Available online at www.sciencedirect.com

ScienceDirect

Vision Research 47 (2007) 2134–2142

Vision Research

www.elsevier.com/locate/visres

The deleterious effect of contrast reversal on recognition is unique to faces, not objects

Marissa Nederhouser [a], Xiaomin Yue [a], Michael C. Mangini [b], Irving Biederman [a,*]

[a] *Department of Psychology and Neuroscience Program, University of Southern California, 3641 Watt Way, Hedco Neurosciences Building, Los Angeles, CA 90089-2520, USA*

[b] *Department of Psychology, University of Notre Dame, 118 Haggar Hall, Notre Dame, IN 46556, USA*

Received 21 October 2006; received in revised form 29 March 2007

Abstract

Faces reversed in contrast cannot be readily recognized, an effect absent in object recognition. Why? Four factors: expertise, reflectance (pigmentation), high similarity, and the need to discriminate metrically varying smooth surfaces have been offered as explanations. Observers achieved expertise on discriminating smoothly shaped, pigmented, non-face blobs with positive contrast, where distractor similarity matched that of a set of faces in shape and reflectance. On a match-to-sample task, reversal of contrast between sample and matching images had no effect when matching such blobs, but markedly degraded performance when matching faces suggesting that this effect is unique to faces.

Keywords: Face recognition; Contrast reversal; Object recognition; Perceptual expertise; Blob discrimination

1. Introduction

The individuation of faces has posed a unique perceptual problem over our evolutionary history in that few other classifications require that subtle metric distinctions be made over a large number of instances (individuals). Recognizing a face presented in reversed contrast—so that light become dark and dark becomes light—is extremely difficult, whereas no effect is found for common, non-face objects (Galper, 1970; Subramaniam & Biederman, 1997). Could this be a function of stimulus and/or expertise variables that, if applied to objects that do not resemble faces, would render their recognition equally vulnerable to reversed contrast? [We will use the term "contrast reversal" when to-be-matched stimuli differ in their direction of contrast, one being positive and the other being negative. We will use the term "contrast negation" when the to-be-matched stimuli are both of negative contrast. An inability to recognize a familiar face from a negative image implies a comparison between that negative image and the previously experienced face, which will always be positive.] An alternative account is that whereas the representation of objects specifies "moderately complex features," typically based on non-accidental specification of orientation and depth discontinuities (Biederman, 1987; Kobatake & Tanaka, 1994) or surface markings (Biederman, Subramaniam, Bar, Kalocsai, & Fiser, 1999), the neural coding of faces is unique to faces in that it retains the original multiscale, multiorientation filtering (Biederman & Kalocsai, 1997), and this representation is especially vulnerable to contrast reversal. We assume that this representation is translation and scale invariant with scale expressed as cycles per face (within limits) rather than cycles per degree, although how the invariance is achieved is still an unresolved question. A theoretical account of why the representation of faces would uniquely retain aspects of their spatial coding—and thus suffer from reversed contrast—is presented in the Section 6.

---

[*] Corresponding author. Fax: +1 213 740 5687.
E-mail address: bieder@usc.edu (I. Biederman).

0042-6989/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2007.04.007

Faces differ from most subordinate-level object classes that we distinguish in that we are often called to individuate them on the bases of small, metric differences in the magnitudes and curvature of their smooth surfaces, or their variations in reflectance (also termed "pigmentation") but more appropriately considered as reflectance (see Russell, Biederman, Nederhouser, & Sinha, 2007) and shadowing within relatively fixed ordinal relations between different regions (Sinha, 2002). Natural classes with highly similar exemplars, such as the birds found on the same page in a bird guide or different makes of cars, are typically distinguished by small, non-accidental features, such as whether the shape of a taillight or logo is round or square or whether a given species has a dark patch on a light underbelly (Biederman et al., 1999). Discriminating exemplars by non-accidental differences may engage different representations than that for the discrimination of faces, which rarely vary in the presence or absence of non-accidental features, but, instead, in the degree of their surface curvature and size of their protuberances (i.e., nose and cheekbones) and magnitudes of reflectance in local regions.

Subramaniam and Biederman (1997) compared the costs of contrast reversal and contrast negation on the same–different matching of faces and chairs of equal average shape dissimilarity on "different" trials. There was almost no cost of contrast reversal on matching chairs but a massive 20% increase in error rates (chance = 50%) and an 80 ms increase in reaction times when matching faces of opposite contrast polarity. They reported only a slight effect of contrast negation on either class of stimuli. A recent study by Vuong, Peissig, Harrison, and Tarr (2005) compared the effects of reversed contrast on the same–different matching of pigmented faces and Greebles. For faces, reversed contrast led to an average increase of 11.5% in error rates (positive and negative trials combined) which produced a difference in d's of .87 ($d' = 2.24$ for matched contrast and 1.37 for reversed contrast). Although the cost of reversed contrast in terms of d's for Greebles was a moderate .39 ($d' = 2.13$ for matched contrast and 1.74 for reversed contrast) the difference in error rates for positive and negative trials combined was only 1% (Vuong et al., 2005, Table 2). Whether the effect of reversed contrast on pigmented Greebles is interpreted as moderate or negligible depends on which measure is taken as more appropriate. In both of these studies it is possible that the non-face stimuli, chairs or Greebles, differed in small shape features—often non-accidental—that could be readily conveyed by edges marking orientation and depth discontinuities that would be invariant to the direction of contrast and thus could have reduced the effect of contrast reversal. The present study compared the effects of contrast polarity of faces to non-face stimuli designed to require the same kinds of smooth surface processing as faces. Rather than same–different matching, in which response criteria for responding "same" (or "different") complicate the data analysis and interpretation, a match-to-sample paradigm was employed in which one of the two matching stimuli was identical to the sample, thus eliminating the problem of where to set the criterion (for responding "same" or "different"). Another methodological feature of the present investigation was that the physical similarity of the distractor and matching stimuli was scaled so the effects of contrast polarity could be evaluated over overlapping ranges of physical similarity. This scaling of physical similarity (see Russell, Sinha, Biederman, & Nederhouser, 2006; Russell et al., 2007; Yue, Tjan, & Biederman, 2006) provides a potential solution to the "apples and oranges" problem of comparing qualitatively different classes of stimuli in their familiar orientations.

There is now considerable evidence that although both the shape of a face and its pigmentation are employed in recognition (e.g., Russell et al., 2006), the deleterious effects of contrast reversal or negation on face matching is a function of the pigmentation of the face in that non-pigmented faces (Bruce & Langton, 1994) or faces all of the same pigmentation but differing only in shape (Russell et al., 2006), show little or no cost in matching from contrast negation. This is also true for the (non-face) blobs used in the present experiment in that when they are non-pigmented they show no cost of contrast reversal and negation (Nederhouser, Mangini, & Biederman, 2001). In the present study, therefore, both the face and non-face stimuli were pigmented, as they were in Vuong et al. (2005).

Humans are face experts, because of their need to individuate faces earlier and more often in life than any other objects in their environment (Gauthier & Tarr, 1997). This intensive face-training experience is exclusively confined to faces of positive contrast so that, for example, the iris is always experienced as being darker than the cornea. Might the cost of contrast reversal for faces be the result of a kind of "face expertise?" Advocates of an expertise account of face processing have held that the perceptual expertise for distinguishing among members of a class can be achieved by adults in about 3000 trials (Gauthier & Tarr, 1997). They have also argued that it is the expertise developed solely with faces of positive contrast that renders face recognition so vulnerable to contrast reversal (Gauthier, Williams, Tarr, & Tanaka, 1998). Would the disruptive effect of contrast reversal and negation, so evident when distinguishing faces, be witnessed for subjects who had achieved expertise on positively contrasted, non-face stimuli, designed to require the same low-level processing as faces?

2. Methods

2.1. Participants

Sixteen University of Southern California right-handed graduate and undergraduate students, eight females and eight males, ages 19–26 years, participated for monetary compensation. None of the participants had seen the stimuli prior to the experiment. Four students were randomly assigned to the "Expert" condition in which they were trained to become blob "Experts," four were assigned to the "Novice" condition in which they received no training on positive blobs prior to being tested on contrast reversed blobs, and eight were assigned to the "Faces" condition.

2.2. Stimuli

2.2.1. Generation of blobs

Stimulus generation was inspired by Shepard and Cermak's (1973) production of a set of complex, asymmetric, 2D novel shapes formed by adding different orientations of the 2nd and 3rd harmonics of a circle to the circle and 4th harmonic. We created 3D versions of these shapes by adding different orientations of the 2nd and 3rd harmonics of a sphere to a sphere and the 4th harmonic of a sphere using Matlab (Fig. 1). The second and third harmonics, 3D shapes with either two or six equally spaced convex lobes, respectively, were added together in one of eight equally spaced orientations, producing a toroidal shaped stimulus space (Fig. 2a). Matlab code for generating the blobs can be found at http://geon.usc.edu/blobcode. The space is toroidal, as was Shepard and Cermak's, in that it curves around on itself so that the top and bottom rows are identical and the left and right columns are identical so there are no boundaries delineating the edges of the space (which could produce enhanced discriminabilities with the absence of near neighbors on one side). Four of these blobs (circled, Fig. 2a, with maximal diagonal distances for a given pair) were selected to form the seeds of the experimental stimuli in which the amplitudes of the 2nd and 3rd harmonics were varied along eight different values (Fig. 2b), so that the stimuli mimicked the way in which faces systematically vary in the magnitude of their protuberances within the same general shape, i.e., all have cheekbones, a nose, etc., in approximately fixed positions and orientations.

The surface contrast variation of the base sphere was specified by an albedo patch derived from a face (described below). Once each blob was specified, it was rendered as illuminated from two point-source lights in front (set at infinity to prevent cast shadows), at 45° above the x-axis, and 90° apart on the y-axis, and its material was set as dull so that the blob would reflect more diffuse light with no specular highlights. The rendered images of the objects were then converted into 8-bit grayscale images at 72 dpi using Adobe Photoshop 5.0 and presented on an Apple Macintosh G3 computer at a resolution of 1024×768 pixels at a refresh rate of 75 Hz.

Even though the blobs had the smooth shape characteristic of faces with face-derived albedo variation and a range of stimulus similarity that matched the set of faces employed in the experiment, their asymmetry and lack of configural resemblance to faces do not invite a face-like interpretation. These characteristics render the blobs a preferred non-face control stimulus set with which to investigate processes that may or may not be unique to face processing,

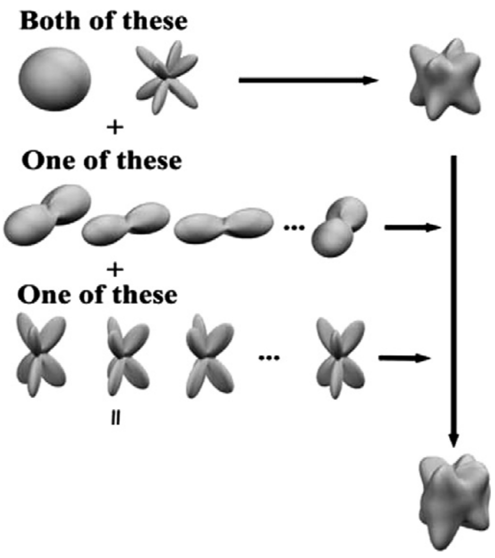

Fig. 1. Generation of the blob stimuli. Blobs were generated by combining the 2nd and 3rd harmonics of a sphere in eight different orientations.

2.2.2. Reflectance variation

Faces also vary in their surface lightness or darkness. Some of this variation is produced by differences in pigmentation (e.g., iris and cornea) and some from shadows (e.g., around eye sockets or in the nostrils). We will refer to all such variation of lightness as reflectance. Eight images of faces served as input to an algorithm (Portilla & Simoncelli, 2000) that "texturized" the images by markedly altering the higher order structure, so the images were no longer recognizable as faces, but preserving the image statistics (luminance, contrast, skewness, kurtosis) and wavelet subbands (space, orientation, scale) of the original images (Fig. 3). The original face images were produced by taking morph steps between a male and female face that had been cropped to minimize boundary face shape differences so that they primarily differed in reflectance. The morph step sizes were selected to equate the image differences produced in the "texturized" versions of the images to that of the differences in blob shape. These texture patches were then projected onto the surface of the sphere that was ultimately deformed to produce the blob. This meant, however, that the visible portion of the surface contrast was lower than that of the faces. The visible surface contrast was further reduced so that the texture patches and the contrast produced by the variation in blob shape matched the contrast energy of the faces (In Experiment II, the texture patches were projected only onto the visible surface of the sphere so that the surface contrast of those blobs matched the contrast of the faces.). Each of the patches was then mapped onto the 64 blobs (selected for the experimental trials) for a total of 512 face patch–blob combinations.

2.2.3. Face stimuli

The stimuli for the face-matching task were gray-level photographs of 86 normal, male and female faces from the University of Stirling web site (http://pics.psych.stir.ac.uk), taken to have neutral expressions and minimal local identity cues such as moles. The pictures were cropped to exclude the hairline and ears, so that only shape and surface reflectance variations could be used to distinguish the stimuli (Fig. 4b).

2.2.4. Stimulus scaling

The physical similarity of pairs of blobs and faces was scaled by a wavelet similarity measure (Lades et al., 1993; Gabor-jet model). This model correlates, over pairs of images, activation values of Gabor wavelets arranged in a rectangular lattice of columns (or "jets") with each jet composed of wavelets at five scales and eight orientations with odd (sine) and even (cosine) phases yielding a vector with 80 values. The correlation for a pair of jets (occupying the corresponding positions in the two lattices, one for each image) is computed as the cosine of the angle between them and the overall similarity value is the average cosine (×100) over the set of jets, with 100 being the maximum correlation (of an image with itself). Many of the phenomena unique to face recognition can be derived from this representation (Biederman & Kalocsai, 1997). These similarity values were highly correlated ($r = .998$) with the distances in the space of amplitude changes of the spherical harmonics (Fig. 2b). The Gabor-jet measure correlates extremely well with human performance ($r > .90$ for error rates and RTs) in same–different matching of metrically varying faces and novel, metrically varying complex shapes (Biederman & Kalocsai, 1997; Biederman & Subramaniam, 1997).

The similarity of the sample to the distractor in the match-to-sample trials for faces varied between Gabor-jet values of 80.0–94.0 in Experiment I, a range bracketed by that of the blob stimuli (69.0–97.0). The analysis thus allowed an evaluation of the effects of contrast reversal for those blob similarities that exactly matched that of the faces. In Experiment II the values for the blob stimuli were 65.8–96.8. It is important to note that the ranges in Experiment I were matched both for the external (silhouette) contrast variation as well as for the full stimuli. An ideal observer analysis (Yue et al., 2006) showed that there were no inherent differences in the information available for discriminating faces and non-pigmented blobs. That is, if there were to be differences in performance between the two classes of stimuli, the differences would arise from within the observer, not the stimuli.

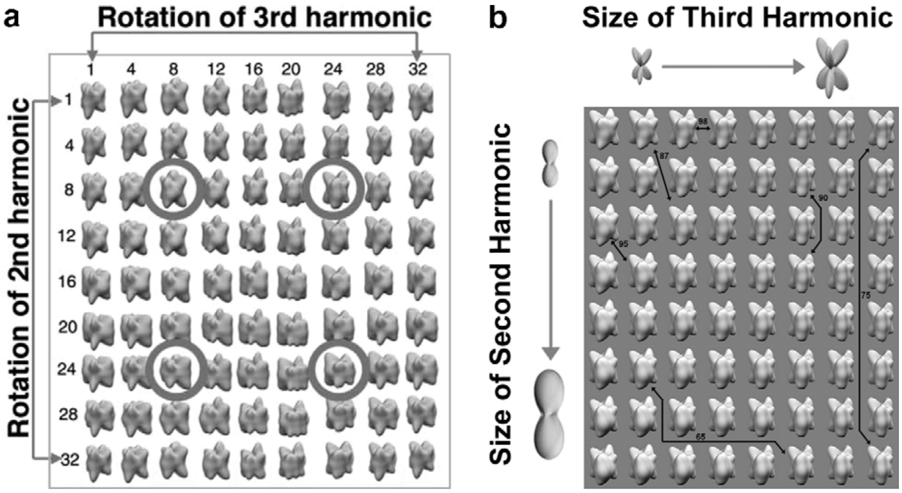

Fig. 2. (a) Blob space produced by combining different orientations of the 2nd and 3rd harmonics, as shown by the orientations above and to the left of the blob space. Proximate blobs were highly similar in shape and those distant were less similar, as confirmed by a Gabor-jet similarity measure (Lades et al., 1993). The four circled blobs were the seeds used to generate the blob spaces, defined by variation in the sizes of the 2nd and 3rd harmonics. (b) A blob space generated by holding constant the orientation of the harmonics but only varying their size, as shown on the axes of the blob space. The illustrated space is generated from the upper left circled blob in (a). The variations in sizes of the harmonics are taken to somewhat mimic the variation in the sizes of facial parts. Both experts and novices were tested with one of the four spaces but the experts gained their expertise on a space defined by a seed diagonally opposite to their test space as shown in (a). Numbers along arrowed lines pointing to pairs of blobs show the Gabor-jet similarity values for those pairs. Figs. 1 and 2 adapted from Yue et al. (2006).

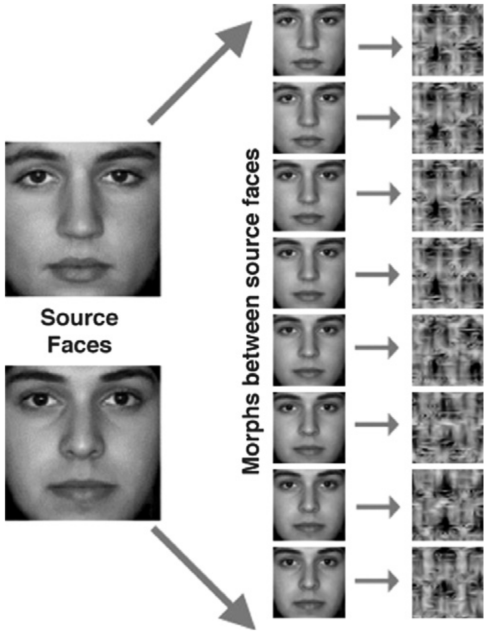

Fig. 3. Production of the variation in reflectance. A male and a female face were cropped to eliminate most of the external boundary and then eight morph steps were computed between them, chosen so the variation in the similarity values of their texturized counterparts would equal those of the blob shape similarity values. The faces were then converted to textures by the Portilla and Simoncelli (2000) algorithm.

3. Procedure

On each trial in a forced-choice, match-to-sample task, three stimuli, all blobs or all faces, were presented simultaneously for 2000 ms (Fig. 4a and b). The sample stimulus was above the two possible matches, one of which was identical to the sample in shape and pigmentation, but possibly not in contrast polarity, when the matching and distractor stimuli would both always be of reversed contrast polarity on those trials when contrast was reversed. It was never the case that the matching stimuli differed from each other in their direction of contrast. For blob and face stimuli, trial types were balanced within subjects for contrast polarity (all stimuli positive, all negative, sample positive/target and distractor negative, sample negative/target and distractor positive) and stimulus similarity over various values. The relatively long presentation duration was chosen to provide the subjects sufficient time to process all three images. The diagonal position of the sample to each of the matching stimuli undoubtedly rendered the task more difficult than if the stimuli were arranged vertically or horizontally. Pilot work had suggested that with shorter durations, particularly on trials where the distractor was highly similar to the matching stimulus, subjects would often only have time to compare one of the two test images to the sample, thus converting the match-to-sample task to a same–different task. Subjects were to respond only after the display was terminated after 2 s. Error feedback was provided by a tone immediately following an incorrect response.

3.1. Expertise training

After a brief familiarization phase of 64 trials, novice subjects ran in 960 test trials of the conditions, with equal numbers of trials with contrast matched and reversed. The participants were asked to judge as quickly and as

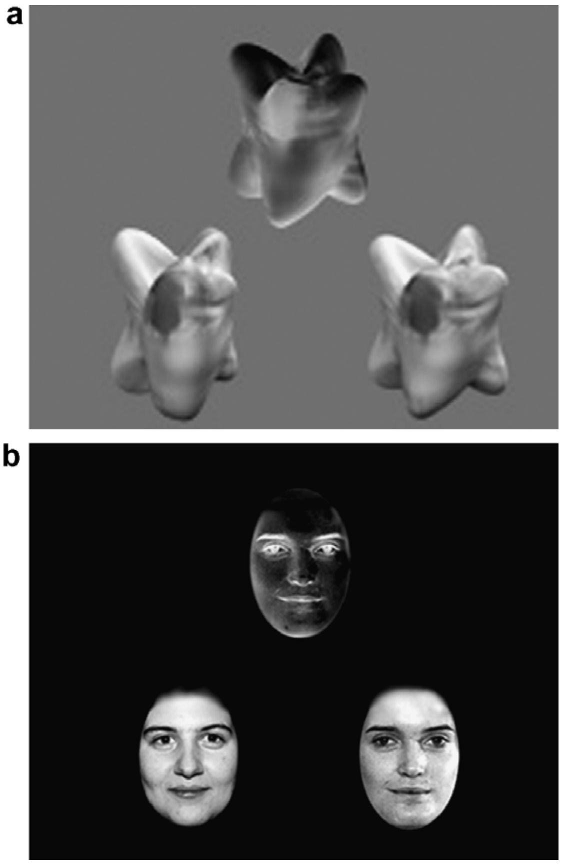

Fig. 4. Sample contrast-reversal trials for blobs (a) and faces (b). Subjects pressed a left or right key according to which of the lower stimuli exactly matched the sample (ignoring direction of contrast). Distractor and matching blobs were equated by Gabor-jet similarity with the faces both overall and in silhouette. The face texture for these blobs was projected onto the whole seed sphere. Thus, some was projected onto non-visible surfaces. In addition, the surface contrast of the blobs was reduced so the contrast energy of the blob shape and surface contrast matched that of the faces. Correct blob and face are on the right.

accurately as possible which one of the two bottom "target" stimuli (left or right) in a given trial matched the top "sample" stimulus by key press, ignoring possible differences in the direction of contrast.

Expert subjects performed this same task as the Novices, but first received training by performing eight sessions of 1024 trials each consisting only of the positive contrast images, for a total of 8192 trials which required about 8 h, far more than has been reported to be necessary to produce expertise in similar object recognition tasks (Gauthier & Tarr, 1997). After attaining expertise, these subjects then completed the test session with images of both positive and negative contrast, identical to that of the Novices. The first half of the test session for the Experts was performed with the same blobs on which they achieved their expertise; the second half was performed with blobs with a new orientation of the harmonics (which matched that of the novices) in order to test for a transfer of expertise to stimuli other than those on which they gained their expertise.

There was a clear benefit of training for the experts, resulting in a drop in error rates from the first session (mean 26.4%) to the eighth (mean 10.2%) across all similarity values (Fig. 5). The training transferred to the untrained stimuli tested in the second half of the ninth session (trained stimuli, 11.1% errors, untrained 12.6%, indicating that the expertise was not one limited to discriminating a particular set of (trained) blobs but was generalizable to discriminating any set of blobs. The experts made the discriminations more quickly and accurately than the novices: Their mean correct reaction times (RTs) was 406 ms shorter with an error rate that was 6.9% lower than that of the novices; $F(1,7) = 71.8$, $p < .001$) for RTs, and $F(1,7) = 53.0$, $p < .001$ for error rates.

4. Results of Experiment I

4.1. Faces

There was a marked increase in error rates for matching faces when the sample and matching stimuli differed in the direction of contrast compared to when they were both positive, from an average of 12.1% to 33.2%, $F(1,7) = 82.2$, $p < .001$, (Fig. 6). This marked cost of contrast reversal was witnessed for every subject, with increases in error rates that ranged from 15.2% to 26.7%. All these differences were significant at least at the $p < .001$ level.

4.2. Blobs (with face textures of lower total surface contrast than the faces)

If what normally produces the costs in contrast reversal when viewing faces is the extensive expertise developed

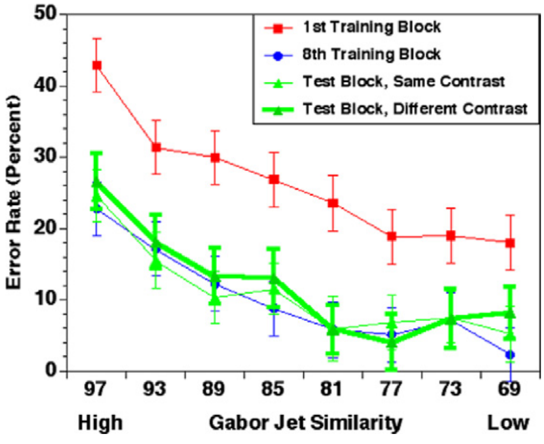

Fig. 5. Error rates for the first and last blocks of expertise training (all with positive blobs) and the testing block (with a new, non-practiced, configuration of the harmonics) as a function of similarity between distractor and matching stimulus and same (sample and test both positive) vs. different directions of contrast (on testing block). The equivalence of the last training block and the first test block implies a complete transfer of expertise. The lack of an effect of contrast reversal on the testing block was apparent over the full range of similarity. The error bars are the means of the standard errors of the individual subjects for each condition and thus do not include the between subjects variance.

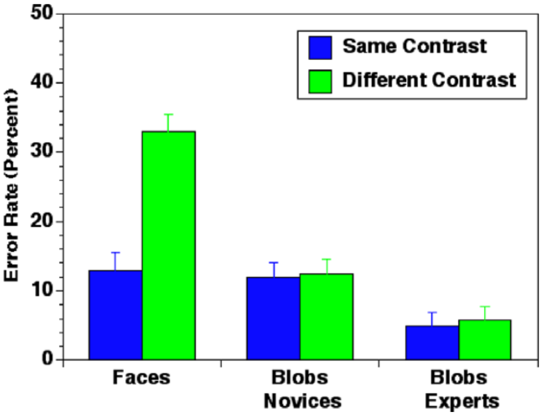

Fig. 6. Error rates for matching faces and blobs for both experts and novices. Whereas the matching of faces with different directions of contrast showed a large increase in error rates there is no effect of contrast reversal when matching blobs. Error bars as described in Fig. 5.

with faces of positive contrast, and that such expertise can be acquired by adults through several thousand discrimination learning trials as maintained by Gauthier and Tarr (1997), we should expect that our experts would show similar costs when discriminating blobs of reversed contrast. However, for both experts and novices there was only a minimal cost in matching blobs of opposite contrast polarity (Fig. 6) that ranged from −1.8% to 5.4% for the novices and from 2.4% to 4.7% for the experts. For none of these eight blob-matching subjects was the cost of contrast reversal significant. This lack of any cost for contrast reversal was apparent throughout the full range of blob similarity (Fig. 5), both for Experts and Novices so the differences in the costs of contrast reversal for faces and blobs cannot be attributed to those regions where the blob similarity values did not match that of the faces. The correlation between the Gabor-jet similarity values of matching and distractor stimuli and error rates was extremely high ($r = .910$, $p < .001$ for blobs, $r = .943$, $p < .001$ for faces) providing a strong justification for the relevance of the Gabor-jet method of scaling the physical similarity of metrically varying complex shapes.

5. Experiment II: Blobs with face textures equal in contrast energy to that of the faces

To address the possibility that the lack of a contrast reversal effect with the blobs compared to the faces in Experiment I was a result of lower surface contrast energy of the blobs, Experiment II was run in which *all* the face contrast was projected onto the half of the sphere that was viewable.

Six novices and two experts performed in this experiment. The blobs were the same as those employed on the test trials in Experiment I except that the contrast energy (mean squared pixel values) from the texturized faces was not reduced to match that of the blob shapes without the textures. Sample stimuli are shown in Fig. 7 and it is evident that they are of higher surface contrast than those shown for the blobs in Experiment I (Fig. 4a). Data for individual subjects and the overall results are shown in Table 1. In this experiment, the experts were not reliably more accurate but were markedly faster than the novices, by 178 ms. As in Experiment 1, there was a strong positive correlation between Gabor-jet similarity (of matching to distractor stimuli) values, 0.87 and 0.81, for error rates and RTs, respectively. None of the subjects had an error rate or RTs for the reversal condition that was reliably greater ($p < .05$ by z-test) than the condition in which both sample and test stimuli were positive. In fact, the error rates for both the Experts and Novices were essentially equivalent for both kinds of trials and their RTs were (non-significantly) longer than on the positive contrast trials than on the reversal contrast trials. Individually, a difference in error rates of 6.71% and in RTs of 51 ms would have produced a significant effect. There was also no effect of whether the stimuli were positive or negative on same contrast trials.

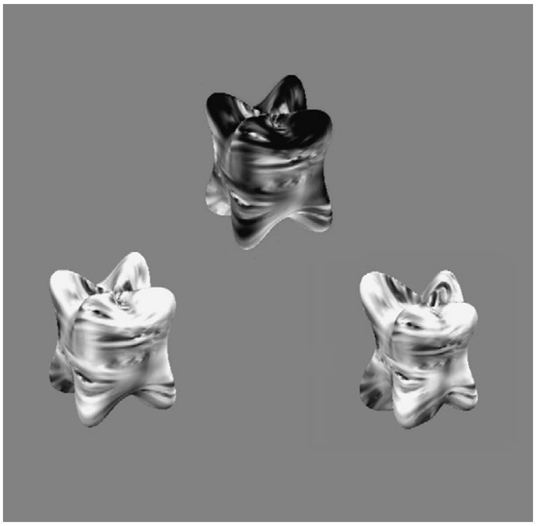

Fig. 7. An example of a match-to-sample blob trial from Experiment II illustrating surface contrast that is equal to that of the face stimuli.

6. Discussion

Here, we have made every attempt to create a class of non-face objects that vary in shape and pigmentation in a manner and extent that matches such variation in faces. Consistent with prior reports (e.g., Galper, 1970; Subramaniam & Biederman, 1997), contrast reversal of our face stimuli resulted in a marked increase (~20%) in error rates in a task in which chance would be 50%. Yet, even after extensive training in matching blobs of only positive contrast, experts (as well as novices) recognized the blobs in a manner invariant to the direction of contrast. The absence of any effect of contrast reversal held true for the full range of similarity between the distractor and matching blobs, which completely overlapped the similarity of

Table 1
Percent correct and mean correct reaction times (RTs) for individual subjects in Experiment II, in which the blob surface contrast matched that of the faces

	Subject	Accuracy			Reaction time (ms)		
		All positive	All negative	Reversal	All positive	All negative	Reversal
Experts	1	0.9268	0.9365	0.9160	304.92	292.87	297.76
	2	0.9059	0.9050	0.9007	345.53	312.75	329.46
	Mean	0.9164	0.9207	0.9083	325.23	302.81	313.61
Novices	1	0.8843	0.8940	0.8810	540.88	554.67	536.71
	2	0.9347	0.9190	0.9374	436.04	456.24	429.08
	3	0.8710	0.9019	0.8839	567.96	571.06	564.03
	4	0.9072	0.8776	0.9032	408.83	424.30	408.44
	5	0.9506	0.9547	0.9578	479.08	469.51	458.46
	6	0.7729	0.7512	0.8121	660.74	633.20	614.27
	Mean	0.8873	0.8833	0.8962	515.59	518.16	501.83

distractor and matching faces. Our match-to-sample tasks precluded the problem of criterion setting present in same–different tasks. The strong association between Gabor-jet similarity values and error rates justified this method of scaling physical similarity. The equating of physical similarity between matching and distractor stimuli for faces and blobs means that the sizable effect of contrast reversal when matching faces cannot be attributed to greater similarity among the face stimuli.

When both sample and test blobs were negative, they were not reliably more difficult to match than when both sample and test blobs were positive (Table 1). Same–different matching tasks with faces, including one run in our own laboratory (Subramaniam & Biederman, 1997), do show a cost of contrast negation, although one that is markedly smaller than the costs of contrast reversal. The cost of matching negative faces is not an effect that would directly be expected from the matching of Gabor activation values although the reported improvement of recognition of contrast-negated faces, when lit from below, is an expected effect (Liu, Collin, Burton, & Chaudhuri, 1999). Why a match-to-sample task would not show a cost of negation but a cost would be manifested in same–different tasks is currently unresolved but one possibility is that the same–different tasks have been run sequentially—requiring memory—whereas sample and test stimuli are available simultaneously in the match-to-sample task as run here, thus reducing the memory requirements. Another possibility is that the match-to-same task was run at a much longer duration—2000 ms—than the same–different task, which was run at 250 ms. It would be a simple matter to assess whether the cost of negation would appear with shorter durations or sequential presentation in a match-to-sample task.

6.1. Why the recognition of faces, but not objects, would suffer from contrast reversal

Biederman and Kalocsai (1997) argued that many of the phenomena associated with face recognition could be derived from the Lades et al. (1993) Gabor-Jet model (see Wiskott, Fellous, Krüger, & von der Malsburg, 1997, as well), which specifies a representation that retains aspects of the original spatial filter (Gabor) values in a retinotopic space. Unlike the Gabor-jet computer-based system which essentially converts simple cell activations to complex cell activations, as an aid to reduce the effect of small image variations on matching, our application to biological vision retains the original simple cell activations, thus rendering the representation sensitive to variations in the direction of contrast. Although all visual stimuli would undergo the same initial filtering, objects would ultimately be represented by "moderately complex features (Kobatake & Tanaka, 1994)," such as geons (Biederman, 1987), which are largely independent of their initial spatial values (Biederman & Cooper, 1991; Fiser & Biederman, 2001; Hayworth & Biederman, 2006; Kobatake & Tanaka, 1994; Yue et al., 2006). Typically, these features would be based on non-accidental differences of orientation and depth discontinuities, and, in the case of highly similar exemplars from a category, discontinuities (at a small scale) in surface marking which would be invariant to direction of contrast. That is, whether a contour is straight or curved or whether there is a spot in the middle of a surface is unaffected by direction of contrast. Such local processing for faces is typically seen only with poor face-recognizers (Cesa, 1994).

The Biederman and Kalocsai (1997) account would also predict that there would be a greater deleterious effect on faces than objects from variations in direction of lighting (Hill & Bruce, 1996; Liu et al., 1999).

Yue et al. (2006) recently reported an fMRI-Adaptation test of the hypothesis that the representation of faces, but not objects, retains aspects of its initial spatial filtering. They created complementary pairs of faces and blobs in the Fourier domain. They did this by filtering each face or blob into eight scales and eight orientations. They then assigned the content of every other orientation-scale combination to one member of a complementary pair and the remaining content to another. Thus each member of a complementary pair was composed of all scales and orientations but in different combinations. During fMRI scanning, subjects performed a same–different task in which they judged whether a face or blob was the same person or blob, ignoring whether it was an identical image or a Fourier complement. As in the present experiment, half the subjects were blob experts and the other half novices (with the same training experience as in the present study). Other than a lower error rate for the experts when matching blobs, the performance and hemodynamic responses in the Fusiform Face Area (FFA) were identical for experts and novices. For both groups, blobs produced markedly less activity in the FFA than faces. Most important, matching complementary pairs of faces (on Same trials) or different faces (on different trials) led to a release from adaptation of the BOLD response in the right FFA compared to when an identical pair of images was shown. For the blobs, there was no release from adaptation in FFA from either a change from one complementary member to another of the same blob or to a change in the identity of a blob. Matching members of a complementary pair of faces produced a marked increase in error rates relative to when the faces were identical. There were no costs from matching complementary (vs. identical) images of blobs although the hemodynamic response for blob experts was greater than blob novices in regions of the lateral occipital complex (LOC) posterior to FFA. Moreover, a change in the identity of a blob (on different trials), produced a release from adaptation for the experts, but not the novices.

The behavioral results with blobs of Yue et al. (2006) thus replicated those of Biederman and Kalocsai (1997) who found no cost from complementation when matching chairs but high costs when matching faces.

Presently, we can only speculate on the mechanisms by which faces wind up with a representation that retains aspects of the original simple-cell coding, rendering their recognition so vulnerable to contrast reversal (and inversion and variations in direction of lighting). Considering the human visual system from an information theoretic perspective we never have more information about the light coming from our world (through reflectance or emission) than what is captured by our retinae. Through successive stages of computation the visual system estimates the necessary information to appropriately interact with our world. Scenes can be navigated, objects recognized and manipulated, speeds and trajectories estimated, all with little or no consequence of the particular lighting conditions in which they occur. The "early information" that accurately represents our visual surroundings is quickly forsaken by these systems in order to provide meaningful, robust, interpretations for actions and generalizable memories. Presumably there is a biological cost for carrying the "early information" forward to successive stages, akin to what is often called the combinatorial explosion. But consider the goal of successful face interaction: accurately storage and discrimination among thousands of highly similar exemplars, with the further requirement for processing minute changes within individuals (to detect emotional states, aggression, even deception). Here may be a case where highly accurate, early representations outweigh the biological costs and lack of robustness (Murray, Kersten, Olshausen, Schrater, & Woods, 2002).

This hypothesis does not necessitate a specialized face system. Instead a flexible, general, system could access early spatial filter coding for any stimulus which must be differentiated from many highly similar exemplars. Along these lines Gauthier and Tarr (1997) argue that visual expertise is likely to cause such coding in any population of stimuli. That only faces appear to access such coding is not a theoretical proposition, but an empirical finding. Here, experts on objects that were as similar to each other as faces, discriminable only by smooth alterations in shape with face-like pigmentation, showed robustness to contrast reversal for the objects but not faces.

7. Conclusions

To summarize the empirical contributions of the present investigation, sources of image variation (e.g., smooth surfaces pigmentation), similarity, and expertise (gained in adulthood), have been ruled out as possible causes of the huge contrast reversal effect when recognizing faces. We are left with the conclusion that the representation that mediates the recognition of faces, unlike those for any other class of objects, is uniquely sensitive to contrast polarity. Human recognition of non-face objects is not sensitive to changes in contrast polarity, even for subjects who have gained expertise in discriminating such objects when the objects have been matched in stimulus similarity and image information (shape and pigmentation) to that of faces. These results are consistent with the account, initially advanced by Biederman and Kalocsai, that the representation of faces—but not objects—retains aspects of the original spatial filter coding which render them highly sensitive to variations in viewpoint, inversion, and direction of contrast.

Acknowledgment

We thank Shuang Wu and Richard Russell for their helpful inputs to this work. Supported by NSF 0420794, 0426415, 0531177, and 0617699.

References

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

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23*, 393–419.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society London: Biological Sciences, 352*, 1203–1219.

Biederman, I., & Subramaniam, S. (1997). Predicting the shape similarity of objects without distinguishing viewpoint invariant properties (VIPs) or parts. *Investigative Ophthalmology & Vision Science, 38*, 998 (Abstract).

Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P., & Fiser, J. (1999). Subordinate-level object classification reexamined. *Psychological Research, 62*, 131–153.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. *Perception, 23*, 803–822.

Cesa, I.L. (1994). Of attraction librarians and lop-sided faces: Development and testing of a training procedure for improving the accuracy of eyewitness (memory, recognition). Unpublished Doctoral Dissertation, Department of Psychology, University of Southern California.

Fiser, J., & Biederman, I. (2001). Invariance of long-term visual priming to scale, reflection, translation, and hemisphere. *Vision Research, 41*, 221–234.

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science, 19*, 207–208.

Gauthier, I., Williams, P., Tarr, M., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38*, 2401.

Gauthier, I., & Tarr, M. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673.

Hayworth, K. J., & Biederman, I. (2006). Neural evidence for intermediate representations in object recognition. *Vision Research, 46*, 4024–4031.

Hill, H., & Bruce, V. (1996). Effects of lighting and the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 986–1004.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology, 71*, 856–867.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., et al. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers, 42*, 300–311.

Liu, C. H., Collin, C. A., Burton, A. M., & Chaudhuri, A. (1999). Lighting direction affects recognition of untextured faces in photographic positive and negative. *Vision Research, 39*, 4003–4009.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academic Sciences of United States of America, 99*(23), 15164–15169.

Nederhouser, M., Mangini, M. C., & Biederman, I. (2001). Object—but not face—matching is invariant to differences in contrast polarity for both naïve and expert subjects. *Journal of Vision, 1*, 92a.

Portilla, J., & Simoncelli, E. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40*(1), 49–71.

Russell, R., Biederman, I., Nederhouser, M., & Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Research, 47*, 157–165.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception, 35*, 749–759.

Shepard, R. N., & Cermak, G. W. (1973). Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology, 4*, 351–377.

Sinha, P. (2002). Qualitative representations for recognition. In *Lecture Notes in Computer Science. LNCS 2525* (pp. 249–262). Berlin: Springer-Verlag.

Subramaniam, S., & Biederman, I. (1997). Does contrast reversal affect object identification? *Investigative Ophthalmology & Visual Science, 38*, 998.

Vuong, Q., Peissig, J., Harrison, M., & Tarr, M. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and Greebles. *Vision Research, 45*, 1213–1223.

Wiskott, L., Fellous, J.-M., Krüger, N., & von der Malsburg, C. (1997). Face matching by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*, 775–779.

Yue, X., Tjan, B., & Biederman, I. (2006). What makes faces special? *Vision Research, 46*, 3802–3811.