Stimulus inversion and the responses of face and object-sensitive cortical areas

Geoffrey Karl Aguirre, CA Rajiv Singh and Mark D’Esposito

Department of Neurology, University of Pennsylvania School of Medicine, 3400 Spruce Street, Philadelphia, PA 19104-4283, USA

Key words: Face perception; fMRI; Fusiform face area; Stimulus inversion

Introduction

A number of experimental findings support the assertion that face recognition is subserved by a system that is both anatomically and computationally specialized. The evidence for anatomical specialization is provided by neuropsychological studies that have found double-dissociations in face and object recognition impairments following brain damage [1–3] and by neuroimaging studies that have found distinct face-responsive (the so-called fusiform face area; FFA) and object-responsive cortical areas [4–6]. The evidence for computational specialization has been drawn from psychophysical studies of visual recognition that find that stimulus inversion disproportionately impairs face recognition as compared to recognition of other stimulus types [7].

Further insight into the nature and interaction of neural systems supporting visual recognition has been provided by studies of stimulus inversion involving brain-damaged patients. Patients who suffer from prosopagnosia, arguably caused by damage to the FFA, do not have an impairment relative to normal controls for inverted face recognition [8] and can even demonstrate a reverse face inversion effect: they perform better with inverted than with upright face stimuli [9]. Conversely, a recent series of studies performed on patient CK, who displays intact face recognition in the presence of a dense object agnosia, found an exaggerated face inversion penalty [3]. These findings suggest a model of visual recognition systems with (at least) the following properties: (1) separate cortical regions necessary for the perception of faces and general objects; (2) non-involvement of the FFA in inverted face perception (suggested by CK’s impairment with inverted faces); and (3) involvement of the object area in inverted face perception (suggested by prosopagnosics’ facility with inverted faces).

Let us assume that there is a link between the involvement of a cortical region in the perception of a given stimulus type and the magnitude of the neural response of that region to that stimulus, such that the region is tuned to respond maximally to stimuli for which it provides necessary processing for perception (note that this does not have to be so [10]). If this is the case, and given the properties of the model enumerated above, we would predict that the FFA would respond with a greater intensity of neural firing to the presentation of upright as compared to inverted faces, while an object area would have the opposite pattern of response.

We used an initial fMRI scan to define in each of eight subjects cortical regions that display greater responses to faces than to general objects, and regions that have the opposite pattern. Within these regions, we then examined the fMRI signal obtained during the presentation of upright and inverted faces and cars. Several features of the experimental design merit comment. First, a set of car pictures was studied along with faces. Cars were used as they constitute exemplars of the same category, as do faces, and were included to determine if any inversion effects are specific for faces or are observed for
other stimulus types as well. Second, subjects performed a detection task that produces inversion effects (RT penalties) for both face and object stimuli [11]. The clarifying effect that this design choice has upon inference is considered in the discussion section. Finally, it should be noted that the event-related [12] task design that was used permitted the randomization of stimulus order, removing the possibility of long-term behaviors associated with the perception of inverted stimuli confounding functional responses to the stimuli themselves.

Materials and Methods

Task design: After obtaining informed consent, eight subjects (two males, all right-handed, mean age (± s.d.) 21 ± 2 years) performed a simple visual detection task upon stimuli presented every 16 s. Each trial began with the presentation of a central fixation cross (black on a white screen) for 500 ms. The fixation cross was immediately followed by the presentation for 200 ms of an 8-bit grayscale picture. This brief duration was intended to minimize saccadic eye movements during stimulus presentation. The pictures were drawn, in a pseudo-random order, from one of three categories: upright stimulus, inverted stimulus, and scrambled stimulus (Fig. 1). An inter-trial-interval followed during which a blank, white screen was presented. Subjects were instructed to make a bilateral button press only in response to scrambled stimuli, thus preventing overt motor responses from contributing to signal changes associated with the perception of upright and inverted stimuli.

Each subject performed the task twice during a single session, once with face stimuli and once with car stimuli (order randomized across subjects). There were 184 face trials (83 upright stimuli, 83 inverted stimuli, and 18 scrambled stimuli) and 92 car trials (same proportions). The greater number of face trials was designed to provide additional sensitivity for functional face vs inverted face comparisons that are not the subject of this report. While the smaller number of car trials reduces the sensitivity of tests that compare upright and inverted cars, the results (in which significant differences were detected between upright and inverted cars) suggest that this was not of practical importance.

Because the response protocol used during scanning precludes the ability to measure reaction times to upright and inverted stimuli, a version of this task was also conducted on a separate group of five subjects (three males, all right-handed, mean age (± s.d.) 19 ± 1 years) outside of the scanner. Testing conditions mimicked the scanning environment with two notable alterations. First, subjects were instructed to respond on every trial, pressing one pair of buttons in response to scrambled stimuli and another pair to intact, upright or inverted stimuli. Second, a full complement of 184 car trials was conducted.

MRI data acquisition and general processing: Our imaging protocol and data pre-processing have been described in detail previously [13]. Briefly, echoplanar imaging was conducted on a 1.5 T SIGNA scanner (TR = 2000 ms, TE = 50 ms, 21 axial slices, 3.75 x 3.75 mm resolution in plane, 5 mm through plane with no skip). After image reconstruction the data were since interpolated in time to correct for the fMRI acquisition sequence [13], motion corrected [14], and then processed to remove spatially coherent signal changes attributable to bulk head motion [15]. No spatial smoothing of the data was undertaken as we wished to retain maximal anatomic resolution.

Definition of functional regions of interest (ROIs): ROIs were defined for each subject by a two-step process. For the first step, subjects participated in an initial scan that was separate from the detection task scans. During this preliminary fMRI study, subjects passively viewed gray-scale faces or general, non-food objects (different from the stimuli used in the primary task), presented every two seconds, arranged into 20 s blocks (20 blocks total). This type of preliminary scan reliably and reproducibly detects the FFA across subjects [5,6]. The data were ana-

![FIG. 1. Examples of stimuli. These 256 x 256 pixel images were taken with a digital camera. The lighting conditions during photography were designed to minimize shadow cues that have been proposed to act as a confounder of face inversion [24]. Image editing software was used to rearrange the position of the internal features to create the scrambled stimuli, which were then filtered to soften and remove the cut-and-paste lines that resulted. Although the same stimuli were used in upright and inverted sets, the randomization of the order of the stimuli (upright and inverted) within and across subjects precludes attribution of differences observed between upright and inverted stimuli to the effect of repetition.](Image 1 to 289x307)
lyzed using a modified general linear model [16], implemented as described previously [15,17]. The resulting SPM(t) map for each subject was thresholded (t > 3.5 in all cases) to identify clusters of voxels in the ventral cortex with greater responses to faces than to general objects, and vice versa. The second step of region definition involved winnowing these ROIs to remove voxels that did not have positive responses to visual stimuli during the main detection task. To do so, the t-value corresponding to the main effect across all stimuli in the model described below was obtained for each voxel. Only those voxels in which this main effect was > 1.65 were retained as part of the ROI. This step guards against the inclusion within the ROI of spurious voxels that are not visually responsive and is statistically valid as all subsequent contrasts evaluated for hypothesis testing are orthogonal to this main effect.

The preliminary face vs object scan was also used to derive a subject-specific model of intrinsic temporal autocorrelation of the fMRI data [13] to be used in the analysis described below.

**Analysis of detection task scans:** The fMRI data from each subject were analyzed separately. The average fMRI time-series was obtained from each of the two ROIs and evaluated with a modified GLM. In addition to a subject-specific model of temporal autocorrelation, the K matrix [16] contained a notch filter designed to remove low frequency confounds (≤ 0.025 Hz) and high frequency noise (≥ 0.244 Hz).

Independent variables were created for each stimulus type (i.e. faces, inverted faces, cars, inverted cars). In addition, all trials in which a scrambled stimulus was presented or in which the subject made a response were modeled by a separate covariate. Thus, the primary covariates for each stimulus category were insensitive to signal changes that might have resulted from overt response or target presentation. The independent variables themselves were formed using the first three principle components (eigenvectors) derived from a set of hemodynamic response functions from an independent group of subjects [13]. The first eigenvector is essentially an across-subject, average hemodynamic response function and it was the relationship between this covariate and the fMRI signal that was of interest. The other two covariates were included to model nuisance variance components. Nuisance covariates modeling the trial means were also included. Application of this analysis method to null hypothesis data has demonstrated an empirical false-positive rate not significantly different from tabular values [6].

Three planned contrasts were conducted on each signal: upright faces–inverted faces, upright cars–inverted cars, and upright faces–upright cars. Each contrast yielded a t-value for each subject. To determine whether a significant effect was present in the population from which our subjects were drawn, a t-test was conducted upon the set of t-values obtained across the subjects.

**Results**

**Behavioral effect of inversion:** Table 1 presents the RT and sensitivity (d') measures for the detection task. An ANOVA performed upon the RT data indicated a significant effect of orientation (F(1,13) = 6.5, p = 0.024) but not stimulus (F(1,13) = 1.8, p = NS), confirming that stimulus inversion produced an RT cost of approximately 40 ms for both faces and cars. Sensitivity (d') measures did not differ between the stimulus categories (F(1,4) = 4.2, p = NS), or between the subjects who performed the behavioral or fMRI scan version of the task (F(1,23) = 0.03, p = NS). The magnitude of the inversion RT penalty, and its presence for both face and car stimuli, is consistent with findings from previous, non-recognition inversion tasks [7,11].

**Region definition:** Figure 2 presents the locations of the centers of the ROIs defined for face and object areas for all subjects. Voxels corresponding to the FFA were identified for all eight subjects. For six subjects, voxels responsive to faces were located either entirely or predominantly in the right hemisphere, within the lateral fusiform gyrus. A smaller number of voxels were also seen in a homologous position in the left hemisphere for three of these

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Behavioral (n=5)</th>
<th>Scan (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT (msec)</td>
<td>d' (upright and inverted)</td>
</tr>
<tr>
<td>Face</td>
<td>438 ± 131</td>
<td>3.29 ± 1.36</td>
</tr>
<tr>
<td></td>
<td>477 ± 131</td>
<td></td>
</tr>
<tr>
<td>Car</td>
<td>414 ± 117</td>
<td>3.69 ± 1.23</td>
</tr>
<tr>
<td></td>
<td>458 ± 116</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Task performance. The first three columns of data present the summary measures of the performance of five subjects studied outside of the scanner, while the fourth column presents the d' scores for the subjects who participated in the fMRI study. As can be seen, detection accuracy was near ceiling.
subjects; only the right hemisphere voxels were included, however, in the ROI. Two subjects had face-responsive voxels exclusively in the left hemisphere. The face ROIs averaged 6 voxels \(0.42 \text{ cm}^3\) in volume. The mean \(\pm \text{s.d.}\) position of the centers of the face ROI for the six subjects with dextral regions was \((x,y,z) = 37 \pm 5, -62 \pm 6, -17 \pm 3\).

Object-responsive voxels were identified in six of the eight subjects. For three subjects, these voxels were found only in the left hemisphere and for three subjects bilateral object-responsive regions of about equal size were found. Voxels in both hemispheres were included in the functional ROIs for the object area. The object-responsive voxels were generally located medial to the FFA. The object ROIs averaged 12 voxels \(0.86 \text{ cm}^3\) in volume. The mean position of the center of the left-hemisphere, object-area ROI was \((x,y,z) = -32 \pm 5, -64 \pm 6, -17 \pm 3\). Two subjects also had a small number of voxels located more anteriorly, near the junction of the collateral and intra-lingual sulci. These voxels were not included in the ROI.

*Functional responses to upright and inverted stimuli:* Figure 3 presents the across-subject, average evoked responses to the different stimulus classes within the two functionally defined areas. Table 2 provides a summary of the results of the statistical contrasts. Within the FFA, upright faces were found to evoke slightly, but significantly, greater signals than did cars. However, no magnitude difference was found between upright and inverted cars or between upright and inverted faces. In contrast, a marked effect of inversion was found within the object region for both face and car stimuli. For both stimulus types, inversion resulted in an increase in the magnitude of the evoked fMRI response in the object area. No significant difference, however, was found between upright faces and upright cars. While this last finding may seem surprising given the apparent mean difference between these stimuli (Fig. 3B), there was considerable variability from subject-to-subject as to its magnitude (and even direction for one subject), resulting in the small across-subject t-value shown.

**Discussion**

The neuropsychological literature considered at the outset led us to predict that inversion of face stimuli would augment the neural response in object regions.
and attenuate the response in face regions. The first of these two predictions was upheld, supporting the assertion that inverted faces are treated as general objects by the visual system. The second prediction was not. Instead, upright and inverted faces were found to evoke the same magnitude of neural activity within the fusiform face area. It should be noted that this finding precludes isolated differences between upright and inverted faces in either the intensity or the duration of evoked neural response, as increases in the duration of neural activity would also be expected to increase the magnitude of fMRI signal response (assuming, however, a linear transform of neural activity to fMRI signal).

The finding of equivalent amplitude responses in the fusiform face area, while not predicted by the neuropsychological literature, was not wholly unanticipated. Neurophysiological studies in monkeys have found that face-selective neurons respond with equivalent firing to upright and inverted face stimuli [18] and event-related-potential studies in humans have found equivalent amplitude waveforms with inversion [19,20]. Additionally, a recent, blocked fMRI study found that inversion of grayscale face stimuli resulted in only a small (albeit significant) change in fMRI signal [21]. How might we reconcile this similarity of response with the neuropsychological and behavioral evidence that inverted faces do not engage face selective mechanisms? A ready explanation is that the region automatically engages in processing that is in some cases unnecessary [10]. While the recognition of inverted faces may depend critically upon computations performed by (for example) general object regions, these stimuli may subsequently and reflexively engage the FFA, perhaps following initial manipulation (e.g., mental rotation) by other regions. Interestingly, the electrophysiological studies cited above have consistently found a neural onset asynchrony between upright and inverted faces, with the response to inverted faces commencing tens to hundreds of milliseconds later than the response to upright faces. One might
plausibly argue that the source of this latency difference is a longer duration of neural processing associated with inverted faces that takes place at some other cortical location. Thus, the activity seen in the FFA in response to inverted faces may be the signature of the eventual perception of the stimulus as a face, even when that perceptual accomplishment is not required for the task in which the subject is engaged. The existence of an inversion neural onset asynchrony within the human FFA is currently the subject of study in our laboratory.

Car inversion was also found to increase the magnitude of evoked fMRI signal within object regions and to have no effect upon the response in face regions. Thus, the effect of stimulus inversion upon evoked activity within the object region is not specific to faces. These increases in signal magnitude might be attributed to increases in the intensity or the duration of neural activity, as the hemodynamic system studied using fMRI effectively integrates neural activity over short (<4 s) time scales. The relative contribution of intensity and duration changes with stimulus inversion is of some interest, as the two types of modulations suggest different computational mechanisms. An increase in the intensity of neural activity in response to stimulus inversion might be taken as evidence that a given cortical area is tuned to (i.e., detects, or represents the features of) the inverted stimulus preferentially, while an increase in the duration of neural activity might imply some computational process (e.g., mental rotation) that proceeds in a serial fashion.

The task design used here produced equivalent RT inversion penalties for faces and cars. A facile explanation for the inversion related increase in fMRI signal in the object area is then simply that a longer duration of neural processing, as indexed by the longer RTs, took place in this region for both faces and cars, suggestive of a serial process. This proposal can be tested in further studies in which a recognition task that does not produce an RT penalty for cars is used. Eliminating the RT penalty associated with stimulus inversion should also eliminate the change in fMRI signal if these changes are attributable to an increase in duration of neural activity.

Finally, it should be noted that this study considered the responses of a cortical area sensitive to general objects, similar to that identified by Kanwisher and colleagues [4]. Recent neuroimaging studies have, however, demonstrated the heterogeneous nature of ventral occipito-temporal cortex for stimulus sensitivity [6,22,23] and have questioned the existence of cortical regions that are truly generally responsive to non-face, object stimuli [6]. These findings raise the possibility that different locales within ventral cortex might have different responses to upright and inverted stimuli. Such hypotheses await higher spatial resolution and further functional definition of regions for testing.

Conclusion
Contrary to predictions based on neuropsychological and behavioral studies, the inversion of face stimuli does not alter the magnitude of neural response within human, face-selective cortical areas. This finding thus argues against the fusiform face area playing a causal role in the generation of the behavioral face inversion effect. Inversion of face stimuli, as well as inversion of car stimuli, does increase the evoked fMRI signal within a functionally defined, object region. These changes may be plausibly attributed to increases in the duration of neural activity associated with the stimulus presentation, as indexed by the reaction time penalty observed for both faces and cars. Apparently, the additional processing required for inverted faces is undertaken by object-responsive regions, in agreement with the behavioral and neuropsychological studies that argue that inverted faces are processed by object perception systems.

References

ACKNOWLEDGEMENTS: This work was supported by grants from the National Institutes of Health (NS01762 and AG13483) and the Charles A. Dana Foundation and the American Federation for Aging Research.

Received 16 September 1998; accepted 8 November 1998