

# Covert Visual Attention Modulates Face-Specific Activity in the Human Fusiform Gyrus: fMRI Study

EWA WOJCIULIK,<sup>1-3</sup> NANCY KANWISHER,<sup>2,3</sup> AND JON DRIVER<sup>4</sup>

<sup>1</sup>Department of Psychology, University of California, Los Angeles, California 90095; <sup>2</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge 02139; <sup>3</sup>Massachusetts General Hospital NMR Center, Charlestown, Massachusetts 02129; and <sup>4</sup>Institute of Cognitive Neuroscience, University College London, London WC1E 6BT, United Kingdom

**Wojciulik, Ewa, Nancy Kanwisher, and Jon Driver.** Covert visual attention modulates face-specific activity in the human fusiform gyrus: an fMRI study. *J. Neurophysiol.* 79: 1574–1578, 1998. Several lines of evidence demonstrate that faces undergo specialized processing within the primate visual system. It has been claimed that dedicated modules for such biologically significant stimuli operate in a mandatory fashion whenever their triggering input is presented. However, the possible role of covert attention to the activating stimulus has never been examined for such cases. We used functional magnetic resonance imaging to test whether face-specific activity in the human fusiform face area (FFA) is modulated by covert attention. The FFA was first identified individually in each subject as the ventral occipitotemporal region that responded more strongly to visually presented faces than to other visual objects under passive central viewing. This then served as the region of interest within which attentional modulation was tested independently, using active tasks and a very different stimulus set. Subjects viewed brief displays each comprising two peripheral faces and two peripheral houses (all presented simultaneously). They performed a matching task on either the two faces or the two houses, while maintaining central fixation to equate retinal stimulation across tasks. Signal intensity was reliably stronger during face-matching than house matching in both right- and left-hemisphere predefined FFAs. These results show that face-specific fusiform activity is reduced when stimuli appear outside (vs. inside) the focus of attention. Despite the modular nature of the FFA (i.e., its functional specificity and anatomic localization), face processing in this region nonetheless depends on voluntary attention.

## INTRODUCTION

Face-specific processing in primate extrastriate cortex provides perhaps the best example of a high-level visual “module.” Psychological accounts for such specialized modules (e.g., Fodor 1983) characterize them by the mandatory response putatively given whenever their triggering input is presented. Neuropsychologists similarly have argued that a dedicated face module becomes mandatorily engaged whenever a face is presented (e.g., Allison et al. 1995; Farah et al. 1995; Puce et al. 1996). However, face-specific neural responses in both monkeys and humans may depend on attention to face stimuli rather than being fully automatic. Covert attention is known to modulate neural responses at several levels of the visual system, both in monkey single-cell recordings (e.g., Maunsell 1995) and in human functional imaging (e.g., Corbetta et al. 1990; O’Craven et al. 1997), but the role of covert attention in the neural responses

to biologically significant stimuli such as faces has never been tested.

The present study tested whether face-specific activity in the human fusiform gyrus is reduced for stimuli presented outside the focus of attention. The fusiform face area (FFA) responds to faces in a highly selective manner as compared with other visual objects (Kanwisher et al. 1997; see also Allison et al. 1994; Haxby et al. 1991; Puce et al. 1996). Prior imaging data also suggest that the fusiform response to faces can be affected by the task given: it is stronger when subjects match faces than when they match colors (Clark et al. 1998) or locations (Haxby et al. 1994) for the same displays. However, subjects were not required to maintain fixation in this previous work, so the findings could be due to foveation of the faces only when they must be matched, effectively changing the visual input across tasks. Moreover, color or location matching does not require any analysis of visual shape, so previous findings might reflect modulation of shape responses in general, rather than of specialized face responses. Our study was designed to reveal whether *face-specific* fusiform activity is reduced when faces are presented outside the focus of covert attention, with retinal stimulation held constant, and with tasks that always required shape comparisons.

## METHODS

### Subjects

Eleven volunteers participated (4 women, 7 men; aged 17–38 yr). Excessive head-motion excluded one. Eight were right-handed, 1 left-handed, and 1 ambidextrous by self-report. All gave informed consent in procedures approved by Harvard University’s Committee on Use of Human Subjects, and Massachusetts General Hospital’s Subcommittee on Human Studies.

### Procedure

In *part 1*, subjects passively viewed sequences of gray-scale face photographs alternating with sequences of photographed common objects. The stimulus epochs (3 of faces and 3 of objects) lasted 30 s each, separated by 20-s fixation epochs. Within each stimulus epoch, 45 different photos of faces or objects were presented centrally (subtending a mean  $15 \times 15^\circ$ ) for 670 ms each. *Part 1* permitted us to localize each subject’s FFA individually as the region of fusiform gyrus responding more to faces than objects under passive viewing. This then served as a region of interest (ROI) for the attentional part of the study. [One subject’s FFA was localized by comparing faces vs. houses instead, which acti-

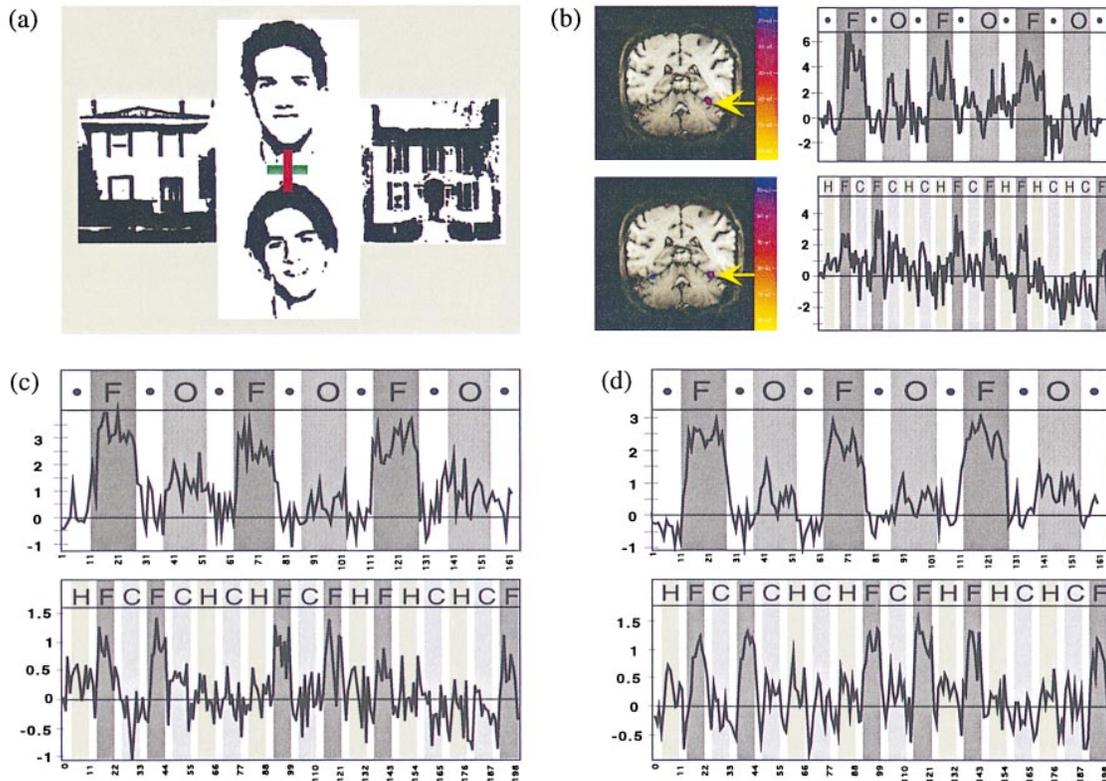


FIG. 1. *A*: sample stimulus display for the attention conditions. Subjects compared the 2 faces, the 2 houses, or the 2 arms of the cross while fixating centrally during the 200-ms exposure. *B*, *top left*: a coronal slice from 1 subject showing the fusiform face area (FFA) in the right hemisphere ( $\rightarrow$ )—the fusiform region that responded more strongly during passive viewing of faces than objects. Kolmogorov-Smirnov (KS) color-coded statistics are superimposed on the anatomic image. Raw percent signal change (PSC) time course of activity in the FFA is graphed *right* of the slice for successive face (F), object (O), and fixation ( $\cdot$ ) epochs for this subject. *Bottom left*: same slice now shows stronger activation in the predefined FFA region of interest (ROI) during active face matching (F) vs. house matching (H), with raw PSC time course again shown to the *right* (C: color matching). *C* and *D*: *top graphs* show mean PSC time courses in the FFA during the faces-objects passive-viewing localizer task; 6 subjects' left-hemisphere FFAs are in *C* and 8 subjects' right-hemisphere FFAs are in *D*. *Bottom graphs*: mean PSC time course in the predefined FFAs during the active attention conditions of matching faces (F), houses (H), or colors (C), averaged across the 6 left-hemisphere FFAs in *C*, and the 8 right-hemisphere FFAs in *D*.

vates the same region as faces vs. objects; see Kanwisher et al. (1997) for further details of these passive-viewing procedures.]

In *part 2*, subjects viewed sequences of displays each comprising two faces, two houses, and a colored fixation cross (Fig. 1A). The faces and houses were two-tone versions of photographs (2-tones allowed greater discriminability in the periphery); the cross was all red, all green, or had one red and one green arm. Note that the faces were different from those of *part 1* in terms of format (2-tone vs. gray-scale), size, and position. In separate epochs, subjects matched concurrent houses, faces, or colors of the cross, pressing a button whenever the relevant stimuli were the same (50% probability). Color matching tested an unrelated hypothesis and is not discussed further. Spatial arrangement was counterbalanced across conditions: faces were above and below fixation with houses to the left and right or vice versa. Displays subtended  $28 \times 21^\circ$ , appearing for 200 ms followed by an isolated cross for 800 ms. Each task epoch lasted 16 s (i.e., 16 trials) preceded by a 6-s visual instruction (the word HOUSES, FACES, or COLOR). There were 18 epochs per run, 6 each in the face (F), house (H), and color (C) matching conditions, ordered H-F-C-F-C-H-C-H-F-C-F-H-F-H-C-H-C-F. Total duration per run was 6 min, 42 s.

Subjects fixated centrally throughout, attending only *covertly* to the relevant stimuli. The 200-ms displays were too brief for deliberate saccades during their presentation. Moreover, eccentric fixation could not aid face or house matching because fixation on one item from the relevant pair made the other much less visible. Eye posi-

tion was recorded for three subjects during scanning by infrared tracker (Ober2, B1200, Permobil Meditech;  $2^\circ$  accuracy). Two showed no saccades; the third made only occasional saccades. All three showed the critical imaging results described below.

Anatomic and functional scans were performed by a 1.5 T GE Signa MRI (Milwaukee, WI), with echo-planar imaging (Instascan, ANMR Systems, Wilmington, MA), using a bilateral quadrature receive-only coil (made by Patrick Ledden). Functional images used an asymmetric spin-echo sequence (TR = 2, TE = 70 ms, flip angle =  $90^\circ$ ,  $180^\circ$  offset = 25 ms). The 10 6-mm-thick near-coronal slices (7 mm for 2 subjects) covered brain areas posterior to the brain stem. Voxel size was  $6$  (or  $7$ )  $\times$   $3.125 \times 3.125$  mm. A bite-bar minimized head motion.

Each subject underwent the passive faces-objects localizer, followed by four runs of the matching tasks and then another localizer. The two localizers were averaged and analyzed by Kolmogorov-Smirnov (KS) test after smoothing with a Henning kernel over a  $3 \times 3$  voxel area, producing an approximate functional resolution of 6 mm. After incorporating a 6-s estimated hemodynamic delay, the KS test was run on each voxel of the localizer data, to identify regions more active during face than object viewing. Each subject's FFA was defined as a minimum of two contiguous voxels in the ventral occipitotemporal fusiform region responding more strongly to faces than objects at  $P < 0.001$ . Areas posterior to the FFA were not pursued, as only the FFA has been shown to generalize across many tests of face-specificity (Kanwisher et al. 1997). Per-

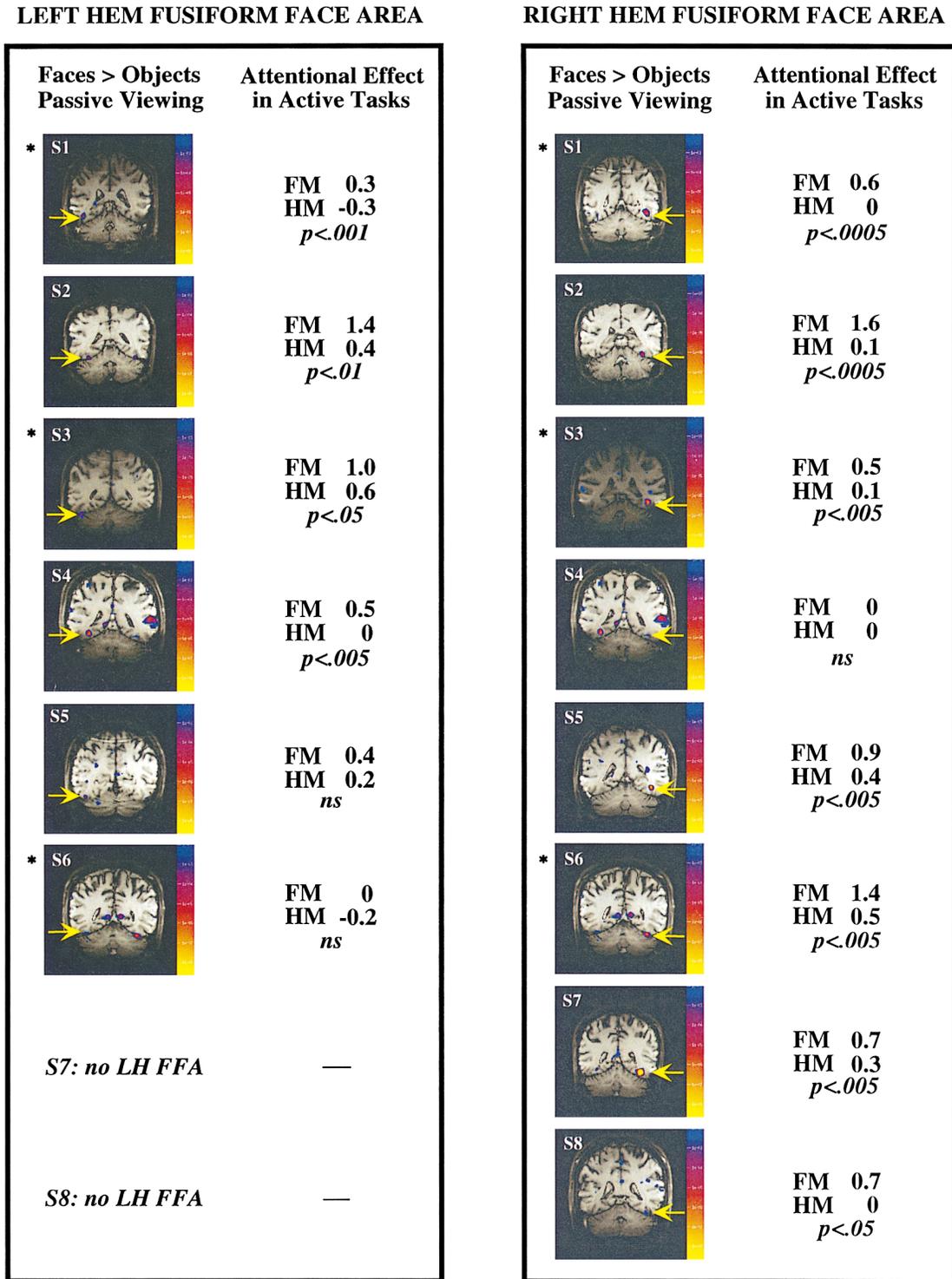


FIG. 2. Brain slices showing the left and right hemisphere FFAs (left and right columns, respectively) for each subject. These ROIs were defined with the passive faces vs. objects localizer test. Results of the subsequent attentional manipulation, with active tasks, for each predefined ROI are shown right of each slice, giving the PSC for face matching (FM) and house matching (HM), and the *P* value of the *t*-test comparing these PSCs during FM vs. HM epochs for each individual. Subjects whose eyes were monitored are marked by \*. S2 is left-handed; all others shown are right-handed.

cent signal change (PSC) was calculated separately for each subject's FFA, averaging over all functional data acquired during each condition.

The attention data were averaged first across four runs and then over all voxels within each subject's ROI (if a reliable ROI pre-

viously had been found for passive viewing). PSC for face- and house-matching epochs were calculated separately for each subject with instruction epochs as baseline. These data were analyzed with two separate paired *t*-tests: first, for each subject individually, PSC during the six face-matching epochs was compared with PSC for

the six house-matching epochs; second, for the group of eight subjects with reliable FFAs, mean PSCs for face matching were contrasted with those for house matching. Note that these analyses did not require corrections for multiple comparisons because they were carried out only in the ROIs, which were defined independently by the FFA localizer test.

## RESULTS

### FFA localizer

Eight subjects showed a region within right fusiform gyrus that responded more strongly to faces (PSC = 2.1) than objects (PSC = 0.6) at  $P < 0.001$  (PSC ratio = 3.5). Six of these subjects showed the same significant pattern in left fusiform gyrus: PSC for faces = 2.2; objects = 0.5; ratio = 4.4; see *top* time course graphs in Fig. 1, *B* (example subject), *C* (group: left-hemisphere), and *D* (group: right-hemisphere). See also brain slices shown in Fig. 2, for anatomic locus of each individually significant FFA.

### Attention conditions

For the eight subjects with reliable FFAs, behavioral accuracy (corrected for guessing) was 77% for face matching and 87% for house matching. Although the tasks did not show equivalent group performance, two subjects with equal behavioral scores across them also showed all the critical imaging results.

Seven of the eight subjects' predefined right FFAs individually showed more response during face than house matching [see *bottom* of Fig. 1*B* for sample time course, and Fig. 2 (*right panel*) for statistics]. The mean PSC across the eight subjects was 0.8 for face matching and 0.2 for house matching (PSC ratio: 4.9). This difference was reliable over subjects [ $t(7) = 4.57$ ,  $P < 0.005$ ; see mean time course graph at *bottom* of Fig. 1*D*].

Of the six subjects with predefined left FFAs, four showed significant attentional modulation individually (see Fig. 2, *left panel*). The mean PSC (across the 6 subjects) was 0.6 for face matching and 0.1 for house matching (PSC ratio: 5.6), and this difference was reliable over subjects [ $t(5) = 3.90$ ,  $P < 0.05$ ; see mean time course graph at *bottom* of Fig. 1*C*].

Note that the present data cannot allow us to determine whether the lower PSC for faces in the attention conditions versus the FFA localizer reflects the difference between two-tone peripheral faces versus grayscale foveal faces or is due to the simultaneous presence of houses in the attention conditions. However, for present purposes, the important point is that the predefined FFA was affected by covert attention toward faces versus houses in the latter conditions.

## DISCUSSION

These results provide the first demonstration that face-specific neural responses can be modulated by covert visual attention. There was significantly less activity in a predefined face-selective ROI within human fusiform gyrus when faces were outside rather than inside the focus of attention. Because central fixation was required, this attentional effect cannot be attributed to differences in retinal stimulation. In addition, the modulation we observed seems specific to face processing in particular, rather than to shape processing in

general, because both face matching and house matching required detailed shape comparison. Thus the face-selective response in this region is not produced in a strictly automatic, stimulus-driven fashion, but depends instead on the allocation of voluntary attention.

As in most physiological studies of attention (e.g. Moran and Desimone 1984; Treue and Maunsell 1996), our attended and unattended stimuli occurred in different spatial locations, raising the possibility that the modulation observed in the FFA might be mediated by an earlier spatial gating mechanism. However, our conclusion that the face-specific fusiform response depends on attention would still stand in this case. Moreover, preliminary results in our lab (Wojciulik and Kanwisher 1997) show similar attentional modulation of FFA activity even when faces and houses are presented in a rapid alternating sequence all at fixation so that purely spatial gating mechanisms could no longer select one category for the matching task.

Effects of covert attention on visual responses have previously been shown for several classes of stimuli, both in animal single-cell recordings (e.g., Moran and Desimone 1984; Treue and Maunsell 1996) and in human functional imaging (Corbetta et al. 1990; O'Craven et al. 1997). However, all of these prior studies employed meaningless nonbiological stimuli (e.g., colored or moving bars), rather than biologically significant stimuli such as the faces used here. Moreover, it often has been argued that face perception in particular depends on specialized module(s) that respond in an obligatory fashion whenever their trigger-stimulus is presented (cf. Allison et al. 1995; Farah et al. 1995; Fodor 1983; Puce et al. 1996). The present results show that although faces still may be special in the sense that dedicated visual machinery exists for them, evidently their perceptual processing depends on attention, just as for other classes of stimuli.

We thank L. Gordon, O. Weinrib, S. Brandt, J. McDermott, J. Culham, and P. Ledden for help.

This work was supported by National Institute of Mental Health Grant MH-56037 to N. Kanwisher, a Biotechnology and Biological Sciences Research Council (UK) grant to J. Driver, and a Human Frontiers grant to N. Kanwisher and J. Driver.

Received 2 May 1997; accepted in final form 3 October 1997.

## REFERENCES

- ALLISON, T., GINTER, H., MCCARTHY, G., NOBRE, A. C., PUCE, A., LUBY, A., AND SPENCER, D. D. Face recognition in human extrastriate cortex. *J. Neurophysiol.* 71: 821–825, 1994.
- ALLISON, T., MCCARTHY, G., BELGER, A., PUCE, A., LUBY, M., KIM, J., SPENCER, D. D., AND BENTIN, S. What is a face? Electrophysiological responsiveness of human extrastriate visual cortex to human and animal faces, face components, and complex objects (Abstract). *Proc. Cognit. Neurosci. Soc.* 2: 49, 1995.
- CLARK, V. P., PARASURAMAN, R., KEIL, K., KULANSY, R., FANNON, S., MAISOG, J. M., UNGERLEIDER, L. G., AND HAXBY, J. V. Selective attention to face identity and color studied with fMRI. *Human Brain Mapping* 5: 293–297, 1997.
- CORBETTA, M., MIEZIN, F. M., DOBMEYER, S., SHULMAN, G. L., AND PETERSEN, S. E. Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* 248: 1556–1559, 1990.
- FARAH, M. J., WILSON, K. D., DRAIN, H. M., AND TANAKA, J. R. The inverted inversion effect in prosopagnosia: evidence for mandatory face-specific processing mechanisms. *Vision Res.* 35: 2089–2093, 1995.
- FODOR, J. *Modularity of Mind*. Cambridge, MA: MIT Press, 1983.
- HAXBY, J. V., GRADY, C. L., HORWITZ, B., UNGERLEIDER, L. G., MISHKIN,

- M., CARSON, R. E., HERSCOVITCH, P., SCHAPIRO, M. B., AND RAPOPORT, S. I. Dissociation of spatial and object visual processing pathways in human extrastriate cortex. *Proc. Natl. Acad. Sci. USA* 88: 1621–1625, 1991.
- HAXBY, J. V., HORWITZ, B., UNGERLEIDER, L. G., MAISOG, J. M., PIETRINI, P., AND GRADY, C. L. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J. Neurosci.* 14: 6336–6353, 1994.
- KANWISHER N., MCDERMOTT, J., AND CHUN, M. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17: 4302–4311, 1997.
- MAUNSELL, J.H.R. The brain's visual world: representation of visual targets in cerebral cortex. *Science* 270: 764–769, 1995.
- MORAN, J. AND DESIMONE, R. Selective attention gates visual processing in the extrastriate cortex. *Science* 229: 782–784, 1984.
- O' CRAVEN, K. M., ROSEN, B. R., KWONG, K. K., TREISMAN, A., AND SAVOY, R. L. Voluntary attention modulates fMRI activity in human MT/MST. *Neuron* 18: 591–598, 1997.
- PUCE, A., ALLISON, T., ASGARI, M., GORE, J. C., AND MCCARTHY, G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J. Neurosci.* 16: 5205–5215, 1996.
- PUCE, A., ALLISON, T., GORE, J. C., AND MCCARTHY, G. Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J. Neurophysiol.* 74: 1192–1199, 1995.
- TREUE, S. AND MAUNSELL, J.H.R. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382: 539–541, 1996
- WOJCIULIK, E. AND KANWISHER, N. Face-specific attentional modulation in the human fusiform face area. An fMRI study. *Soc. Neurosci. Abstr.* 23: 299, 1997.