Role of ordinal contrast relationships in face encoding

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.0812396106">http://dx.doi.org/10.1073/pnas.0812396106</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Thu Jun 22 16:05:14 EDT 2017</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/50254">http://hdl.handle.net/1721.1/50254</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Role of ordinal contrast relationships in face encoding

Sharon Gilad, Ming Meng, and Pawan Sinha

Communicated by Richard M. Held, New England College of Optometry, Cambridge, MA, December 12, 2008 (received for review September 10, 2008)

In principle, a contrast negated image is exactly as informative as its positive counterpart; negation perfectly preserves an image’s 2D geometric and spectral structure. However, as anyone who has had to search through a roll of negatives for a snapshot of a particular person knows, this simple operation has dramatically adverse consequences on our ability to identify faces, as illustrated in Fig. 1 (1–6). Exploring the causes of this phenomenon is important for understanding the broader issue of the nature of information the visual system uses for face identification.

Several researchers have hypothesized that negated face-images are hard to recognize because of the unnatural shading cues in negatives, which compromise shape from shading processes (7–11). The resulting problems in recovering veridical 3D facial shape are believed to impair recognition performance. Although plausible, it is unclear whether this explanation is a sufficient one, especially in light of experimental results showing preserved recognition performance in the absence of shading gradients (12), and theories of face recognition that are based on the use of 2D intensity patterns rather than recovered 3D shapes (13, 14). Another prominent hypothesis is that negation causes faces to have unusual pigmentation (15, 16). However, the adequacy of this “pigmentation hypothesis” has been challenged by data showing that hue negation, which also results in unnatural pigmentation (making the entire face look bluish-green), has little effect on recognition performance (11). It is also unclear whether observers can usefully extract pigmentation information across different illumination conditions (17, 18). We propose an alternative account of negation-induced impairment. Of the infinitely many aspects of facial photometry, our results suggest that the destruction of a small set of stable 2D contrast polarity relationships might underlie negation-induced decrements in face-recognition performance.

In our past work on face representation under variable illumination conditions (19, 20), we have found that polarity of contrast around the eyes is a remarkably stable feature, with the eyes usually being darker than the forehead and the cheeks, as illustrated in Fig. 2. The absolute magnitude of contrast across different regions of a face changes greatly under different imaging/lighting conditions, but the local polarity relationships between the eye regions and their neighborhood are maintained in all but the most unnatural lighting setups (such as lighting a face from below). Dark or oily facial complexion also does not typically disrupt these relationships. Watt (21) and Chen et al. (22) too have remarked on the stability of these polarity relationships. The use of ordinal brightness relationships confers significant tolerance for photometric variations in images and may help explain perceptual and neural invariance to illumination changes (23, 24) and their effectiveness for classifying faces when used in machine vision systems (19, 25).

To the extent that the visual system learns object concepts by encoding regularities in the observed data (26–28), it seems likely that stable contrast polarity relationships would come to be incorporated into the facial representation used by the brain. Mismatches between this internal representation and input data will then be expected to lead to decrements in recognition.

Contrast negation of face images leads to the destruction of otherwise highly consistent polarity relations. We hypothesized that this may be a factor leading to the poor recognizability of negated faces. To test this hypothesis, we created a set of “contrast chimeras.” These are faces that are photo-negatives everywhere except in the eye region (thus preserving local contrast polarity in that neighborhood). Such faces still have unnatural pigmentation and shading cues over much of their extents and present largely the same problems to shape-from-shading processes as the full negatives do. Explanations based on disruptions due to unnatural pigmentation or incorrect shape-from-shading cues would not, therefore, predict significant improvements in performance with such chimeric faces beyond the improvements derived from the intrinsic recognizability of the eyes. However, if performance in negatives is compromised, at least in part, because of the destruction of local polarity relations between the eyes and their neighborhood, performance with contrast chimeras is expected to be significantly better than that with contrast negatives.

Results

Fig. 3A shows a few contrast chimeras of the kind used in our experiments. It is interesting to note the large perceptual difference brought about by reinforcement of the eyes. We used 24 monochrome celebrity face images spatially normalized to have the same inter-
pupillary distance. For each original positive image (P), we created three additional variants: a full-negative (N), positive eyes on a head-silhouette (E), and a contrast chimera (C). These variants are shown in Fig. 3B. The “eyes on head silhouette” condition was included in our set to allow us to determine whether any observed performance gains in the chimeric condition are due merely to the intrinsic recognizability of the positive eyes on their own.

We conducted 2 experiments. The first used a between-subjects design whereas the second adopted a within-subjects approach for a finer-grain analysis of data. Fifteen subjects participated in the first experiment. They were assigned randomly to 3 equal-sized groups. Subjects in a given group were shown all faces in one of the 3 experimental conditions (full-negative, eyes on head-silhouette or contrast chimera). They were asked to name the individuals shown and were subsequently shown all faces in the full-positive condition to determine which celebrities they were familiar with. Performance for each subject was recorded as the proportion of familiar faces recognized in the experimental condition they had been assigned to. The results are shown in Fig. 3B. The mean level of recognition performance we found in the full-negative condition (54.35%) is consistent with results from previous studies that have included in our set to allow us to determine whether any observed performance gains in the chimeric condition are due merely to the intrinsic recognizability of the positive eyes on their own.

The results are shown in Fig. 3C. As in experiment 1, we found that contrast chimeras were recognized at a significantly higher rate than the full negatives (P < 0.01, 1 tailed t test). More importantly, the results showed that the increase in performance could not be explained simply by appealing to the recognizability of the eyes by themselves (P < 0.01).

It is important to note that, whereas performance with contrast chimeras is high, it is slightly below that with the full-positives. There could be a few explanations for this. First, we have considered only one set of polarity relationships—those between the eyes and their neighborhood. More such polarity relationships are likely to exist in other regions of the face. The inclusion of these other relationships in the chimeras can conceivably equalize performance between the chimeras and the original positive images. Second, our explanations based on 2D ordinal relationships, and the pigmentation or shape-from-shading based accounts, are not mutually exclusive. It is possible that unusual pigmentation or unnatural shading cues, may lead to a small decrement in performance, enough to account for the differences between the positive and the chimeric conditions.

Besides the aforementioned behavioral experiments, another approach for examining the significance of ordinal relationships is to determine whether the inclusion or violation of these relations modulates neural responses to face images. To this end, we recorded brain responses to negative, positive, chimeric and eye
images, using functional magnetic resonance imaging (fMRI). Eight subjects were shown 25 instances each of these 4 classes of face images in M-sequences (29) and their brain activations recorded using a rapid event related design. For any one subject, a particular face appeared in only 1 of the 4 conditions. The condition in which a particular face was shown was counterbalanced across different participants. Subjects were asked to continuously monitor a small fixation cross (0.7° in extent) so as to avoid confounds from potentially differing patterns of scan-paths across the different stimulus conditions. They were not required to make any judgments regarding the faces shown.

Previous studies have implicated regions in the fusiform gyri [fusiform face areas (FFAs)], especially in the right cerebral hemisphere, in face perception (30). We identified the right FFA via separate localizer runs for each subject and then examined activations in this region corresponding to the 4 stimulus conditions. Fig. 4 shows the results.

Consistent with past reports (2), fully positive faces led to significantly increased brain activity in the right fusiform face area relative to fully negative faces ($F_{1, 40} = 4.8, P < 0.05$). Interestingly, responses evoked by contrast chimeras are as high as, and statistically indistinguishable from, those corresponding to fully positive faces ($F_{1, 40} = 0.11, P = 0.74$). By contrast, eyes embedded in the head-silhouette lead to minimal activation in the fusiform face areas, much less than positives and chimeras (P vs. E: $F_{1, 40} = 13.76, P = 0.001$; C vs. E: $F_{1, 40} = 7.18, P = 0.01$), and similar to full negatives ($F_{1, 40} = 0.89, P = 0.35$). These results suggest that the regular polarity of contrast around the eyes is important for eliciting significant neural activation to faces. A similar trend is observed in the left fusiform face area, although the level of statistical signifi-
The reason why ordinal relationships around the eyes might be especially significant could be based in the statistics of facial-feature variability and diagnosticity. The eye neighborhood appears to embody a very stable set of relations across many different faces and illumination conditions. This is borne out in analyses that Viola and Jones (25) undertook toward designing a computational face-detection system. They considered many thousands of relationships across various image regions and evaluated each in terms of its ability to indicate whether the underlying image was a face or not. The top two features from this analysis are shown in Fig. 5. It is interesting to note that they both comprise contrast relationships in the eye neighborhood.

This importance of ordinal contrast relationships around the eyes may also explain a few other well-known observations regarding face recognition. A primary one is the difficulty people experience in recognizing faces that are lit from below (making the eye region brighter than the local neighborhood, thereby upsetting the polarity relationships) (10). An interesting prediction that the ordinal-relation hypothesis makes is that for bottom-lit faces, negatives should be easier to recognize than positives, because the ordinal relationships are restored by this manipulation. This counterintuitive prediction is indeed supported by empirical data. Liu and colleagues (31) found precisely such an effect in their studies of lighting direction and negation. (As an aside, it is worth noting that negatives of faces lit from above are not the same as positive images of faces lit from below.)

Another finding that the ordinal-relations hypothesis can help explain is the difficulty of recognizing unpigmented face depictions (32), such as stone busts. It is unlikely that the difficulty arises from disrupted shape-from-shading processes, because such busts present ideal conditions for the operation of these processes (33). However, they do not provide the contrast polarity relationships between the eyes and their neighborhood that real faces do. Interestingly, Bruce and Young have noted that painting dark irises on such busts makes them look much more “life-like” (7, pages...
163–165), further suggesting the role of ordinal contrast relationships in the representation of faces.

How can these findings regarding the perceptual significance of photometric relationships be reconciled with our ability to recognize line-drawings of faces? Intuitively, line-drawings appear to contain primarily contour information and very little photometric information over which to define the luminance relations. However, experimental data suggest otherwise. Studies (34, 35) have found that line-drawings that contain exclusively contour information are very difficult to recognize (specifically, they found that subjects could recognize only 47% of the line-drawings compared with 90% of the original photographs). How can we resolve such findings with the observed recognizability of line-drawings in everyday experience? Bruce and colleagues (7, 36) have convincingly argued that such depictions do in fact contain significant photometric cues, i.e., the density and weight of the lines change the relative intensity of different facial regions. In essence, the contours included in such depictions by artists correspond not merely to a low-level edge-map, but in fact embody a face’s photometric structure. It is the skillful inclusion of these photometric cues that is believed to make human generated line-drawings more recognizable than computer generated ones (37). If the recognizability of line-drawings is governed in part by the photometric cues they contain, then it stands to reason that they would be susceptible to contrast negation, just as continuous-tone images are. This prediction is indeed supported by experimental data (38).

These results point to several interesting questions for future research. We mention three, and offer tentative answers for two. First, is the ordinal encoding strategy a physiologically plausible one? In other words, how can ordinal measurements be extracted by the human visual system? A potential answer might be found in the contrast response profiles of neurons in the primary visual cortex. As Fig. 6 shows, many neurons in mammalian V1 exhibit rapidly saturating responses as a function of contrast. These profiles are approximations of step-functions that would characterize an ordinal comparator. Thus, neurons in the early stages of the mammalian visual pathway provide a plausible substrate for extracting ordinal image relationships.

A second important question for future research concerns the specificity of these results to the domain of faces. Negation effects are known to be more pronounced for faces than for other classes of objects (1). Does this mean that ordinal contrast relationships are involved only in encoding faces? Not necessarily. Faces constitute a particularly homogenous class, with a very consistent set of photometric relationships within them. A more heterogeneous collection of objects, even though comprising a single category such as “houses” or “cars,” might not preserve such consistency. The windows of a car can, for instance, be brighter or darker than the rest of the body. Given this variability, an ordinal code would not privilege one direction over the other. Rather, it will simply include both directions as equally valid, effectively indicating that a difference relationship ought to be present in a given image location, without mandating what the sign of the relationship must be. Such a code will not be affected by negation. Even though this seems to be too permissive an encoding strategy, it has been demonstrated to be quite effective for machine vision systems designed to recognize real-world objects (40).

A third question pertains to the relevance of these results to the study of neurological disorders such as autism, which are believed to be associated with face processing abnormalities. Specifically, individuals with autism are known to avoid eye-contact and tend to focus more on the mouth region (41, 42). If this tendency changes...
the relative saliences of image regions and relationships to be weighted more toward the mouth, we would expect that mouth-chimeras would be more facilitatory for recognition performance of autistic observers rather than eye-chimeras, in contrast to the pattern we observed. Results have described above with neuro-typical observers. Such a study would provide clues about the nature of facial representation in autism.

To summarize, our findings provide a potential answer to the long-standing question of why faces are hard to recognize in negative contrast images. More generally, they suggest that contrast polarity relations between face regions in the vicinity of the eyes might be embodied in the visual system’s facial representations and serve as strong determinants of recognition performance.

Materials and Methods

Stimuli. Chimeric face images were generated by digitally compositing positive eye-regions on negative faces. The eye-region was approximately spectacle shaped, extending from the top of the eyebrows to the lower margin of the eyes and from the left to the right tips of the eyebrows. To minimize the creation of edge artifacts in the cropping process, we first changed the overall intensity of the face image so as to make the skin be midgray, and feathered the interface between positive and negative regions.

fMRI Experiment. Eight adult volunteers in the fMRI experiment. Scanning was performed on a 3.0-Tesla Siemens scanner, using a standard head coil at the participant. Stimuli were rear-projected onto a screen in the scanner bore. Experimental stimuli consisted of 100 face images in negative, positive, chimeric and eye-only conditions. In each experimental run, these 100 faces mixed with 24 fixation-only trials were presented in a different M-sequence (29). Each trial was 2 s long. Each face image was presented for 300 ms followed by 1700 ms ISI. In addition, each run began and ended with 12 s of fixation-rest. For each participant, a particular face was presented in only 1 of the 4 conditions with 25 faces in each of the conditions. The condition in which a particular face was shown was counterbalanced across different participants. In summary, each participant saw the same 6 sequences of the same 100 faces but the faces were shown in different counterbalanced conditions. During the experiments, a green dot was randomly presented onto a random location of the screen. The fixation cross was also presented randomly in red or green. Participants were instructed to continuously monitor these changes and when the fixation was green report whether the previous display had a green dot.

fMRI Analysis. All fMRI data underwent 3-D motion correction, slice time correction and analysis, using SPM2 (http://www.fil.ion.ucl.ac.uk/spm/software/spm2) and custom-routines in Matlab. Slow drifts in signal intensity were removed by linear detrending, but no other temporal smoothing was applied. fMRI data of each participant were normalized into SPM-MNI space and were spatially smoothed (FWHM ~ 6 mm). ROI localization was accomplished by 3 separate block-designed runs of full color faces vs. objects. These faces were different from the face images used in the experimental runs. ROI was defined as the set of contiguous voxels in fusiform gyrus that showed significantly stronger activation (P < 0.05, uncorrected) to faces than to objects (fusiform face area (FFA) (30)). Right FFA was reliably located in all 8 participants, whereas left FFA was found in 6 of the 8 participants. Event-related average of fMRI activity in these ROIs was sorted according to the 8 experimental conditions. Percentage MR signal change was calculated by using activation at t ~ 0 s of each trial as the baseline. Peak amplitude of 4 to 6 seconds after stimulus onset was used to perform statistics.

ACKNOWLEDGMENTS. We thank Irving Biederman, Daniel Kersten, Ami Klin, Thomas Papathomas, and Roger Tootell for their helpful comments on this work. This work was supported by an Alfred P. Sloan Foundation Fellowship in Neuroscience (to P.S.) and by the Simons Foundation.