range of center points: 42.4%–55.2% Face 2), and thus predicted category boundaries were determined on an individual subject basis, separately for upright and for inverted.

We also examined the slopes of the fitted sigmoids at the center point. Although slopes tended to increase with practice (particularly for upright faces), there was a clear inversion effect on slopes. Across all subjects, 36 classification blocks were presented in high noise. Of these blocks, 33 showed higher slopes for upright faces than for inverted faces (upright, $M = 4.57$, $SD = 2.57$; inverted, $M = 2.29$, $SD = 1.25$). Given that identical physical images were used with upright and inverted presentation, this result argues that the configuration of the upright face does, as we have proposed, allow more reliable identification of intermediate morphs. However, as described earlier, binary classification slopes do not bear any direct relationship to the presence or absence of CP. Furthermore, note that the slope differences between upright and inverted faces are only relative and that even inverted slopes were well above 1. As with most inversion effects in face recognition, then, differences in classification slope are *not* suitable for isolating configural processing.

### *ABX Discrimination Task*

Figure 6 shows ABX discrimination data for three subjects. Recall that, in this task, we plot accuracy (percentage correct) of

---

[2] We examined only the sign of $h$ rather than its value. The exact value is highly unreliable in the case of no peak, because a peak is described by width as well as height: If the function is in fact flat, any value of $h$ can be fitted as long as the peak is chosen to have zero width.

[3] We tested 1 additional trained subject, for whom results are not presented. This participant showed unsatisfactory learning of the endpoints even after seven blocks of classification trials (e.g., for inverted stimuli, he classified the Face 1 endpoint as Face 2 on 20% of trials). He was also unable to perform the better likeness task (e.g., his JNDs exceeded the full range of the continuum even for upright faces at midpoints of 40% or 60% Face 2), consistent with his poor knowledge of the endpoint faces.

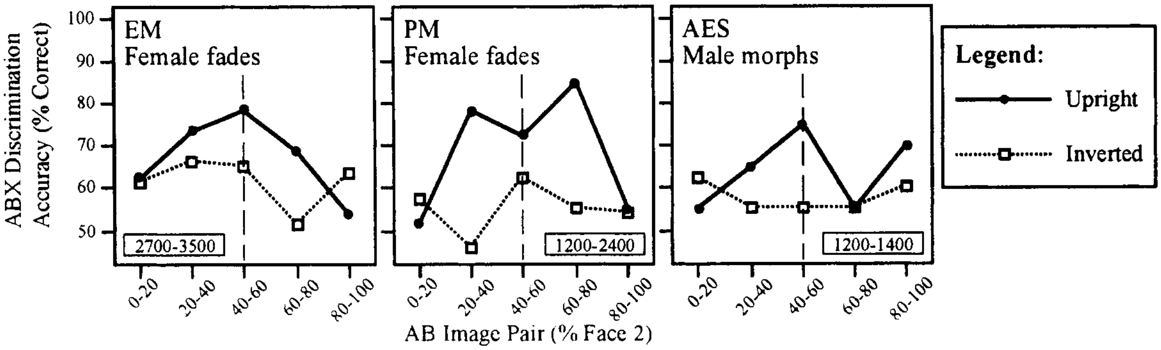


*Figure 6.* Experiment 1: ABX discrimination data for 3 individual subjects (E.M., P.M., and A.E.S.), showing categorical perception (CP) for upright but not inverted faces. Chance performance is 50% accuracy. In all cases, category boundaries determined from the classification task were close enough to 50% Face 2 that the predicted category boundary was the 40%–60% pair. The number of trials at the beginning and end of testing is indicated in the rectangle.

matching X to A or B against fixed differences between pairs of faces taken from different regions of the continuum. The results were generally consistent with CP for upright faces (i.e., more accurate discrimination across the category boundary predicted from the classification data than away from the boundary); in contrast, there was no indication of CP for inverted faces. For upright faces, differences in the percent correct across the continuum were statistically significant for E.M. on female fades, $\chi^2(4, N = 400) = 13.90, p < .01$, and for P.M. on female fades, $\chi^2(4, N = 600) = 10.14, p < .05$, but not for A.E.S. on male morphs, $\chi^2(4, N = 200) = 5.56, p > .1$. For inverted faces, differences did not approach significance in any case: E.M. on female fades, $\chi^2(4, N = 400) = 4.90, p > .2$; P.M. on female fades, $\chi^2(4, N = 600) = 7.14, p > .1$; and A.E.S. on male morphs, $\chi^2(4, N = 200) = 0.82, p > .2$. Note that sigmoid-derivative fit analyses were not carried out because only five points across the continuum were tested.

The data reported in Figure 6 were collected relatively early in practice. Attempts to test other subjects, or to retest the same 3 subjects at later points in practice were less successful. When subject A.D. was tested with male morphs (data not shown), her ABX discrimination was at chance levels for all conditions (even upright pairs crossing the predicted boundary), despite clear evidence of CP on the better likeness and similarity rating tasks (as described subsequently). Subjects tested after more practice with the images often showed ceiling effects, with accuracy greater than 90% in many conditions. When subjects indicated that their excellent performance relied simply on learning to match local image information (i.e., they could remember a locally bright region, or local texture, from the images presented 400 ms earlier), we attempted to force matches based on face identity by having X differ from both A and B in noise assignment, contrast, and overall brightness. These efforts failed when performance fell to chance in all conditions. Various manipulations of presentation duration and successive versus simultaneous presentation did not solve the problem.

Thus, although the ABX discrimination task produced some evidence of CP for upright faces, the effect was rather unreliable. Importantly, however, no subject tested ever showed any suggestion of CP for inverted faces (i.e., all curves for inverted faces were as flat as those shown in Figure 6). In general, we found the ABX task to be prone to methodological problems of floor and ceiling effects. We suspect the reason for this is twofold: (a) the explicit task requirement in ABX is of a match in image rather than necessarily of a match in face identity; and (b) the fact that the matching is made to only a short-term memory representation makes it possible for low-level image information to become the primary determinant of performance.

### *Better Likeness Task*

Figure 7 shows just noticeable differences in the better likeness task. Recall that, in this task, the subject determined which of A or B was more like a long-term memory representation of Face 1, and we plotted JNDs between A and B as a function of the position of their midpoint on the face continuum. From Figure 7, it can be seen that all subjects showed CP for upright faces (i.e., better discrimination around the predicted category boundary), whereas none showed any suggestion of CP for inverted faces. This was the case even though subjects were given trial-to-trial feedback on their accuracy in selecting the stimulus most like Face 1 and even though 5 of the 6 had completed 7,000–10,000 trials with the target faces by the end of data collection on this task.

Sigmoid-derivative fits were made to the better likeness data shown in Figure 7 for each subject except P.G. (who was tested at only five points along the continuum). These fits provided strong statistical support for the presence of CP for upright faces (see Appendix for details). The peak height parameter *h* was positive in all cases, and the sigmoid-derivative-plus-linear fits explained between 33% and 96% ($M = 68\%$) more variance in the data than did linear-only fits. This improvement was significant in all cases ($ps < .05$). Throughout this article, all cases of significant peaks occurred in the correct position; specifically, $x_c$ obtained from the discrimination data agreed with $x_c$ predicted from the classification data within 1.5 standard errors. The fits also supported the absence of CP for

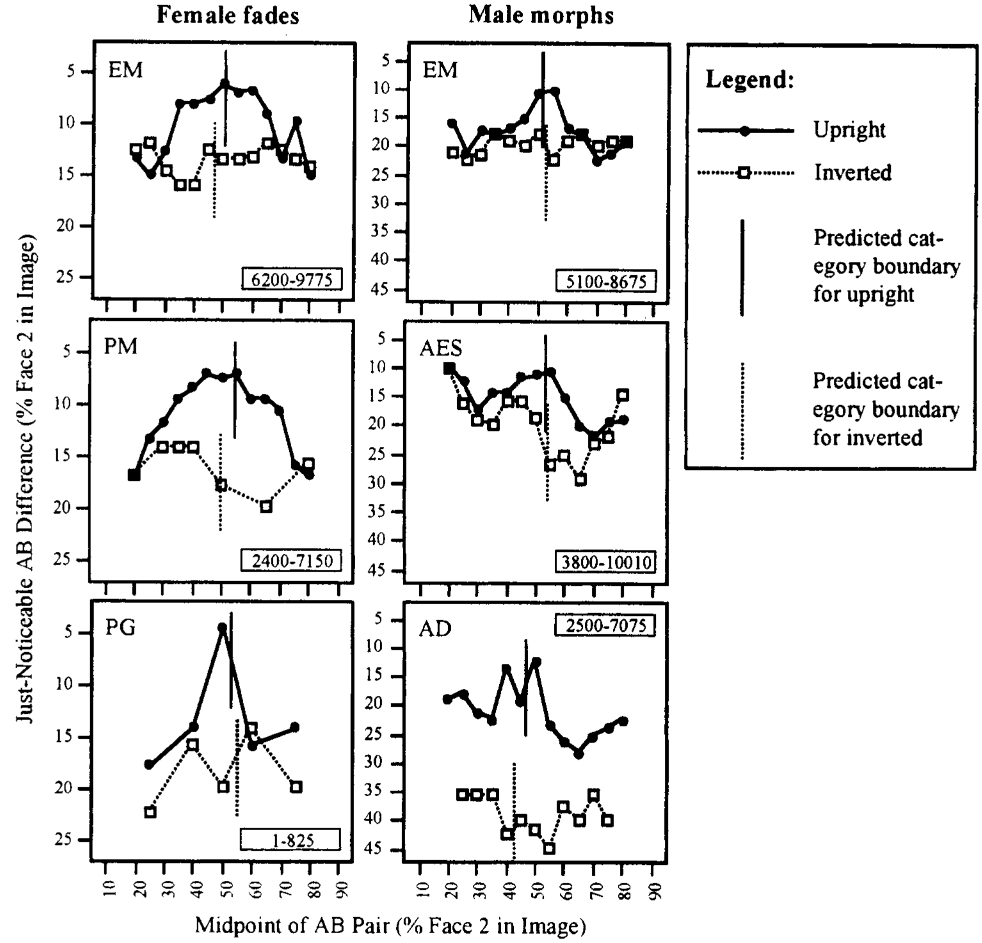


*Figure 7.* Experiment 1: Just-noticeable differences (JNDs) in the better likeness task, again showing categorical perception (CP) for upright but not inverted faces. Data are shown for 3 subjects tested on the female fades (E.M., P.M., and P.G.) and for 3 subjects tested on the male morphs (E.M., A.E.S., and A.D.). Note the different *y*-axis scales for the two sets of stimuli. To assist in interpreting the JND measure, consider the upright data for E.M. on female fades. The JND in the morph images was approximately 6% at the predicted category boundary; that is, the minimal-difference pair discriminated with 82% accuracy was the 47%–53% pair. Away from the category boundary, JNDs were more than twice as large; around the 20% Face 2 point, the JND was approximately 14%, indicating a minimal-difference pair comprising the 13% and 27% morphs.

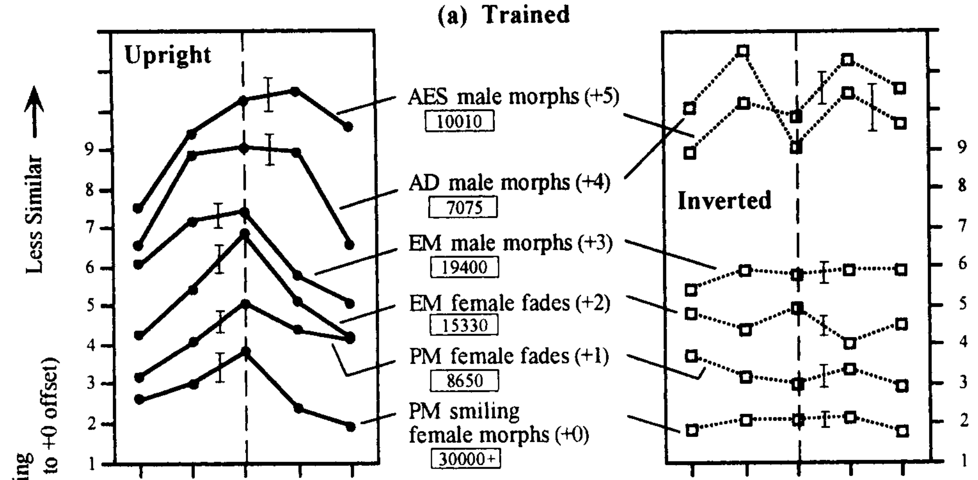
inverted faces: *h* was negative (four cases) at least as often as positive (one case), and 0% additional variance was explained by the sigmoid derivative in four cases, with the exception (A.D. on male morphs) demonstrating a negative *h*.

Although they do not affect the conclusion of CP for upright but not inverted faces, some peculiarities in the data can be noted. For example, subject A.E.S. on the male morphs had an overall left-to-right trend for both upright and inverted faces, whereby she was more sensitive to differences between face pairs the more those pairs fell toward the Face 1 end of the continuum. For this subject, the peak in discriminability for upright faces was superimposed on this overall trend. Second, subject A.D. on the male morphs had much worse overall performance with inverted faces than upright faces (to the point where we could not determine thresholds around the 20% and 80% midpoints as a result of trials overflowing the 0% and 100% points on the continuum). In contrast, the other subjects showed similar thresholds for upright and inverted faces away from the category boundary.

### *Similarity Rating Task: Trained Subjects*

Similarity ratings of pairs differing by 20% of the face continuum were examined to assess CP for upright and inverted faces. The raw data comprised ratings for each of the 17 pairs (0%–20%, 5%–25%, 10%–30%, and so on). To show data from multiple subjects in the same figure, we grouped the pairs into five points along the continuum according to their distance from the subject's own predicted category boundary, separately for upright and for inverted. Ratings from all pairs crossing the predicted boundary were averaged to provide the cross point. The four pairs at the extreme Face 1 end of the continuum were averaged to give the far (F1) point, and the four pairs at the extreme Face 2 end produced

(a) Trained



(b) Control

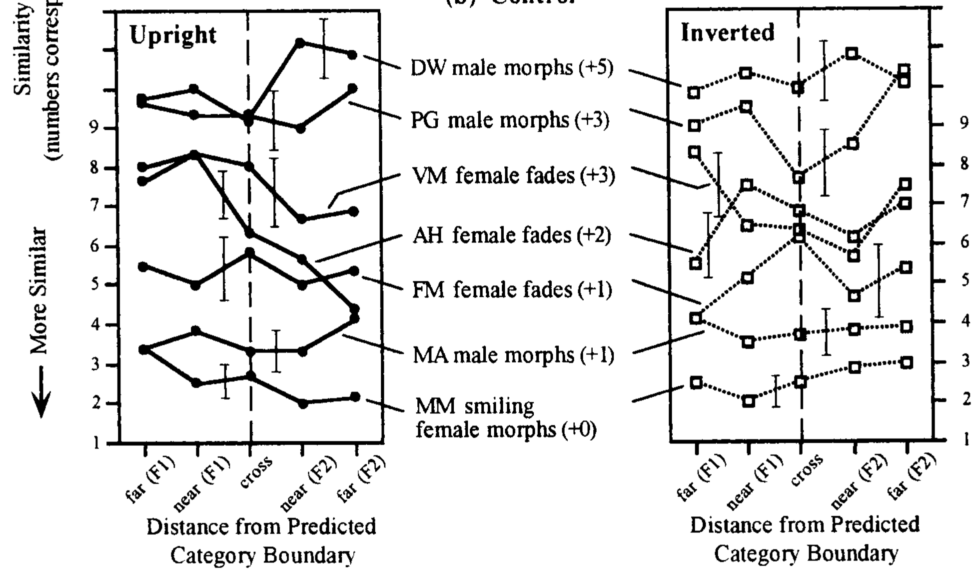


*Figure 8.* Experiment 1: Similarity ratings for (a) trained subjects (A.E.S., A.D., E.M., and P.M.), showing categorical perception (CP) for upright but not inverted faces, and (b) untrained controls (D.W., P.G., V.M., A.H., F.M., M.A., and M.M.) showing no distortion around the center of the continuum without prior exposure to the endpoint stimuli. Curves for individual subjects have been separated by adding offsets to the obtained ratings (e.g., +1 to all scores). For trained subjects, numbers of exposures before collection of the similarity rating data are indicated in the rectangles. Error bars indicate the average standard error of the means for each curve (i.e., ±1 *SEM*). Note that data are more variable for controls because fewer blocks of ratings were taken. See the text for an explanation of the distance from predicted category boundary groupings.

the far (F2) point. The remaining intermediate pairs were averaged to give the near (F1) and near (F2) points.[4]

Figure 8 (upper panels) shows six sets of similarity rating data. All demonstrate a pattern consistent with CP for upright faces, although subject A.E.S. on the male morphs again had this superimposed on an overall trend. No subject, however, showed any suggestion of CP for the same faces when presented inverted. Fit results (based on the 17 points actually tested across the face continuum) supported these conclusions. For upright faces, in all cases *h* was positive and the sigmoid-derivative-plus-linear fit explained significantly more variance than the linear-only fit (range = 43%–82% extra variance; $M = 60\%$). For inverted faces, *h* was scattered on either side of zero, and at most a nonsignificant 6% additional variance was explained by the sigmoid-derivative fit (range = 0%–6%; $M = 2\%$).

---

[4] A minor adjustment was made to this procedure for subjects with category boundaries more than 5% away from the middle of the continuum. Specifically, the closest point from the relevant far grouping was moved to the near grouping to ensure that ratings were averaged from at least two of the original points for each grouping.

We also examined the average similarity ratings for the pairs differing by 0%, 20%, and 40% of the continuum. This is of particular importance for the inverted faces, wherein we wished to attribute the flat functions of Figure 8 to a lack of CP rather than simply to a lack of sensitivity even to objective differences for inverted faces. Figure 9 (open squares) shows mean rating scores for the 0%, 20%, and 40% pairs for inverted faces in trained subjects and demonstrates the expected pattern; that is, increasing objective difference between the pairs produced higher ratings, reflecting less perceived similarity. (Figure 9 shows only means across all trained subjects; however, each subject considered individually showed the same trend.) Furthermore, the magnitude of the effect indicates that subjects' similarity ratings were sufficiently sensitive to have shown CP for inverted faces had CP been present.

### *Similarity Rating Task: Untrained Control Subjects*

Similarity ratings from untrained control subjects were examined to confirm that the presence of CP could not be attributed simply to nonlinearities produced by the fading or morphing procedures (specifically, to larger objective differences around the middle of the continua than toward the ends). In fact, the lack of any CP for inverted faces already argues that such stimulus artifacts cannot account for the CP obtained for the same images when presented upright. Ratings from controls led to the same conclusion. Figure 8 (lower panels) shows controls' ratings of 20% pairs with no previous exposure to the endpoint faces. There were no peaks in dissimilarity in the middle of the continua, as confirmed by fit results. Both for upright and for inverted faces, *h* was negative as often as positive, and sigmoid-derivative fits explained only a nonsignificant 0%–7% more variance than linear-only fits. Controls' mean rating scores for pairs differing by 0%, 20%, and 40% of the continuum are shown in Figure 9. These scores confirm appropriate ratings of less similarity for pairs objectively further apart, for both upright and inverted faces.

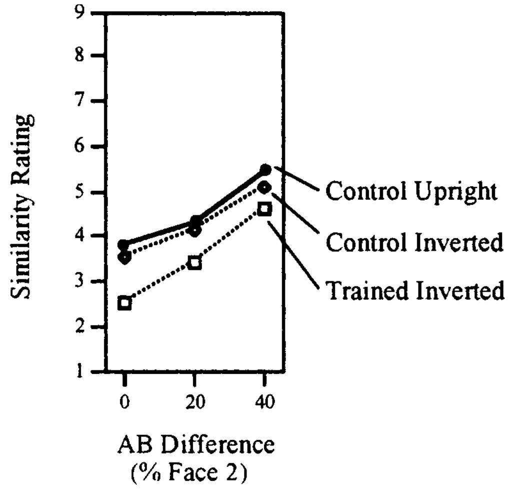


*Figure 9.* Experiment 1: Mean similarity ratings for all images differing by 0%, 20%, and 40% Face 2, showing less perceived similarity with greater objective difference. Data are not presented for the trained upright condition because it is inappropriate to average pairs from different regions of the continuum in the presence of distortion across a category boundary.

### *Summary*

Experiment 1 examined CP for the identity of similar faces presented in noise. Under these circumstances of poorly specified local information, we found clear evidence of CP for upright whole faces using two of three discriminability measures, namely better likeness JNDs and similarity ratings. The remaining discriminability measure, ABX discrimination, showed some evidence of CP but was generally beset by problems of floor and ceiling effects on performance. Similarity ratings obtained from control subjects demonstrated that the peaks found across the predicted category boundary after training could not be attributed to intrinsic nonlinearities in the stimulus continuum produced by the fading or morphing procedures.

For inverted faces, in contrast, we found no CP for any subject in any task. The lack of CP could not be attributed to a failure to learn the endpoint faces adequately when inverted (endpoint classification accuracy was above 95%), to insufficient time to process inverted stimuli (unlimited viewing time was provided in the similarity rating task), to a failure to use the scale for inverted faces in the similarity rating task (appropriate ratings were given to pairs that objectively differed in degree of similarity), or to a lack of statistical power in detecting a peak in the inverted data (the number of trials used was sufficient to show clearly significant peaks for upright; also, *h* was negative as often as positive, and the sigmoid-derivative-plus-linear fits explained almost no more variance in the data than did simple linear fits). Most important, the lack of CP could not be attributed to slower learning of CP for inverted faces than for upright faces. Trained subjects failed to learn CP for inverted faces despite (in four of the five cases) 7,000–30,000 exposures to faces from the target continuum. Furthermore, these exposures included 1,500–3,000 trials for inverted faces in the better likeness task, on which trial-by-trial feedback was provided. Table 1 clarifies the exact number of exposures to inverted faces at the final point of testing for CP across all experiments. It appears that our subjects were not able to learn CP for inverted faces in noise, no matter how well they knew the endpoint faces or how many times they had seen the intermediate images.

Thus, our results show a qualitative difference between performance for upright and inverted faces, with CP of faces in noise obtained *only* for upright faces. This is in contrast to the standard partial inversion effect shown in the previous literature, in which good residual performance with inverted faces indicates a nonconfigural contribution to overall performance. Thus, we have demonstrated that configural processing can operate in isolation from feature-based identification, not only after brain injury (as in patient C.K.) but in normal subjects as well.

## Experiment 2: CP for Faces Versus Features in High and Low Noise

In Experiment 1, we demonstrated that categorical perception of face identity in heavy noise exists for upright faces but not for inverted faces. In Experiment 2, we examined CP for whole faces and for a single feature presented alone at high and low levels of

Table 1
*Amount of Practice With the Trained Face Continuum at the Point of Last Testing of Categorical Perception for Inverted Faces in Noise, Across All Experiments*

| | | Number of trials | | |
|---|---|---|---|---|
| Subject | Face continuum | Total | Inverted whole faces | Time period |
| P.G. | Female fades | 1,120 | 560 | 3 days |
| A.E.S. | Male morphs | 10,200 | 5,100 | 3 months |
| P.M. | Female fades | 11,700 | 3,766 | 6 months |
| A.D. | Male morphs | 12,800 | 4,125 | 6 months |
| E.M. | Female fades | 20,500 | 4,648 | 6 months |
| E.M. | Male morphs | 24,400 | 6,711 | 6 months |
| P.M. | Smiling females | 32,000 | 15,000 | 2.5 years |

*Note.* No categorical perception was observed in any case. Shown for each subject are (a) the total number of trials with any images from the trained continuum, including whole faces in any orientation, along with face parts for subjects who participated in Experiment 2; (b) the number of trials from the trained continuum specifically with inverted whole faces; and (c) the time period over which subjects were tested with faces from the trained continuum.

noise, both upright and inverted. There were three related reasons for examining CP under these various conditions.

One reason was to assess whether adding noise to the stimuli, as in Experiment 1, is in fact necessary to isolate configural processing. Levin and Beale (2000) have found that without noise, CP occurs for inverted faces (and thus fails to isolate configural processing). Their endpoint stimuli, however, were much more distinct than ours (e.g., hair included, age differences, and mustache in one endpoint but not the other). For our very similar faces, it might be that noise is not necessary to remove nonconfigural contributions to CP; if this were the case, CP should be restricted to upright faces even when the noise level is reduced. In contrast, if noise is necessary to isolate configural processing, even with our stimuli, then we should find that CP occurs for both upright and inverted faces in low noise.

A second aim of Experiment 2 was to explore the source of any CP for inverted faces. In the introduction, we claimed that Levin and Beale's (2000) finding of this effect (without noise) could well have relied on CP for local shape information and thus did not necessarily indicate anything about *face*-like processing of the inverted stimuli. If this were the case, then CP should be found even for a single feature presented alone, as long as it is shown at a low level of noise. Furthermore, this single-feature CP should be obtained inverted as well as upright.

Finally, Experiment 2 evaluated our original rationale for adding noise. In the introduction, we proposed that the effect of heavy noise was to make any local contour or feature, taken alone, an unreliable indicator of face identity for the intermediate morphs, thus forcing the use of information integrated across large regions of the face. The classification slope data from Experiment 1 provide some indirect support for this idea, showing that classification of intermediate morphs was less reliable for inverted faces than for upright faces. In examining CP for a single feature presented alone, Experiment 2 addressed this issue more directly. If our explanation of the role of noise is correct, then no CP should be found for a single feature presented in high noise (e.g., the nose alone). Furthermore, CP should be impossible even when the single feature is presented upright and even when that feature is the one indicated by subjects to be the most informative for differentiating the intermediate morphs.

## *Method*

### *Subjects and Design*

Four independent data sets were obtained from three subjects, one of whom (A.D.) was naive as to the purposes of the experiment. All subjects had participated in Experiment 1 and were tested in Experiment 2 with a face continuum on which they were already highly trained. One subject (E.M. on male morphs) was tested for CP on the better likeness task, and three subjects (A.D. on male morphs, E.M. on male morphs, and P.M. on smiling female morphs) were tested for CP on the similarity rating task. The female fade continuum was not used in Experiment 2 because the blended images were unsuitable for presentation in low noise. Discrimination performance was examined in eight conditions for each subject. These conditions were formed by crossing the whole face versus a single feature, high noise versus low noise, and upright versus inverted presentation. All subjects were also tested on the classification task to determine the predicted category boundary for each condition.

### *Stimuli and Noise*

In Experiment 1, we used two forms of producing noise, both of which were effective in removing CP for inverted faces. To assess the impact of noise on feature-based CP, we again included both forms here. As in Experiment 1, 25% of pixels in the image were replaced from the noise frame for male morphs in the better likeness task, and Photoshop uniform noise of strength 70 was added for both face continua in the similarity rating and classification tasks. These manipulations composed the high noise level. For the low noise level, 5% of pixels were replaced for male morphs in the better likeness task, and Photoshop uniform noise of strength 15 was added in the similarity rating and classification tasks.

A low level of noise, rather than no noise at all, was used to disguise differences in skin texture along the continua arising from the morphing procedure. Specifically, the full skin texture of the two endpoint images rapidly becomes smoothed with small amounts of morphing away from the endpoints and remains smooth throughout the central two thirds or so of the continuum. Examples of the stimuli are shown in Figure 10. Note that, unlike the high noise, the low noise did not noticeably disrupt information about the shape of local contours (see also Figure 2).

### *Procedure*

For the whole face in high noise conditions, most of the data were taken from Experiment 1. For E.M. and A.D. in the similarity rating task, these

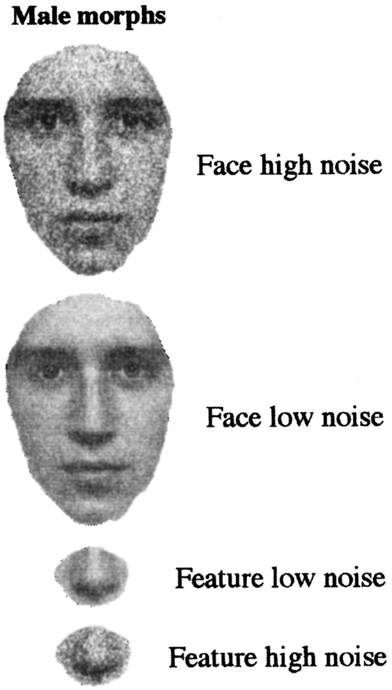


*Figure 10.* Examples of whole face and single-feature stimuli used in Experiment 2. Subject gave consent for his photograph to be used in scientific studies.

data were supplemented with additional runs taken for the same continuum (with different random noise assignments) at the end of Experiment 3 to allow us to present the data at a higher resolution than that used in Experiment 1.

The whole face in low noise conditions (upright followed by inverted) were the first conditions tested specifically for Experiment 2. After testing with low noise, subjects were asked to indicate which feature they found most useful in discriminating the faces, particularly when inverted. For our stimuli, this feature was reported to be the nose in all cases. This feature was then cut out from the whole face (using the same shaped window for all morphs on a given continuum), as shown in Figure 10. Next, subjects were tested in the CP task for the feature presented alone, first in low noise (upright and then inverted) and then in high noise (upright and then inverted). The single feature was shown at the same absolute size and viewing distance when presented alone as when presented as part of the whole face. Finally, participants completed two blocks of the classification task for each of the six conditions not tested in Experiment 1.

The procedure for all tasks was the same as in Experiment 1. In the similarity rating task, the anchor point images shown to indicate the use of 1 and 9 on the scale were the same as those used in Experiment 1 but were presented in a form that matched the test stimuli in each condition; for example, in the low noise single-feature condition, the anchor images also showed cutouts of noses presented in low noise.

In the better likeness task, discrimination data were based on a mean of at least two QUEST runs at each point, with different random noise used on every trial. In the similarity rating task, between 6 and 10 runs (one rating per point per run) were averaged. For this task, between two and four different random noise assignments were used in every condition. The total testing times for the six new conditions of Experiment 2 were approximately 21 hr for the one subject who performed the better likeness task and 6 hr for each of the three subjects who performed the similarity rating task. Testing was spread over approximately 2 weeks.

### *Results and Discussion*

Figure 11 shows discrimination functions in Experiment 2 for the whole face and for the single feature at high and low levels of noise. Similarity rating data are shown for all of the 20% difference pairs tested across the continuum rather than being collapsed into five groupings as was done in Experiment 1. Predicted category boundaries determined from the classification task are also indicated. According to the logic described in detail in Experiment 1, results of sigmoid-derivative fits (see Appendix) in all cases unambiguously supported the presence or absence of CP, as claimed subsequently.

#### *CP for Faces Versus Features*

The top row of Figure 11 shows the results of Experiment 1, updated by additional runs in cases in which these were taken. For whole faces in high noise, all subjects showed CP (i.e., a peak around the predicted category boundary) for upright but not inverted faces.

The second row of Figure 11 shows that, when the noise was reduced, CP emerged for inverted faces as well as for upright faces. In three data sets, CP was as strong for inverted faces as for upright faces, whereas P.M. (on smiling female morphs) showed a weaker but still significant effect for inverted faces.

The third row of Figure 11 shows that the CP for inverted faces in low noise does not need to be attributed to any face-like structure of the stimulus. In low noise, all subjects also showed CP for a single feature presented alone. Furthermore, there was no inversion effect for the single feature, with CP for the inverted nose as strong as that for the upright nose.

The fourth row of Figure 11 shows data for the single feature presented in high noise. For these stimuli, no CP was found in any case, either upright or inverted. (We suspect that the unusual *dip* for E.M. in the better likeness task arose from skin texture changes toward the end of the continuum that remained visible in the 5% pixel-moved images.) Figure 12 confirms that this lack of CP cannot be attributed to failure to use the scale in the similarity rating task: When the 0%, 20%, and 40% difference pairs were examined, two subjects (A.D. and P.M.) showed good discrimination of objective differences in similarity of the single feature in high noise, although one subject (E.M.) showed somewhat weaker discrimination.

#### *Summary*

Experiment 2 has demonstrated that, in low noise, CP occurs both for an inverted face and for a single inverted feature. In high noise, however, CP is not found for either type of stimulus and occurs only for an upright whole face. This confirms that CP for faces in high noise isolates configural processing: If the phenomenon reflects contributions of configural processing only, then it should not occur under any circumstances in which the face configuration is broken, whether this is through inversion (as was shown in Experiment 1) or through the presentation of a single isolated feature (as has now been shown in Experiment 2).

Three further conclusions can be drawn from Experiment 2. First, we have demonstrated that adding heavy noise is essential to isolate configural processing in the CP task, even when similar

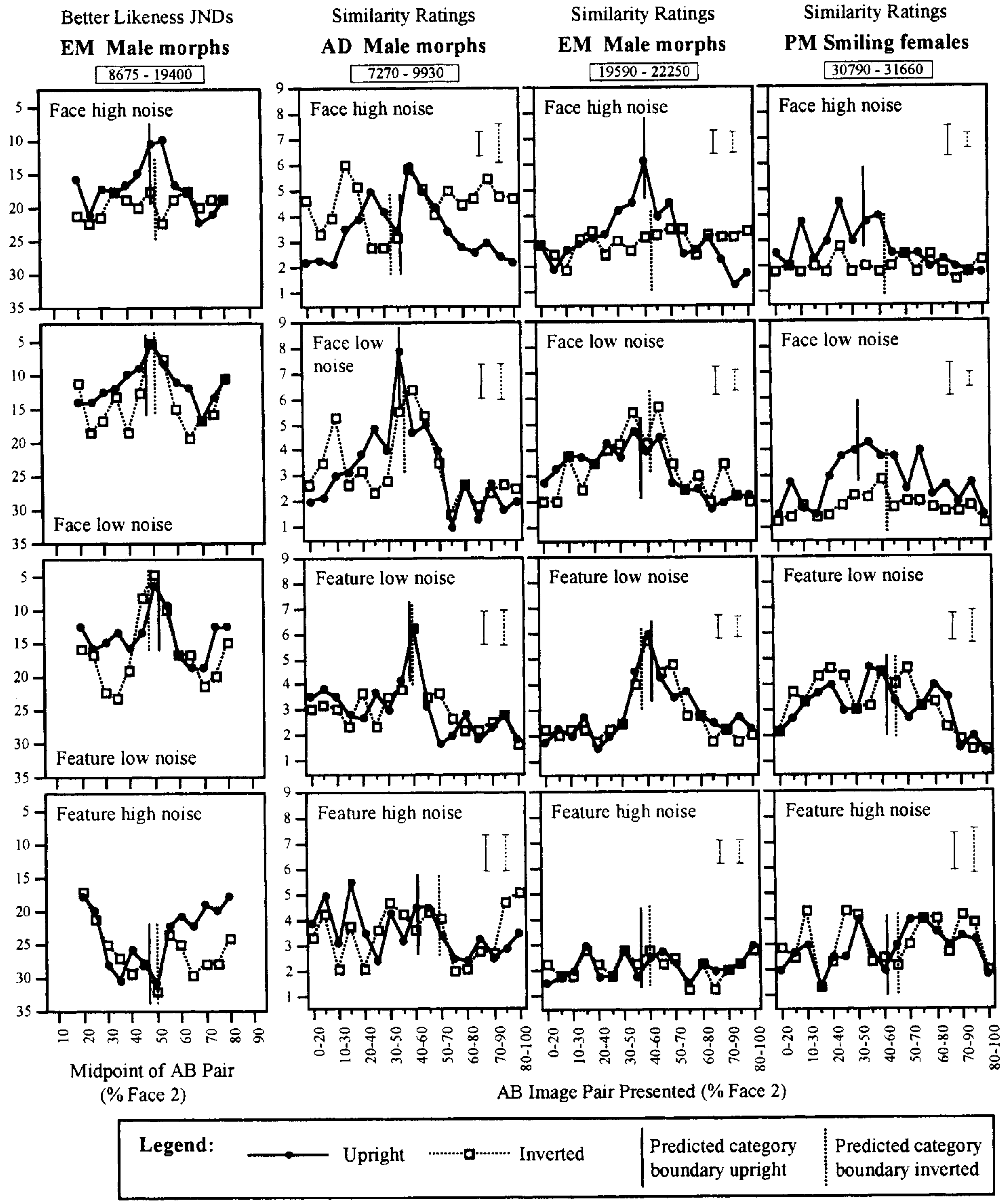


*Figure 11.* Experiment 2: Categorical perception for the whole face and for a single feature at high and low levels of noise. Data are shown for 1 subject in the better likeness task (E.M.) and for 3 subjects in the similarity rating task (A.D., E.M., and P.M.). Rectangles show number of trials at the beginning and end of data collection specifically for Experiment 2. Error bars (solid = upright; dotted = inverted) show ± 1 *SEM*, averaged across all points tested for each curve. The *y*-axes show just-noticeable AB differences (% Face 2) in the better likeness task or similarity rating scores (1 = most similar; 9 = least similar) in the similarity rating task. JNDs = just-noticeable differences.

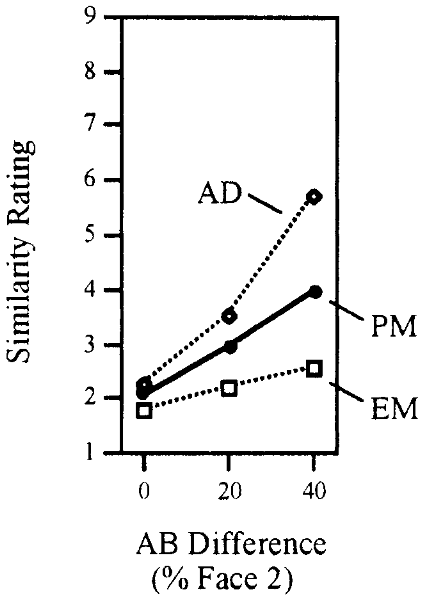


*Figure 12.* Experiment 2: Mean similarity ratings for images differing by 0%, 20%, and 40% Face 2, averaged across the single feature conditions that did *not* show categorical perception in the 20% difference data (i.e., upright and inverted nose in high noise). Data for A.D., E.M., and P.M.

endpoint stimuli are used (e.g., no hair, no distinguishing marks, and overlapping feature positions). In low noise, we obtained CP for inverted faces and for single features; thus, a high level of noise was needed to limit CP to upright whole faces.

Second, we have demonstrated that CP for inverted faces without noise, such as that observed by Levin and Beale (2000), does not need to be attributed to CP for face structure. The fact that CP occurs for many forms of nonface stimuli (e.g., Livingston et al., 1998) already argues that any CP for faces will not necessarily reflect the *face*-like properties of the stimulus, at least when clear pictures are used. Our finding of CP for a single feature in low noise confirms that CP can rely on local image-shape properties that have nothing to do with the structure of the whole face.

Finally, our results provide experimental support for the rationale that led to our original selection of noise in Experiment 1. We argued that, with high noise, different random noise assignments could change local contour shape sufficiently to shift an intermediate feature morph to the wrong side of the category boundary (see Figure 2). Experiment 2 confirms that the noise did indeed render unreliable the identify information extracted from any single local image region. In high noise, insufficient information remained in a single feature to support CP. This was true even though the feature chosen was that which subjects reported to be the most discriminating for our stimuli.

## Experiment 3: Image Plane Rotation

Experiments 1 and 2 demonstrated that CP for faces in high noise reflects configural processing only, with no confounds from feature-based identification. This is in striking contrast to most previous methods of investigating face recognition. The usual finding of some (often substantial) residual performance with inverted faces argues that performance for upright faces relies on a mixture of configural and featural processing. Furthermore, these two mechanisms could combine in many different ways to produce a given level of observed performance, and there is no principled method of determining the relative contribution of each. With the provision of a method for isolating configural processing, therefore, many questions about the nature of configural processing become tractable to experimental study in normal subjects for the first time.

In Experiment 3, we wished to demonstrate the usefulness of CP in noise as a technique by applying it to at least one question regarding the nature of the internal representations underlying configural processing. The question we addressed was how closely tuned the configural "prototype" is to upright, that is, the degree to which a face stimulus can be rotated in the image plane and still activate configural processing mechanisms. The answer to this question places constraints on any theoretical model of configural face recognition, in that it determines the degree of orientation mismatch that can be tolerated between the stimulus and the internal face representation. To investigate the orientation tuning of configural processing, we tested only whole faces in high noise and examined the presence or absence of the signature CP phenomenon at various rotations of the stimulus in the image plane.

### *Method*

#### *Subjects and Design*

Four independent data sets were obtained from three subjects, one of whom (A.D.) was naive as to the purposes of the experiment. Subjects were tested with face continua on which they had been highly practiced in Experiment 1 and, in two cases, in Experiment 2 as well. One subject (E.M. on female fades) was tested for CP with the better likeness task, and three subjects (E.M. on male morphs, P.M. on female fades, and A.D. on male morphs) were tested for CP with the similarity rating task. The smiling female morphs were not used in Experiment 3 because differences in angle of the smile and crookedness of the nose left some doubt about the exact orientation of the faces even when the eyes were aligned. All subjects were also tested on the classification task to determine the predicted category boundary for each condition. Six to seven orientation conditions were tested in 22.5° steps of clockwise rotation in the image plane, from 0° (upright) to 180° (inverted).

#### *Materials and Procedure*

The female fades and male morphs were presented as whole faces in high noise, as in Experiment 1. The procedure for each task was the same as in Experiment 1. In the similarity rating task, anchor point images were presented at the same orientation as the test faces.

Data taken at the 0° and 180° orientations in Experiment 1 were reused in Experiment 3. The first orientation tested specifically for Experiment 3 was 90°. If this orientation showed CP, 112.5° was tested. Other conditions were then tested in the order 45°, 67.5°, and 22.5°. For each orientation, two blocks of the classification task were performed. Subsequently, JNDs for the better likeness task were determined from the mean of at least two runs at each point, or similarity ratings were determined from six runs involving two different random noise assignments at every orientation. Further similarity rating data were obtained in most conditions (in a different order and with new random noise assignments) to clean up the data; in total, similarity ratings were based on an average of 6 to 10 runs at each orientation, with two to four different random noise assignments.

After testing of all other orientations, 0° and 180° were retested in case of any overall practice effects or changes in patterns with additional practice after Experiment 1. For the better likeness task, at least one additional QUEST run was taken at each point; for the similarity rating task, at least four additional runs were taken (with new random noise). These data were averaged with those collected in Experiment 1.

The testing times required to collect the new data for Experiment 3 were approximately 16 hr for E.M. in the better likeness task and 6 hr for each of the three subjects who performed the similarity rating task. Testing was spread over approximately 10 days.

### *Results and Discussion*

Figure 13 shows JNDs in the better likeness task for one subject, and similarity ratings for three subjects, as a function of the amount of clockwise rotation of the stimulus from upright. Predicted category boundaries determined from the classification task are also shown. Results of sigmoid-derivative fits to each data set, provided in the Appendix, supported the interpretations described subsequently.

### *CP and Image Plane Rotation*

In all cases, CP disappeared well before the face became fully inverted, as can be seen by scanning down each panel of Figure 13 in turn. For subject E.M. (on both female fades and male morphs), statistically significant CP was present with up to 45° rotation from upright. For male morphs, E.M. showed no CP beyond this point; for female fades, there was ambiguous evidence for CP at 67.5° (a nonsignificant but quite substantial 26% additional variance explained by the sigmoid-derivative fit) and no CP at 90° and beyond. For subjects P.M. (female fades) and A.D. (male morphs), CP endured until 90° but had disappeared by 112.5°.

The disappearance of CP cannot be attributed simply to a failure of subjects to use the similarity rating scale at more extreme rotations. Figure 14 shows mean ratings for pairs differing by 0%, 20%, and 40% of the continuum, averaged across all conditions for which a subject did not show CP. As can be seen, ratings of objective differences in the stimuli were sensitive enough for CP to have emerged had it been present.

Note that from the data presented in Figure 13, we do not wish to say anything about exact tuning curves for configural processing. To do so, our data would need to provide a direct comparison of the magnitude of the CP effect (e.g., the exact height of the cross-boundary peak) at different orientations. In fact, our data were suitable for comparing discriminability only within each orientation in turn, as a result of the blocked testing of orientation. In the similarity rating task, subjects tended to rescale their judgments to use a similar range of the scale at each orientation in turn, making scores in this task unsuitable for comparisons of perceived similarity across orientation conditions.[5] In the better likeness task, there was an overall practice effect (i.e., smaller JNDs) combined with a ceiling effect on performance across the category boundary, which together produced a spurious flattening of curves at the 45° and 22.5° orientations (which were tested later) as compared with upright (which was tested first). Indeed, the 0° data collected at the end of the experiment did not look very different from those shown for 45°. Thus, the nature of our data limits us to drawing conclusions regarding simply the presence or absence of CP at each orientation rather than its exact magnitude.

### *Upright and Inverted at the End of Experiment 3*

The data shown in Figure 13 for the 0° and 180° conditions represent an average of runs taken for Experiment 1 and runs taken at the end of Experiment 3, after substantially more practice with the target continuum. Note that the pattern of data from the end of Experiment 3 did not differ in any meaningful way from that collected in Experiment 1. Most important, there was no sign of any development of CP for inverted faces.

### *Summary*

Using the presence of CP in noise as an index of the existence of configural processing, the results of Experiment 3 show that configural processing of faces survives rotation in the picture plane to somewhere between 45° and 90° from upright. This is a new result. The orientation tuning of configural processing could not be assessed through previous methods (such as identification accuracy, as in Rock, 1974, or famous–nonfamous reaction times, as in Valentine & Bruce, 1988) that confounded, in an unknown manner, contributions of feature-based and configuration-based identification to performance at different orientations.

Although limited to one noise level and suggesting some individual differences, our data provide the first look in normal subjects at how closely tuned an isolated configural processing mechanism is to upright. Interestingly, our results agree with those obtained from patient C.K. (Moscovitch et al., 1997), who demonstrated an organic isolation of configural processing following brain injury. With informal testing in which a famous face was first presented inverted and then rotated until C.K. could identify it, Moscovitch et al. (1997) also found that "a deviation between 45° and 90° from the upright is sufficient to prevent a face from engaging face-recognition mechanisms" (p. 599; M. Moscovitch, personal communication, November 1998). Thus, converging evidence argues that configural processing can operate with a remarkably high degree of mismatch between the stimulus and the internal "prototype" centered on upright. It appears, however, that configural processing is not possible beyond 90°, even after thousands of trials with the same inverted face stimuli.

Finally, our result provides an important constraint on theoretical models of configural face processing, in that any viable model must be able to tolerate an orientation mismatch between the stimulus and the stored representation of up to 90° but no more than 90°. Note that this constraint is more relevant to future model development than to current theories, given that current theories have little, if anything, to say about the orientation tuning of configural processing. Most theories (e.g., Rhodes et al., 1993; Tanaka & Farah, 1993) are phrased in descriptive terms only (e.g.,

---

[5] We disagree here with the approach of Stevenage (1998) and Livingston et al. (1998), who compared actual similarity rating scores (rather than just the pattern of ratings) across different blocked conditions. Both of these studies made claims about whether learned CP reflects expansion of differences across the category boundary or compression of differences away from the boundary by comparing absolute pretraining and posttraining ratings. Our argument is that subjects probably rescale all of their ratings to match the range of perceived similarities in the stimulus set with which they are presented, and thus no direct comparison of ratings is possible.

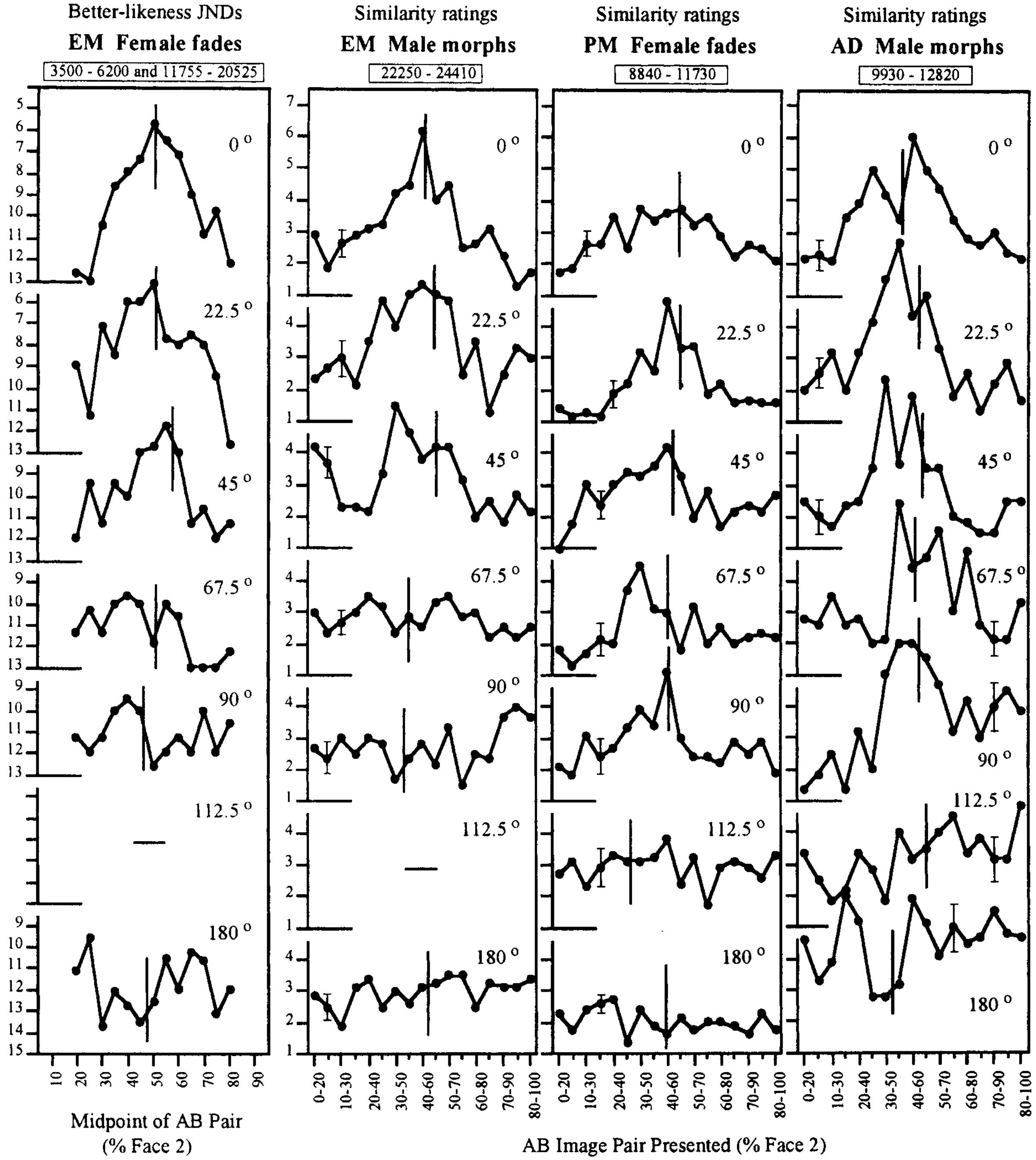


*Figure 13.* Experiment 3: Categorical perception as a function of rotation away from upright in the image plane for whole faces in high noise. Data are shown for 1 subject in the better likeness task (E.M.) and for 3 subjects in the similarity rating task (E.M., P.M., and A.D.). Rectangles show number of trials at the beginning and end of data collection specifically for Experiment 3. Note that for E.M. on female fades, data were collected in two phases (classification followed by better likeness). Error bars show ±1 average *SEM*. The $y$-axes show just-noticeable AB differences (% Face 2) in the better likeness task, or similarity rating scores (1 = most similar; 9 = least similar) in the similarity rating task. JNDs = just-noticeable differences.

a simple statement that inverted faces do not activate the configural processing mechanism) and do not have the computational specificity required to make predictions about orientation tuning. Although various computational approaches have begun to appear in the psychological literature (particularly principal-components analysis; Hancock, Bruce, & Burton, 1998; O'Toole et al., 1994), they often do not explain even the most basic behavioral data, including the inversion effect itself. The principal-components

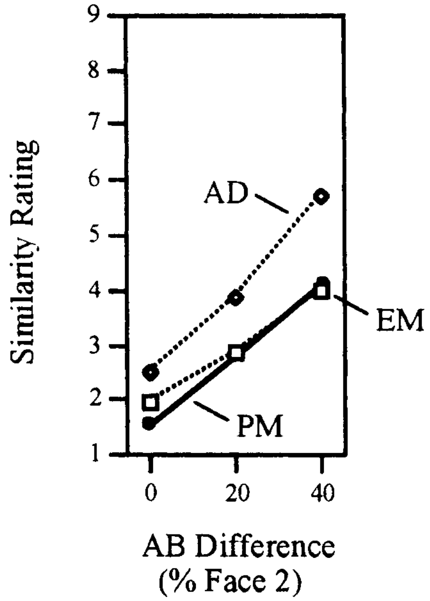


*Figure 14.* Experiment 3: Mean similarity ratings for all images differing by 0%, 20%, and 40% Face 2, averaged across orientations that did not show categorical perception in the 20% difference data (i.e., 67.5°, 90°, and 180° for subject E.M.; 112.5° and 180° for subjects P.M. and A.D.).

analysis approach, for example, requires prealignment of the image to upright before processing can begin, and there is no reason why this (unspecified) image alignment procedure should fail at any particular degree of misorientation.

## General Discussion

In our experiments, faces were presented with external and nonface features removed (e.g., no hair, no spectacles) and in noise designed to make information from any single internal face feature (e.g., eye, nose, mouth, eyebrow) or more local region (e.g., wrinkle, patch of skin texture) an unreliable cue to the identity of the face. Experiment 1 showed that, even under these conditions, subjects familiar with the endpoint faces could perceptually distort a continuum between these faces to form a categorical perception of identity. This was possible, however, only when the faces were presented upright. When the same faces were inverted (Experiment 1) or a single isolated feature was shown (Experiment 2), no CP was found for any subject in any task, even with very large amounts of practice spread over several months. Thus, we have demonstrated a qualitative difference between upright and inverted face processing whereby a particular phenomenon exists *only* for upright whole faces. This shows that configural processing can operate in isolation from feature-based processing, not only after brain injury (Moscovitch et al., 1997) but in normal subjects as well. In addition, we have provided one example (Experiment 3) of using CP in noise as a technique to study the nature of configural representations unconfounded from nonconfigural influences.

We now consider the implications of our results for several issues arising from the previous face recognition literature. This is followed by a discussion of exactly how it is that CP in noise "works" to isolate configural processing.

### *Interpretation of Previous Face CP Studies*

Several previous studies have investigated CP for faces, for either identity (Beale & Keil, 1995; Levin & Beale, 2000; Stevenage, 1998) or expression (Calder, Young, Perrett, Etcoff, & Rowland, 1996; de Gelder, Teunisse, & Benson, 1997; Etcoff & Magee, 1992; Young et al., 1997). With the exception of the Stevenage (1998) study, all involved a set of morphs with no noise. Our results have shown that, under these circumstances, subjects are able to learn CP based on even a single local feature. This implies that, as with other standard techniques used to study face recognition, CP for face morphs without noise can confound influences of both configural and more local information. Thus, the presence of CP for "faces" in these studies may not necessarily indicate anything about *face* processing per se.

To take one example from the previous literature in detail, de Gelder et al. (1997) examined CP for facial expression (e.g., happy–sad). They tested inverted faces as a "control" to demonstrate that the morphing procedure had produced linear changes in the continuum, based on the argument that expression is not perceived in inverted faces. (With expression, unlike identity, it is not possible to use similarity ratings from untrained subjects to confirm morphing linearity, because all people are familiar with the endpoint expressions.) Their results were rather mixed. CP was statistically significant for upright faces but not for inverted faces. However, the actual magnitude of the CP effect was nearly as large for inverted faces as for upright faces (and no significance tests comparing upright and inverted were presented). Thus, it may well be that CP was possible even when faces were inverted. By the logic of de Gelder et al. (1997), such a result could only be interpreted to mean either that the morphing produced nonlinear changes or that expression can be perceived in inverted faces. Neither of these would be desirable conclusions. By our argument, however, CP for inverted "expression" could simply reflect CP for local shape information. This would carry no theoretical implication of expression perception in inverted faces, although it does indicate that inverted faces (without noise) do not provide a useful control for morphing artifacts in studies of CP for facial expression.

### *Qualitative Differences Between Upright and Inverted Face Processing*

As reviewed by Valentine (1988), early studies of face processing did not produce particularly convincing evidence of qualitative differences between the processing of upright and inverted faces, despite Yin's (1969) demonstration of the unusually large effects of inversion on recognition memory for faces. This led Valentine (1988, 1991) to argue that upright and inverted faces are, in fact, processed in qualitatively similar ways. In the past 15 years, far more convincing evidence of qualitatively different modes of processing for upright and inverted faces has been produced (e.g., Bartlett & Searcy, 1993; Leder & Bruce, 1998; Martini & Nakayama, 1999; Rhodes et al., 1993; Tanaka & Farah, 1993; Young et al., 1987). Our results add to this growing literature.

### *Expertise and the Lack of Configural Processing for Inverted Faces*

Whereas many authors have argued that face processing is "special" and distinct from the processing of objects and face parts (e.g., Farah et al., 1998; Moscovitch et al., 1997), others have argued that people are simply more expert with faces than other objects. It has been argued (e.g., Diamond & Carey, 1986; Gauthier & Tarr, 1997) that configural processing will occur for nonface objects as long as (a) the task requires individuation of members of a class all sharing a common basic configuration (e.g., recognizing an individual Scotch terrier among many Scotch terriers) and (b) the subject is an expert in that stimulus domain. It has also been reported that this expert recognition involves the same brain areas as does face recognition (Gauthier, Skudlarski, Gore, & Anderson, 2000). If this interpretation of specialized face effects as expertise effects is correct, then it might be expected that any stimulus domain meeting the requirements just described would have the potential to show configural processing, as long as the subject is given sufficient exposure to that domain.

One such stimulus domain is inverted faces. Our results argue, however, that it is not possible to learn configural representations of inverted faces, at least with the type of practice we used. This was the case even with thousands or tens of thousands of trials. The failure to learn configural processing for inverted faces has at least two possible theoretical interpretations, both interesting from the point of view of the relationship between face and object processing. One interpretation is that faces may be genuinely special and may be the only stimulus class for which configural representations in the inverted form cannot be learned; this could arise, for example, if the need for innate perception of expression (e.g., see Ekman, 1994) required an innate coding of the upright form of faces that cannot be overcome. Alternatively, it may be that configural processing of the inverted form can never be learned for any stimulus once the upright form has been overlearned. The first of these proposals predicts that it should be possible for, say, bird experts to learn configural processing of inverted birds, whereas the second predicts that this would not be possible. We suggest that exploring expertise effects on configural processing for noncanonical forms would be a fruitful line of future research.

### *Relationship Between Featural and Configural Processing*

The previous literature on configural face processing has focused almost entirely on showing that configural processing can be selectively disrupted while feature processing remains intact. Single dissociations such as this can always be argued simply to reflect differential difficulty, with the more difficult processing being more susceptible to brain damage, memory limitations, and so on. Our present results, consistent with Moscovitch et al.'s (1997) patient C.K., demonstrate the reverse dissociation. In fact, our results show that configural processing can be *easier* than processing of single features, in that more noise can be "seen through" to make fine discriminations of identity. We have argued that there is a clear theoretical reason for such a mechanism in face recognition: In the real world, variation from one image to the next is so great that reliable identification of individuals is not possible by, for example, the nose alone. Instead, a more sensitive mechanism is needed, namely one that integrates information across the entire face region (see also Martini & Nakayama, 1999) rather than merely operating on local regions in a piecemeal fashion.

Given that it is possible for a configural processor to operate in a manner functionally independent from feature-based identification, how can this be understood in terms of underlying mechanisms? Perhaps the most natural interpretation (e.g., Farah et al., 1998; Moscovitch et al., 1997) is to assume two parallel systems that independently scan the visual scene: a part-based "object recognition" system that processes objects as well as isolated face features and a (more sensitive) "face recognition" system searching for whole faces. In this view, configural face representations are intrinsically holistic and are not parsed into feature-based parts, although they still code detailed information about all regions of the face to allow discrimination of individual faces (e.g., Hancock et al., 1996; O'Toole et al., 1994; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). This proposal sits perhaps most easily with the double dissociation in the patient data between the processing of objects and faces (e.g., see Moscovitch et al., 1997, p. 557, for a review) and is also compatible with our own data.

An alternative view is of a hierarchical arrangement in which features and groups of features lead to configurations, and face features form a distinct level of representation. Such hierarchical processing has been proposed in a general-purpose pattern recognizer by Perrett and Oram (1998). Initial parsing of faces into feature-based parts is also implied by theories of face recognition in which configural representations code higher order spatial relationships among feature components (e.g., Rhodes, 1988) or compare the spatial arrangement of feature components with a norm representing the average upright face (e.g., Diamond & Carey, 1986). The hierarchical approach might appear to fit less immediately with the finding that configural processing can operate in isolation, because this implies that configural processing can be "above threshold" when performance for each component part leading into the configuration is "below threshold." Note, however, that such an idea might be more reasonable than it first appears: It could be that configural processing is more sensitive than would be predicted from merely a collection of unrelated parts, simply because the configural representation integrates information over a larger spatial area than does each single feature representation (i.e., the "receptive field" is larger). Overall, then, the isolation of configural processing is potentially consistent with either of two views—"two systems" or "different stages in a single hierarchy"—regarding the relationship between featural and configural processing.

### *Exact Nature of Configural Representations*

The previous literature has demonstrated that configural face representations are *holistic* (i.e., integrate information over the entire internal face region) and are *tuned to upright.* Although these two basic properties are accepted by most researchers, little is known about the more detailed nature of configural representations. An important factor in this has been limitations in the techniques available, in which configural processing is examined only indirectly (i.e., through what aspects of performance break down without it). With the technique introduced in the present article, the nature of configural processing can now be examined directly in normal subjects.

Our results so far delineate two additional properties of configural representations. The first regards the extent to which configural face representations code detailed information about local parts of the face (e.g., exact feature shape). Views in which configural processing is based on spatial relationships among feature components (e.g., Diamond & Carey, 1986; Rhodes, 1988) can be read as suggesting that configural information consists only, for example, of a collection of numbers coding distances between abstract markers representing the locations of the key features; in this interpretation, all detail about the appearance of individual face parts is lost. In contrast, several authors have proposed that the configural representation includes *all* information about a face, including details of individual feature shape. Two sources of prior data support the latter of these views in showing that individual feature processing is better in the context of a learned face. Tanaka and Sengco (1997; see also Tanaka & Farah, 1993) showed that forced-choice recognition memory for a face part—"choose the old nose"—was most accurate when the two noses were each presented as part of the original studied face and least accurate when the noses were presented alone. Moscovitch et al. (1997) presented famous faces with a single part missing and required subjects to choose the correct part (i.e., the correct nose) of two alternatives; both patient C.K. and normal control subjects were able to choose the correct part reliably when they identified the face. This above-chance performance was not found either for inverted faces or for upright faces that the subject failed to identify.

Our own results provide further support for the view that configural representations include detailed information about face features. CP between highly similar individuals surely requires reference to detail (particularly to discriminate, say, 45% and 55% morphs), and yet it is exactly this information that was disrupted in the image by our noise. Thus, we suggest that subjects must have used their representations of Face 1 and Face 2, stored in memory, to "fill in" the missing detail necessary to support CP. Crucially, because our results showed that this was possible only for upright faces, we argue that it was specifically the *configural* representation of each endpoint face that coded the relevant knowledge of detail. Thus, we suggest that configural representations do not simply code distances between feature positions but instead code all levels of information about a face, including exact feature shape and the spatial relationships among features. (Note this view would appear to be more consistent with the "holistic" approaches discussed in the preceding section than with approaches assuming initial parsing into feature components.)

The second property of configural representations demonstrated by our results is that configural processing operates with up to 45°–90° of image plane rotation from the upright, but not beyond this point. Comfortingly, this result agrees with that demonstrated by patient C.K., who showed an organic, rather than experimental, isolation of configural processing.

Our finding regarding rotation effects was obtained by examining where the signature phenomenon of configural processing dropped out. In future research, many further questions about the exact nature of configural processing can be addressed through a similar method. Interesting issues might include, for example, the properties of the stimulus necessary to activate configural representations (e.g., how much of the face is needed, whether forming a perceived surface is essential, and how much the face can be warped or split apart) as well as more general issues (e.g., whether configural representations are race specific, whether there are individual differences in configural processing ability in normal subjects, and how configural processing develops across childhood). Many such questions have, of course, received considerable attention in the literature; in most cases, however, measures have been used that confound performance based on configural processing with that based on identification of local features (e.g., recognition memory for clear pictures of full heads). Tracking the magnitude of CP for faces in noise, in contrast, would allow the effects of size, warping, childhood development, and so on to be determined specifically for *configural* processing.

### *Exactly How Does CP in Noise Isolate Configural Processing?*

In our experiments, we found that adding noise was necessary to isolate configural processing in the CP task. We now consider the role of this noise in more detail. There are two reasons for doing so: First, we wish to dispel a common misinterpretation of our purpose in adding the noise, and, second, we wish to clarify the circumstances necessary to ensure the isolation of configural processing in future studies.

Among researchers with some exposure to "perceptual" (as opposed to "cognitive") approaches to face recognition, there appears to be a rather common myth that associates "configural processing" with "low spatial frequency information." This myth seems to have arisen from a misinterpretation of early studies (e.g., Fiorentini, Maffei, & Sandini, 1983; Sergent, 1985) on spatial frequency contributions to face recognition. These studies are often cited (at least informally) as showing that face recognition relies primarily on low spatial frequency components of the image. In fact, this is not the case. First, although it has been demonstrated that low spatial frequencies can provide reasonably reliable face identification when higher spatial frequencies are removed (e.g., Costen, Parker, & Craw, 1994; Sergent, 1985), Fiorentini et al. (1983) also tested high-pass faces and found equally reliable identification based on *higher* frequencies when low frequencies were removed. Second, the studies just cited used as stimuli only a limited set of faces with hair; thus, the low spatial frequency contribution might reflect hair recognition as much as face recognition. Third, Sergent (1985) demonstrated that high spatial frequencies were not necessary in a male–female discrimination task but were very important in a task requiring discrimination of individuals within one sex; thus, she argued that although gross discriminations can be made with low spatial frequencies, finer discriminations of identity require high spatial frequencies. Fourth, to our knowledge, no one has directly tested whether performance with low spatial frequency components reflects *configural* processing by testing inverted faces. Finally, the arguments put forward in the preceding section make it clear that there are several reasons to believe that configural representations code full details of faces, not just gross-scale information.

Thus, we wish to emphasize that our rationale for adding noise to the stimuli was not to isolate configural processing by removing high spatial frequencies (although the type of noise we added did, of course, disrupt primarily these frequencies). Indeed, had this been our aim, we could have low passed our face stimuli and left out the noise. We suspect, however, that constant stimuli such as these would not have isolated configural processing; subjects would be able, with sufficient practice, to learn to take advantage of small differences between the intermediate morph images, because these differences would be consistent across all presentations of a given morph.

Our presumption is that the critical aspect of adding noise was to make local information *variable*. From previous studies, we know that CP is possible (without noise) for a wide range of image properties. These include low-level properties such as textures (Pevtzow & Harnad, 1997) and brightness–saturation combinations (Goldstone, 1994), and also, more relevantly, shape information such as that distinguishing the cloacae of male and female chickens or two individual, line-drawn, microorganismlike objects (Livingston et al., 1998). We therefore presume that, for faces, subjects will use local shape information to support CP if it is possible to do so. Our proposal, then, is that the key property of our heavy noise was that different random noise assignments made local shape information an unreliable indicator of position with respect to the category boundary. This claim is supported by the illustration in Figure 2 that, under some high noise assignments, even the 40% and 60% morphs could appear on the wrong side of a 50% boundary if the nose region is considered alone (note that this was not true when the low noise level was used). It is also consistent with the evidence from Experiment 2 that CP was not possible for the nose alone when it was presented in high noise.[6]

Finally, we emphasize that there should be other ways of forcing CP to rely entirely on configural information that do not involve adding noise. Our noise was merely an easy-to-implement analogy for the type of image variability across different views, expressions, and so on of the same person that occurs in the real world. It is possible to include this variation in the stimulus set directly, as was done by Stevenage (1998). Her stimuli were multiple, different photographs of identical twins taken with different poses and expressions. By our logic, such a stimulus set should create a situation in which any one local region of an image provides an unreliable indicator of identity, thus forcing configural integration over the whole face. Unfortunately, Stevenage did not test inverted faces or isolated face parts, and so although we suspect that her stimulus set might have isolated configural processing in the CP task, it is not possible to confirm this empirically.

## *Conclusion*

Most experimental methods of examining face processing potentially confound performance based on a configural representation of the whole face with performance based on lower level information, such as identification of local features. To the extent that researchers have attempted to disentangle these two components in normal subjects, they have usually done so by trying to disable configural processing while leaving feature processing intact, for example by inverting the face or separating the face into parts. This approach can be seen to be analogous to that of studying prosopagnosia following brain injury; that is, in both cases, the interest is in what aspects of performance break down without configural processing. Although this approach has demonstrated that configural processing exists and plays an important role in face identification, it has not as yet allowed the exact nature of the configural representations of upright faces to be determined. Moscovitch et al. (1997) argued for the importance of approaching configural processing from the other direction as well, that is, investigating what can be done through configural processing alone. Their study of patient C.K. provides an example of this alternative approach in a case of organic isolation of configural processing following brain injury. The technique we have described in this article provides an implementation of the same idea in normal subjects. In examining the signature phenomenon of categorical perception of face identity in noise, we have investigated face recognition performance in a situation in which feature-based identification has been disabled while configural identification is left intact. It is our hope that this approach will allow researchers to look more directly at the nature of configural face representations than has been possible to date.

---

[6] This implies that the type of noise selected should match the particular stimuli used. With our low-contrast, detailed gray scale photographs, we chose "snowlike" noise. Line drawings of faces, for example, contain much stronger local contour information than photographs; for these stimuli, it might be necessary to try random curved line segments as noise. This also applies to objects. In principle, CP in noise could be used to compare configural processing for nonface objects with that for faces. However, objects such as cars, dogs, and "greebles" are largely defined by the shape of high-contrast external contours. Such contours, especially if they are long and curve only slowly, will remain visible and distinctive through the types of noise we have used. Thus, adding snowlike noise is unlikely to disable CP based on nonconfigural cues, either for line drawings of faces or for most nonface objects.

## References

Adobe Photoshop (Version 4.0) [Computer software]. (1997). San Jose, CA: Adobe Systems Incorporated.

Angeli, A., & Gerbino, W. (1998). A disproportionate effect of inversion in cross-category morphs of familiar faces. *Perception, 27*(Suppl.), 126.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25,* 281–316.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57,* 217–239.

Behrman, M., Winocur, G., & Moscovitch, M. (1994). Intact visual imagery and impaired visual perception in a patient with visual agnosia. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 1068–1087.

Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 535–565). New York: Cambridge University Press.

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

Bruyer, R., & Coget, M. C. (1987). Features of laterally displayed faces: Saliency or top-down processing? *Acta Psychologica, 66,* 103–114.

Burns, E. M., & Ward, W. D. (1978). Categorical perception—Phenomenon or epiphenomenon: Evidence from the experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63,* 456–468.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3,* 81–117.

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

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments and Computers, 25,* 257–271.

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception, 23,* 129–146.

de Gelder, B., Teunisse, J.-P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion, 11,* 1–23.

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

Ehret, G. (1987). Categorical perception of sound signals: Facts and hypotheses from animal studies. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 301–331). New York: Cambridge University Press.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4,* 99–109.

Ekman, P. (1994). Strong evidence for universals in facial expression: A reply to Russell's mistaken critique. *Psychological Bulletin, 115,* 268–287.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expression. *Cognition, 44,* 227–240.

Farah, M. J. (1996). Is face recognition "special"? Evidence from neuropsychology. *Behavioral Brain Research, 76,* 181–189.

Farah, M. J., Tanaka, J. N., & Drain, M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 628–634.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research, 35,* 2089–2093.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105,* 482–498.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception, 12,* 195–201.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3,* 191–197.

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

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123,* 178–200.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research, 38,* 2277–2288.

Hancock, P. J. B., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principle components analysis. *Memory & Cognition, 24,* 26–40.

Harnad, S. (1987a). Category induction and representation. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 535–565). New York: Cambridge University Press.

Harnad, S. (1987b). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1–25). New York: Cambridge University Press.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51*(A), 449–473.

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, cross-race faces, and inverted faces. *Perception & Psychophysics, 62,* 386–401.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358–368.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 732–753.

Martini, P., & Nakayama, K. (1999). Spatial summation of facial images reveals holistic processing. *Investigative Ophthalmology and Vision Science, 40,* S389.

Morph (Version 1.0) [Computer software]. (1992). San Diego, CA: Gryphon Software.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9,* 555–604.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping and the recognition of occluded objects. *Perception, 18,* 55–68.

O'Toole, A. J., Deffenbacher, K. A., Valetin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other race effect. *Memory & Cognition, 22,* 208–224.

Perrett, D. I., & Oram, M. W. (1998). Visual recognition based on temporal cortex cells: Viewer-centred processing of pattern configuration. *Zeitschrift fur Naturforschung, 53C,* 518–541.

Pevtzow, R., & Harnad, S. (1997). Warping similarity space in category learning by human subjects: The role of task difficulty. In M. Ramscar, U. Hahn, E. Cambouropolos, & H. Pain (Eds.), *Proceedings of SimCat 1997: Interdisciplinary workshop on similarity and categorization* (pp. 189–195). Edinburgh, Scotland: Department of Artificial Intelligence, Edinburgh University.

Postma, P. (1998, July). *Developmental prosopagnosia: Should it be taken at face value?* Paper presented at the meeting of the Experimental Psychology Society, York, England.

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

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

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

Rock, I. (1974). The perception of disoriented figures. *Scientific American, 230,* 78–85.

Rumiati, R. I., Humphreys, G. W., Riddoch, J. M., & Bateman, A. (1994). Visual object agnosia without prosopagnosia or alexia: Evidence for hierarchical theories of visual recognition. *Visual Cognition, 1,* 181–225.

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

Sergent, J. (1985). Influence of task and input factors on hemispheric involvement in face processing. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 846–861.

Sinha, P., & Poggio, T. (1996, December 5). I think I know that face . . . . *Nature, 384,* 404.

Stevenage, S. V. (1998). Which twin are you? A demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology, 89,* 39–57.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46*(A), 225–245.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25,* 583–592.

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception, 9,* 483–484.

Tong, F., & Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1016–1035.

Valentine, T. (1988). Upside-down faces: A review of the effects of inversion upon face recognition. *British Journal of Psychology, 79,* 471–491.

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

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Memory & Cognition, 16,* 556–566.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113–120.

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

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

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63,* 271–313.

(*Appendix follows*)

# Appendix

## Discrimination Curve Fit Results

Tables A1–A3 show the following: the sign of $h$ (+ indicates peak; − indicates dip), the proportion of variance explained by a sigmoid-derivative-plus-linear fit ($R^2_{\text{sd+lin}}$) and by a linear-only fit ($R^2_{\text{lin}}$), and $F$ values for significance of the improvement in $R^2$ values with the addition of the sigmoid-derivative component. See the text for interpretation of these values as evidence for the presence or absence of categorical perception. B-L = better likeness task; Sim = similarity rating task; ff = female fades; mm = male morphs; sfm = smiling female morphs. For the $F$ test of the difference between $R^2_{\text{sd+lin}}$ and $R^2_{\text{lin}}$, $df = 3, 8$ for better likeness and $df = 3, 12$ for similarity ratings.

Table A1

*Experiment 1: Whole Faces, High Noise*

| | Upright | | | | Inverted | | | |
|---|---|---|---|---|---|---|---|---|
| Group | $h$ | $R^2_{\text{sd+lin}}$ | $R^2_{\text{lin}}$ | $F$ | $h$ | $R^2_{\text{sd+lin}}$ | $R^2_{\text{lin}}$ | $F$ |
| Trained subjects | | | | | | | | |
| E.M., ff, B-L | + | .81 | .01 | 11.23** | − | .02 | .02 | <1 |
| P.M., ff, B-L | + | .96 | .00 | 64.00** | − | .18 | .18 | <1 |
| A.D., mm, B-L | + | .73 | .23 | 4.94* | − | .58 | .08 | 3.17 |
| A.E.S., mm, B-L | + | .79 | .46 | 4.17* | + | .26 | .26 | <1 |
| E.M., mm, B-L | + | .83 | .04 | 12.39** | − | .20 | .20 | <1 |
| E.M., ff, Sim | + | .82 | .00 | 18.22** | + | .11 | .11 | <1 |
| P.M., ff, Sim | + | .77 | .22 | 9.57** | − | .22 | .16 | <1 |
| A.D., mm, Sim | + | .76 | .01 | 12.5** | − | .00 | .00 | <1 |
| A.E.S., mm, Sim | + | .88 | .36 | 17.33** | − | .47 | .42 | <1 |
| E.M., mm, Sim | + | .76 | .23 | 8.83** | + | .11 | .11 | <1 |
| P.M., sfm, Sim | + | .63 | .20 | 4.64* | − | .00 | .00 | <1 |
| Untrained controls | | | | | | | | |
| A.H., ff, Sim | − | .72 | .72 | <1 | − | .07 | .07 | <1 |
| V.M., ff, Sim | − | .18 | .18 | <1 | + | .00 | .00 | <1 |
| F.M., ff, Sim | + | .00 | .00 | <1 | − | .08 | .08 | <1 |
| M.A., mm, Sim | − | .17 | .10 | <1 | + | .00 | .00 | <1 |
| D.W., mm, Sim | + | .18 | .18 | <1 | − | .01 | .01 | <1 |
| P.G., mm, Sim | + | .00 | .00 | <1 | + | .08 | .01 | <1 |
| M.M., sfm, Sim | + | .41 | .41 | <1 | − | .04 | .04 | <1 |

*$p < .05$. **$p < .01$.

Table A2

*Experiment 2: Faces Versus Features*

| | Upright | | | | Inverted | | | |
|---|---|---|---|---|---|---|---|---|
| Condition | $h$ | $R^2_{\text{sd+lin}}$ | $R^2_{\text{lin}}$ | $F$ | $h$ | $R^2_{\text{sd+lin}}$ | $R^2_{\text{lin}}$ | $F$ |
| Face, high noise | | | | | | | | |
| E.M., mm, B-L | + | .83 | .04 | 12.39** | − | .20 | .20 | <1 |
| E.M., mm, Sim | + | .82 | .04 | 17.33** | + | .26 | .25 | <1 |
| A.D., mm, Sim | + | .74 | .00 | 11.38** | − | .08 | .08 | <1 |
| P.M., sfm, Sim | + | .63 | .20 | 4.64* | − | .00 | .00 | <1 |
| Face, low noise | | | | | | | | |
| E.M., mm, B-L | + | .76 | .00 | 8.44** | + | .61 | .00 | 4.17* |
| E.M., mm, Sim | + | .81 | .33 | 10.11** | + | .71 | .01 | 9.66** |
| A.D., mm, Sim | + | .75 | .06 | 11.04** | + | .73 | .08 | 9.63** |
| P.M., sfm, Sim | + | .67 | .00 | 8.12** | + | .60 | .01 | 5.90* |
| Nose, low noise | | | | | | | | |
| E.M., mm, B-L | + | .63 | .01 | 4.96* | + | .77 | .00 | 8.93** |
| E.M., mm, Sim | + | .85 | .04 | 21.60** | + | .90 | .00 | 36.00** |
| A.D., mm, Sim | + | .84 | .21 | 15.75** | + | .80 | .07 | 14.60** |
| P.M., sfm, Sim | + | .57 | .10 | 4.37* | + | .72 | .25 | 6.71** |
| Nose, high noise | | | | | | | | |
| E.M., mm, B-L | − | .80 | .12 | 9.07** | − | .69 | .20 | 4.21** |
| E.M., mm, Sim | − | .09 | .09 | <1 | − | .00 | .00 | <1 |
| A.D., mm, Sim | − | .24 | .24 | <1 | + | .02 | .02 | <1 |
| P.M., sfm, Sim | − | .11 | .11 | <1 | + | .03 | .02 | <1 |

*$p < .05$. **$p < .01$.

Table A3
*Experiment 3: Rotation*

| Orientation (degrees) | $h$ | $R^2_{\text{sd+lin}}$ | $R^2_{\text{lin}}$ | $F$ |
|---|---|---|---|---|
| E.M., ff, B-L | | | | |
| 0 | + | .93 | .03 | 34.29** |
| 22.5 | + | .78 | .04 | 8.97** |
| 45 | + | .80 | .00 | 10.67** |
| 67.5 | + | .65 | .39 | 2.67 |
| 90 | + | .01 | .00 | <1 |
| 180 | − | .00 | .00 | <1 |
| E.M., mm, Sim | | | | |
| 0 | + | .82 | .04 | 17.33** |
| 22.5 | + | .71 | .00 | 9.79** |
| 45 | + | .58 | .11 | 4.48* |
| 67.5 | − | .09 | .09 | <1 |
| 90 | + | .13 | .13 | <1 |
| 180 | + | .26 | .25 | <1 |
| P.M., ff, Sim | | | | |
| 0 | + | .78 | .01 | 14.00** |
| 22.5 | + | .84 | .02 | 20.50** |
| 45 | + | .70 | .00 | 9.33** |
| 67.5 | + | .79 | .01 | 14.86** |
| 90 | + | .62 | .00 | 6.53** |
| 112.5 | + | .00 | .00 | <1 |
| 180 | + | .13 | .13 | <1 |
| A.D., mm, Sim | | | | |
| 0 | + | .74 | .00 | 11.38** |
| 22.5 | + | .85 | .04 | 21.60** |
| 45 | + | .72 | .02 | 10.00** |
| 67.5 | + | .49 | .01 | 3.76* |
| 90 | + | .86 | .25 | 17.43** |
| 112.5 | − | .37 | .37 | <1 |
| 180 | − | .08 | .08 | <1 |

$^{*}p < .05$. $^{**}p < .01$.

Received June 1, 1999
Revision received April 3, 2000
Accepted August 23, 2000 ■