J. Willard Marriott Library
University of Utah
Electronic Reserve Course Materials

The Fusiform "Face Area" is Part of a Network that Processes Faces at the Individual Level

Isabel Gauthier
Vanderbilt University

Michael J. Tarr
Brown University

Jill Moylan, Pawel Skudlarski, John C. Gore, and Adam W. Anderson
Yale School of Medicine

Abstract

■ According to modular models of cortical organization, many areas of the extrastriate cortex are dedicated to object categories. These models often assume an early processing stage for the detection of category membership. Can functional imaging isolate areas responsible for detection of members of a category, such as faces or letters? We consider whether responses in three different areas (two selective for faces and one selective for letters) support category detection. Activity in these areas habituates to the repeated presentation of one exemplar more than to the presentation of different exemplars of the same category, but only for the category for which the area is selective. Thus, these areas appear to play computational roles more complex than detection, processing stimuli at the individual level. Drawing from prior work, we suggest that face-selective areas may be involved in the perception of faces at the individual level, whereas letter-selective regions may be tuning themselves to font information in order to recognize letters more efficiently. ■

INTRODUCTION

Many extrastriate areas respond to images in a category-selective manner. For instance, Puce, Allison, Asgari, Gore, and McCarthy (1996) found that an area in the occipito-temporal and inferior-occipital sulci responds selectively to nonsense letter strings. They and others have also identified an area of the fusiform gyrus, which responds selectively to faces (McCarthy, Puce, Gore, & Allison, 1997; Allison et al., 1994; Haxby et al., 1994; Sergent, Otha, & MacDonald, 1992). Since then, numerous experiments have tested whether this selectivity is the result of various differences in image properties between faces and letter strings. The "fusiform face area" (FFA) was found to be selective for faces even compared to scrambled faces (with, nonetheless, different spatial frequency components than normal faces) as well as to a variety of nonface object classes (Kanwisher, McDermott, & Chun, 1997; Kanwisher, Stanley, & Harris, 1999). The face selectivity of the FFA appears to be relatively generic, with almost any image perceived as a face being an effective stimulus. However, despite numerous studies on the FFA, its computational role (as well as that of other category-selective areas) remains unclear.

Faces and letters are particularly important stimuli for which efficient routes to recognition, for example, by-passing general object processing, could be acquired through experience (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999a; Gauthier, Skudlarski, Gore, & Anderson, 2000). Thus, some category-selective areas may be involved in detecting exemplars from a category, for instance, whether faces or letters are present in the visual field, so that specialized routines may be engaged (i.e., for reading, or to recognize an individual or their facial expression). We refer to this as the "detection hypothesis." Arguments in favor of this hypothesis include the apparent automatic activation of category-selective areas. Specifically, the FFA exhibits category selectivity even during passive viewing (Kanwisher et al., 1997; McCarthy et al., 1997), as does the "parahippocampal place area," which is selective for scenes and spatial layouts (Epstein & Kanwisher, 1998). Automaticity may be expected from areas that would detect even unforeseen stimuli. Moreover, some stimulus manipulations with large effects on face recognition behavior do not appear to influence the selectivity of the FFA. For example, several groups have reported only small or no effect of face inversion in the FFA (Aguirre, Singh, &

D'Esposito, 1999; Haxby et al., 1999; Kanwisher, Tong, & Nakayama, 1998). This is surprising given that inversion dramatically impairs face recognition performance (Diamond & Carey, 1986; Yin, 1969), by preventing the use of configural cues thought to be acquired through expertise with faces (Rhodes, Brake, & Atkinson, 1993; Diamond & Carey, 1986). Because the inversion effect in the right FFA is much larger for two-tone than grayscale faces, Kanwisher et al. (1998) suggested that the FFA is involved in detecting faces and not in their identification (note, however, that if a stimulus is not perceived to be a face, it also cannot be processed as an individual face). Similarly, N200 potentials recorded in the extrastriate cortex during the presentation of letter strings are insensitive to word type (e.g., concrete nouns vs. unpronounceable nonwords, Allison et al., 1994), suggesting an early, prelexical processing stage through which word-like stimuli may be detected.

Some evidence point to a more complex role for face-selective areas. Small but replicable effects of inversion for faces can be obtained in the FFA (Gauthier et al., 1999a; Kanwisher et al., 1998), consistent with the fact that although impaired, performance with inverted faces remains largely better than chance (Diamond & Carey, 1986). Critically, a comparable inversion effect can be obtained for novel nonface objects ("Greebles")i n the FFA, and this effect increases with expertise (Gauthier et al., 1999a). Bird and car expertise have also both been shown to recruit the right FFA (Gauthier et al., 2000). In other studies (Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997; Gauthier, Tarr, Moylan, Anderson, & Gore, 1999b), the FFA was more active for subordinate-level processing of familiar nonface objects compared to basic-level processing of the same images (e.g., recognizing the same object as a pelican vs. a bird). Moreover, a brain area corresponding in location to the FFA shows priming for negative contrast faces previously seen in their normal versions (George et al., 1999). This effect is specific to famous faces, suggesting that retrieval of names and/or semantic information may be involved (Dubois et al., 1999). A final argument against the detection hypothesis is that in actuality, a relatively large amount of extrastriate cortex responds selectively to faces. The currently popular hypothesis of the face-selective area as a neural and cognitive module specialized for face perception may bias the discussion to a single area in the right fusiform gyrus (Kanwisher et al., 1997; McCarthy et al., 1997). However, the comparison of fMRI images acquired while subjects view faces as opposed to nonface objects (the comparison typically used to define the right FFA) activates a diverse set of areas in the temporal lobe, including regions within the left-middle fusiform gyrus (Gauthier et al., 1999a, Gauthier et al., 2000), bilaterally in the anterior fusiform gyrus (Gauthier et al., 1999a; Sergent et al., 1993) and an area in the occipital lobe (Dubois et al., 1999; Halgren et al.,

1999; Haxby et al., 1999). It is unlikely that such a large network is dedicated only to the relatively computationally simple task of face detection.

In this study, we consider areas consistently detected in the localizer comparisons that we perform routinely to define face-selective areas: the FFA as well as a face-selective region in the occipital lobe. Because this early visual area is defined using the same localizer as the FFA, we call it the "occipital face area" (OFA). This area can be seen quite clearly in figures showing the FFA in several studies (Dubois et al., 1999, Figure 2a; Halgren et al., 1999, Figure 3; Kanwisher et al., 1997, Figures 3–4). Surprisingly, although the OFA shows a selectivity to faces similar (albeit weaker) to the selectivity seen in the FFA, the OFA has not been considered as a category-specific processing module—authors typically assuming that such an early area may mediate only the simplest and presumably generic stages of object perception. On the other hand, because it is found among relatively early visual areas, the OFA is a prime candidate for a category-selective area concerned with detection.

There is a recent interest in trying to identify different face-selective areas or electrophysiological components with processing stages. For instance, George et al. (1999) suggest that different stages of face processing may be isolated along the occipito-temporal pathway: posterior areas involved in structural encoding of faces and more anterior areas of the fusiform gyrus responsible for recognition of well-known individuals. This proposal is clearly influenced by an earlier model of the cognitive architecture of face processing (Bruce & Young, 1986), which distinguishes structural encoding (the extraction of a3-D invariant representation of an individual) from later recognition stages, including access to names and semantic information. A similar proposal is advanced by Bentin, Deouell, and Soroker (1999), who describe an ERP component (the N170) that they think may reflect an early structural encoding stage for faces (but see Rossion et al., in press).

In contrast, the question asked here is simpler in that we do not attempt to distinguish structural encoding from person recognition, but consider an even earlier stage of processing.[1] We examine whether extrastriate regions can be found with responses that could reflect only the first stage necessary in a category-specific processing stream: that of detecting objects from the category. In so doing, we address the basic issue of the level of processing for faces at two different stages of the occipito-temporal pathway, in the OFA and FFA. Do these face-selective areas behave as face detectors or is their response compatible with processing at the individual level? As a control condition, we use single letters of different fonts (or typefaces, see Figure 1): a letter-specific area (LA) can be defined, and we examine the processing level for both faces and letters in both face- and letter-selective areas.
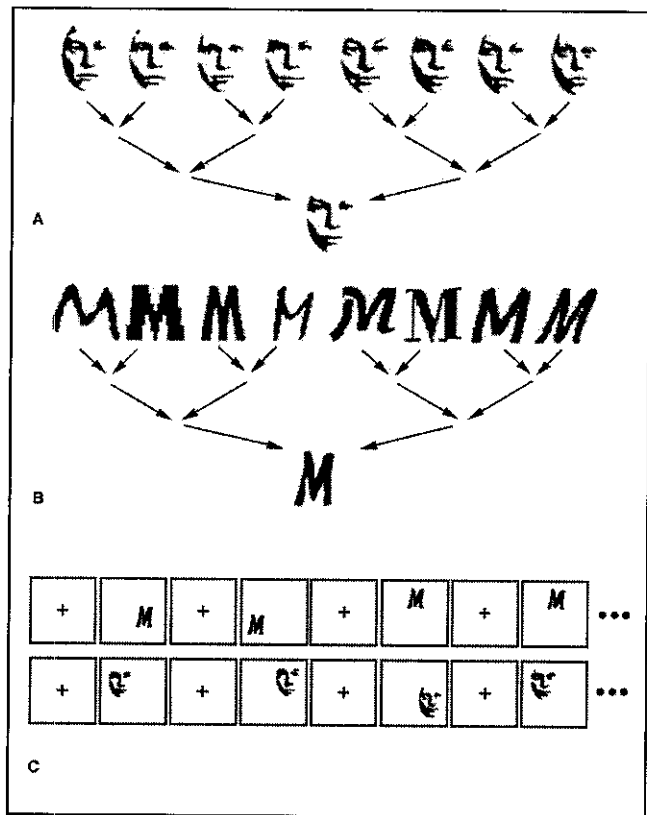
Figure 1. Stimuli and sample trials. Image averaging procedure as applied to (A) a set of eight faces and (B) a set of eight different fonts of the letter M. (C) The nonaverage stimuli or the resulting morphed average were shown in a random position on each trial while subjects attended to their location.

Subjects attended to the location of two-tone faces and letters (matching contrast as much as possible). A task in which subjects are instructed to ignore the objects' identity is most likely to activate potential category detectors without any further processing. In Repeated epochs, a single face or letter (morphed from eight different faces or fonts of the same letter) was used for 16 consecutive presentations, whereas in Different epochs, eight different faces or letters were shown twice each in a random order. A "detector" for faces or letters should not show any difference between these two conditions. In other words, it should be invariant to exemplar differences (Grill-Spector et al., 1999). In contrast, an area processing stimuli at the individual level should habituate more in the Repeated than in the Different epochs, although some habituation may still be expected in the Different condition (Gauthier & Tarr, 1997).

RESULTS

Face- and Letter-Selective Areas

Comparing all of the face versus letter epochs (regardless of condition) we defined three category-selective regions of interest (ROIs) in each hemisphere for each subject: an

FFA, an OFA, and an LA. Some ROIs were not present at the threshold selected for some subjects, but all were found in at least 14 of the 20 subjects. Examples are given for one subject in Figure 2. Note that at the arbitrary threshold selected ($t = 1$, at this threshold, all ROIs could be found in a majority of subjects, and the average size of the FFA was comparable to that reported in other studies), our ROIs may not include all category-selective voxels. Typical localizers use passive viewing, whereas our task involved attending to the objects' location (known to reduce the activation for faces, see Courtney, Ungerleider, Keil, & Haxby, 1996; Haxby et al., 1994), probably accounting for relatively more moderate selectivity compared to other studies. Our ROIs included only those seen in the majority of subjects and of theoretical interest as possible "category detectors." However, they are probably part of a larger computational network (see Haxby et al., 1999).

Following Haxby et al. (1999), who compared faces to houses, we use the terms "face-selective" and "letter-selective" only in reference to the difference between the two categories of stimuli. We do not wish to claim that the selectivity we observe would hold against all other possible stimulus categories. The locations of our ROIs (see Table 1) correspond to those reported in other studies for both the FFA (Kanwisher et al., 1997) and the region known to be activated by letter strings (Puce et al., 1996). Our letter-selective region overlaps with the region activated by letter strings (Puce et al., 1996), although it appears to be somewhat more posterior and larger (the extent of activation across studies and methods is difficult to
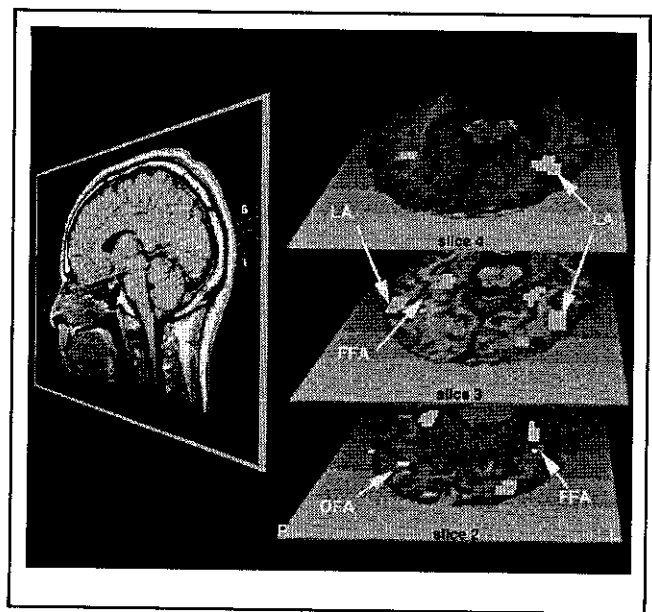

Figure 2. Slices covering the temporal lobe and ROIs for a single subject. The red to yellow voxels were more active during epoch showing faces while the blue to purple voxels were more active during epochs showing letters.

Table 1. Category-Selective ROIs

	N (/20)	Size (number of voxels ± SE)	Talairach coordinates x, y, z
Right FFA	18	9.4 ± 2	35, − 49, − 8
Left FFA	14	7.9 ± 2	− 35, − 56, − 6
Right OFA	19	41.1 ± 7	31, − 75, 0
Left OFA	18	30.4 ± 6	− 30, − 77, 0
Right LA	20	23.4 ± 5	50, − 59, 3
Left LA	20	25 ± 6	− 53, − 62, 3

compare). Here, we argue that this area habituates to font information (in order to facilitate access to letter identity). If this is the case, one would expect that in a study where letters were presented in a single font (Puce et al., 1996), habituation would lead to a smaller LA as compared to our study, in which letters were presented in eight different fonts.

Habituation in the Faces Conditions

The results of analyses over blocks during the presentation of faces in the Repeated and Different conditions are shown in Figure 3. No sign of habituation for faces was found in the LA. In contrast, repeated faces led to more habituation than different faces in the FFA and OFA, bilaterally. These observations are confirmed by a repeated-measures ANOVA conducted for each ROI, with a linear contrast applied to the factor of Blocks (four blocks per epoch) and the crucial effect being the interaction of a linear effect for Block with Condition (Repeated vs. Different). All effects reported are significant at $p < .05$ unless indicated. The main effect of Condition is not of interest, because the baseline in each condition consisted of the average of the four blocks. A main linear effect was obtained in the OFA and FFA bilaterally, but there was no significant linear effect in the LA $(F < 1)$. In all but one face-selective ROI (the right FFA), there was a significant interaction between a linear effect of Block and Condition. In the right FFA, this effect was very close to significant $(p = .051)$. No significant difference was found between the right and left FFA for this interaction.

Habituation in the Letters Conditions

The results of the analyses over blocks during the presentation of letters in the Repeated and Different conditions are shown in Figure 4. For both conditions, there was an overall pattern of habituation in nearly every ROI (suggesting that each of these areas may potentially have access to letter identities, e.g., distinguish an A from an E). The Repeated font condition

clearly led to greater habituation than the Different fonts condition in the right LA. A similar, but less clear, pattern was observed in the left LA and OFA.

ANOVAs revealed significant main linear effects of Blocks in all ROIs except for the right OFA $(F = 2.6, p = .12)$. A significant interaction of Block × Condition was found only in the right LA. The same effect was marginal in the left LA $(F = 3.02, p = .098)$. Interestingly, the analyses reported in the next section indicate that in both the left and right LA, there was more activation during Different than Repeated epochs for fonts.

In summary, the analyses over blocks reveal that two face-selective areas (OFA and FFA) habituate more to a repeated average face than to different faces. In contrast, their response for letters does not change over time, regardless of whether the font is repeated or not. Conversely, LA shows no change in response over time for faces, but does habituate more to letters of the same font than to letters presented in different fonts.

Different versus Repeated Epochs

As another way to measure a difference between the Repeated and Different conditions, the percent signal change was evaluated for each condition using the Repeated condition of the other category as a baseline. We compared both faces/different and face/repeated to font/repeated, whereas fonts/different and font/repeated were compared to face/repeated. Thus, any difference between the Repeated versus Different conditions for
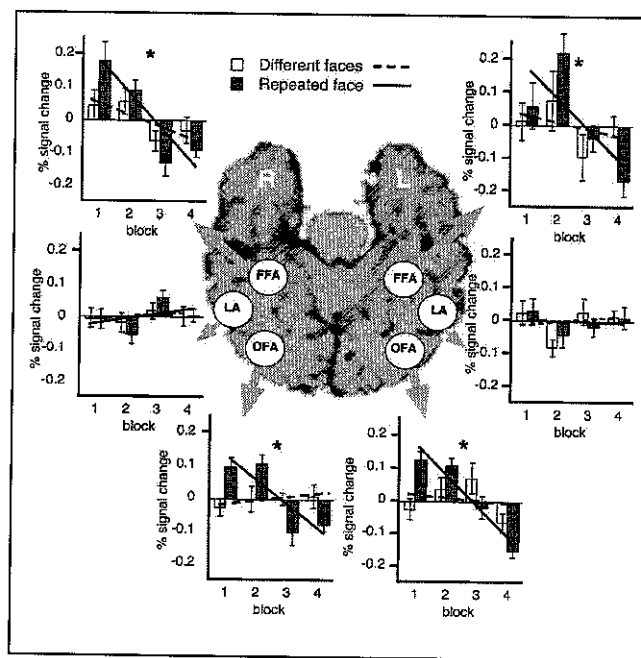

Figure 3. Habituation for faces during the Different and Repeated epochs in three different ROIs. The solid line represents the best linear fit for the repeated face condition while the dashed line shows the same function for the different faces conditions. Asterisks indicate significant Time × Condition interactions.
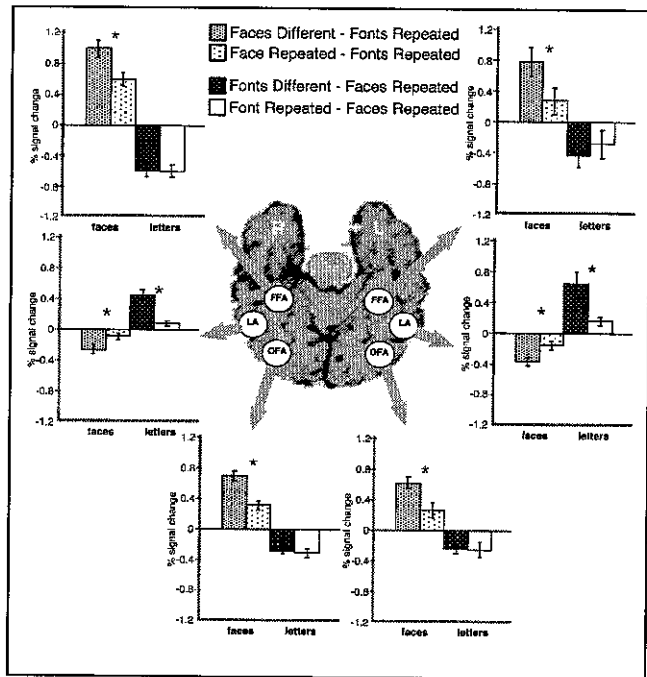
Figure 4. Habituation for letters during the Different and Repeated epochs in three different ROIs. The solid line represents the best linear fit for the repeated font condition while the dashed line shows the same function for the different fonts conditions. Asterisks indicate significant Time × Condition interactions.

these areas in the Different than the Repeated condition for faces (paired t tests, all $p < .02$), and no significant difference for letters (all $p > .20$). In the LA, there is greater activation during the Different than the Repeated condition for letters in both hemispheres ($p < .0001$), consistent with the habituation results. However, in this area, faces produced relatively greater activation in the Repeated than the Different condition (both $p < .02$). Because no difference was obtained in habituation for faces in this area, this effect apparently does not increase during an epoch. In other words, average faces elicit more activation than individual exemplar faces almost immediately and certainly within the first few presentations. One explanation is that the average faces are closer to a prototype or spatial norm for faces (Rhodes & McLean, 1990) or at least close to the average of the set of faces used in our study. We cannot isolate the effect of such a physical difference between the average and exemplar stimuli in areas that also show a difference in their habituation pattern in the Different versus Repeated conditions. However, in areas that do not show differential habituation for Different versus Repeated fonts (OFA and FFA) we find no indication of a difference between the Different and Repeated font epochs. We will consider possible interpretations of these findings in the Discussion.
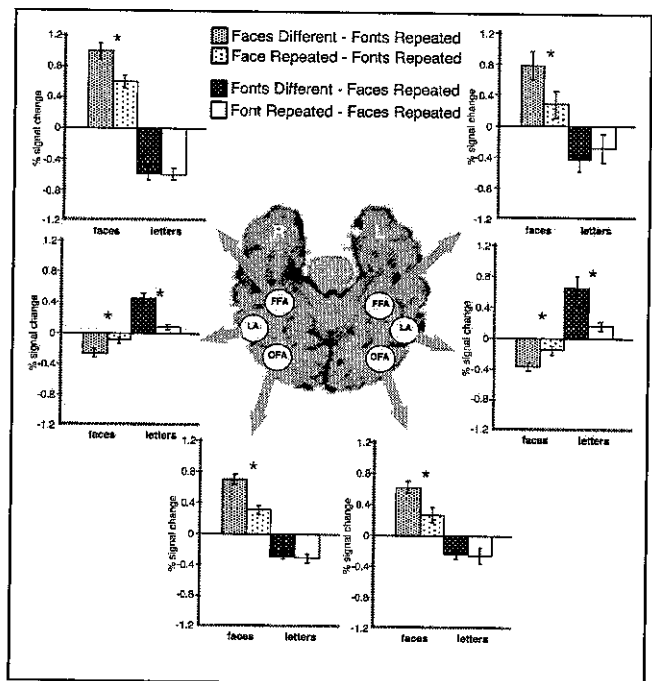
each stimulus class cannot be attributed to the baseline, common to both.

To some extent, this analysis should be consistent with the results of the analyses over blocks, in that if there is greater habituation during a Repeated than a Different epoch, there should be more activation in the Different as compared to the Repeated condition. However, these analyses are not entirely redundant, in that it is possible to obtain effects that are not mediated by habituation within an epoch. Specifically, an area may be sensitive to the difference between averaged and non-averaged stimuli because of perceptual differences rather than because of a level of processing (individual versus average) effect.

The results of this analysis in each ROI are shown in Figure 5. The percent signal change reported are relative to the repeated condition with the contrast category and not to a fixation baseline. Without such a baseline, we cannot quantify the extent to which each ROI responds to a category independently of the other. However, based on other studies (Gauthier et al., 2000; Kanwisher et al., 1999; Malach et al., 1995), both stimulus categories would be expected to activate these ROIs more than fixation. For example, negative value for faces in the LA does not indicate that this area shows a negative response to faces, but simply quantifies how much more it responds to letters than faces.

As in previous analyses, the FFA and OFA show a similar pattern of results: There is greater activation in

Figure 5. Mean percent signal change during the Different and Repeated epochs for faces and letters, relative to a baseline of the Repeated condition for the other category. Asterisks indicate significant differences (paired t tests, $p < .05$) between the two conditions, for a given category. Note that the (face/repeated minus font/repeated) and the (font/repeated minus face/repeated) comparisons are mirror images of each other.

One concern regarding the analyses over blocks is that if face-selective areas do not respond to letters and vice versa, there would be no response and habituation could not occur. However, our evidence indicates that a response was elicited for faces in letter-selective areas and vice versa. First, we found asi gnificant difference between the Repeated and Different conditions for faces in the LA. Second, we found asi gnificant linear effect indicating habituation over blocks (regardless of condition) for letters in the FFA and OFA. There was sufficient response for the less-preferred category in these areas to obtain significant effects over conditions or blocks, suggesting that the lack of habituation cannot be explained by an absent response.

DISCUSSION

The difference between the Repeated and Different conditions seems to reflect the automatic processing of faces and letters as a result of overtraining with these visual categories. Thus, the activity in the Different conditions for faces or letters would not have been qualitatively different even if subjects had been attending to the identity, rather than the location, of the objects. That is, ift he identity of a stimulus is processed automatically when subjects are attending to location, it should be processed at least equally well when subjects are attending to identity. However, attention to the identity of individual stimuli would likely have changed the pattern of habituation in the Repeated condition. In particular, if a subject were monitoring for changes in face identity, the FFA would maintain a more constant level of activity over trials with a repeated face.

Although we were interested in the automatic visual processing of objects, we used an active task—the location task—to assess the level of processing automaticity when attention is diverted from stimulus identity. In contrast, other authors (Kanwisher, Chun, McDermott, & Ledden, 1996) have argued that passive viewing designs provide a better measure of object perception. However, it is unclear which degree stimulus identity is attended to during passive viewing. The difference between passive and active tasks to study automatic processing of object identity appears to be a determining factor when comparing habituation studies. Using ERP, a method vastly different from fMRI in the temporal domain, Puce et al. (1999) found that N200 potentials to faces do not habituate to repeated presentations. Their subjects were not instructed to pay attention to a dimension other than identity and could have been attending to the identity of the faces. In an fMRI study, Epstein, Harris, Stanley, and Kanwisher (1999) found no difference in the FFA between a condition with different faces and a condition where many faces were repeated during the experiment (although the faces varied from trial to trial, unlike in our design).

In Epstein et al.'s study, subjects made 1-back identity judgments and thus attended to face identity.

FFA

Some authors have argued that the FFA is important for the recognition of individual faces (Haxby et al., 1994; Sergent et al., 1992), whereas others have advanced the possibility that the FFA may be involved in face detection only (Kanwisher et al., 1998). Given evidence that lesions in the vicinity of the FFA lead to deficits in individual face recognition (but not detection), it may not be particularly surprising that our results implicate face processing within the FFA at a more specific, and most likely individual, level.

The difference between our results and earlier studies highlights the computational role of the so-called "face-area." This area is strongly activated even by unfamiliar faces, and is, thus, unlikely to be specifically involved in the retrieval of names or other semantic information associated with a specific face. It does not show reduced activation when a face has been seen before (Epstein et al., 1999), and, perhaps, not even when the same face is repeated, but the subject attends to its identity (Puce et al., 1999). Therefore, the FFA is probably not involved in memory judgments per se, for instance, whether a person is familiar or not. On the other hand, when subjects attended to the location of faces rather than their identity, the FFA was more active when different faces are presented in succession than when the same face was repeated (even though all faces were eventually repeated within arun). This indicates that the FFA is involved in the processing needed to perceive a face at a more specific level. Thus, even when a face is familiar, its unexpected presentation should elicit this automatic processing. Moreover, the following prediction can be made: Whenever attention to the identity of as pecific face is elicited on agiv en trial (e.g., in a task where the same face appears on most of the trials, but subjects have to monitor for a new face), there should be no habituation. In addition, the role of the FFA in face perception appears to be sufficiently automatic that an abrupt change in the identity of the current stimulus face should elicit individual-level perception and no habituation relative to the preceding trial (even if that particular face has been seen before). On the other hand, the FFA should show habituation when subjects are not attending to face identity and the identity of the face is constant, as in our Repeated condition.

For nonface objects, the FFA may play a similar computational role as for faces, in a less automatic fashion—automaticity being one marker of expertise. When subjects match the same nonface objects with a label at the subordinate level (e.g., Pelican) versus a label at the basic level (e.g., Bird), the FFA shows increased activation (Gauthier et al., 2000; Gauthier et al., 1997). The FFA

also shows more activity for 1-back location judgments in epochs containing homogeneous stimuli (e.g., all birds or all cars) than in epochs containing objects from various categories (in novices, this effect is reduced for 1-back location judgments, Gauthier et al., 2000). Thus, even when we attend to the basic-level identity of a nonface object, the FFA is not maximally recruited—consistent with the idea that most objects are not perceived at the subordinate level automatically (Archambault, O'Donnell, & Schyns, 1999).

Why is the FFA strongly driven by the automatic processing of faces at the subordinate level? Our experience learning to individuate exemplars of this homogeneous class determines the level of categorization at which faces are preferentially processed (Tanaka & Taylor, 1991). A recent fMRI study of expertise with novel objects (Greebles) suggests that our vast experience with individual face recognition may have tuned the OFA and FFA (Gauthier et al., 1999a). After extensive training discriminating between individual Greebles, subjects (experts) showed more activation in their FFA for Greebles than nonface objects, even during passive viewing. In comparison, for novices the presentation of Greebles led to a response that was much more similar to any other nonface object. Thus, although automatic processing of faces in the FFA could be taken to indicate a "special" status for faces (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999), subordinate-level processing in the FFA is also automatic for nonface objects when subjects are experts for the objects in question (the same is true for bird and car experts, Gauthier et al., 2000).

In summary, the FFA appears to be involved in perceptual processing of visual information important for individual-level recognition and seems to be recruited whenever: (1) subjects attend to the identity of objects; or (2) stimuli are from a category for which subjects have extensive experience processing category members at the individual level. According to this model, it is not surprising that more specific information for letters (font) is not processed in the FFA automatically, particularly given our bias for ignoring this type of information when reading. However, our reading experience also biases us to automatically recognize individual letters (and words). This information appears to be available to some degree in the FFA (and OFA), since habituation was found in these areas for both font conditions equally. Perhaps expertise in processing font would lead letters to be processed at a more specific level of categorization in the FFA.

OFA

We defined an OFA, as observed in many prior studies (Dubois et al., 1999; Halgren et al., 1999; Haxby et al., 1999; Kanwisher et al., 1997). It is localized within the larger LO complex described by Malach et al. (1995), an area responsive to objects (including faces) more than

textures. Prior authors (Dubois et al., 1999; Haxby et al., 1999) suggested that this area is involved in detecting faces (i.e., perceiving faces at the basic level) or categorical distinctions such as gender or age (Sergent et al., 1992; Bruce & Young, 1986), but this interpretation has not been tested directly. Despite the robust co-occurrence of this region with the FFA, authors have been quick to reject the possibility that the OFA is specifically related to face processing. This dismissal probably rests on the assumption that relatively "early" visual areas should not be category-specific and that the OFA does not show a degree of selectivity to faces that is quite as large as that of the ubiquitous FFA. However, in all of our experiments designed to study the FFA and in which we returned to the data to define an OFA in the localizer task (faces minus objects), this area responded more to faces than objects, even those from as ingle homogeneous category such as cars.

LA

Finally, we defined an LA, which responded more to letters than faces (both in two-tone versions). According to Puce et al. (1996), this area is important in early prelexical, presemantic stages of word processing, but our results indicate that letter strings per se are not necessary to activate it. There is no evidence that this area is more important for words than letters, and none that it is not equally engaged by unfamiliar letter-like characters. This is relevant to many studies looking at the neural substrate of word recognition but which do no use control conditions with letters or nonsense characters (Kim et al., 1999).

We compared the response of the LA with that of the FFA/OFA network because our experience with letters is diametrically opposed to that which we have with faces. In face recognition, one typically needs to encode second-order configural information (Diamond & Carey, 1986), which makes it possible to recognize an individual; the first-order information, which defines faces at the basic level (eyes above nose, above mouth), is of no use in this regard. In contrast, many authors agree that abstract letter identity, independent of font and case, is used as the code for recognizing visually presented words (Peressotti, Job, Rumiati, & Nicoletti, 1995; Evett & Humphreys, 1994). In letter recognition, one typically needs to discard the second-order information in order to identify the underlying first-order information which defines the letter (e.g., two oblique lines joined at the top with an intersecting horizontal line); the second-order font information is of little use for telling an X from a V. These biases arise because of our different goals regarding faces and letters: If we only wanted to detect faces (for instance, because they were associated with danger) or if we wanted to remember individually each letter "A" that we encountered, our processing biases would be very different.

The right LA habituated more to a Repeated font than to Different fonts, and showed more activation overall in epochs where the font was varied, clearly indicating that it is sensitive to font information. This statement is agnostic with regard to this area's computational role(s). As discussed in the previous sections, we believe that analogous effects for faces in the FFA/OFA is associated with exemplar-level processing, because of converging evidence from other imaging and lesion studies. For letters, behavioral evidence suggests that during reading, the perceptual system tunes itself to the properties of a specific font in order to partial out this extraneous information. Thus, judgments on CONSISTENT FONT words are easier than those on M*I*x*E*d *F*o*N*T words (Sanocki, 1987, 1988). This process of font tuning (in essence, separating first-order from second-order information in order to access the first-order information more efficiently) could be mediated by the LA. The habituation observed for Repeated fonts would result from the fact that less tuning is required when successive letters are shown in the same font. According to this model, habituation should also be observed when different letters of the same font are shown in succession.

Surprisingly, the LA responded more to average than original faces, although it showed no habituation to either. Average faces may be closer to a representation of the "basic-level face," and the LA may encode basic-level representations for any category. For faces, there is no processing bias to discard individual-level information (as there is for letters), and individual exemplars may not activate a' 'basic-level" face representation as efficiently as an average. This account is obviously speculative. Our results nonetheless indicate that the computational puzzle of object recognition in extrastriate cortex is far more complex than faces simply being processed in the face-selective area, letters being processed in the letter-selective area, and so on for each geometrically distinct stimulus category.

Conclusions

The early impact of neuroimaging on understanding the ventral-temporal pathway was primarily to identify distinct visual areas, several of which behave in a category-specific manner. More recently, authors have begun to turn to the more difficult question of the computational role played by such areas and how they work in concert to produce behavior.

Significant effort has been expended investigating whether the "face area" is specifically dedicated to faces and whether this category preference is the product of nature or nurture. More recent studies have shown that the FFA is sensitive to experience (Gauthier et al., 1999a, Gauthier et al., 2000), level of categorization (Gauthier & Tarr, 1997; Gauthier et al., 1999a, Gauthier et al., 2000) priming (George et al., 1999) and other factors such as

attention (Wojciulik, Kanwisher, & Driver, 1998). Beyond asking what stimuli an area prefers, we can now start to ask what type of computations are performed and to what behaviors such computations contribute.

Here, we have demonstrated that the FFA is not likely to be involved simply in face detection, because it is sensitive to the individual identity of faces. In addition, we have found a face-specific area of the occipital lobe, which exhibits similar sensitivity to face identity. This area is also recruited by object expertise (Gauthier et al., 1999a, Gauthier et al., 2000). The implication is straight-forward: Models of face recognition should not be restricted to the FFA. Moreover, theories must expand beyond stimulus category differences to predict the circumstances under which the neural substrates recruited by faces are or are not engaged by nonface objects. One caveat to our conclusions is that our techniques do not allow us to distinguish the primary role of a brain area (e.g., OFA) from that arising in feedback from other areas (e.g., the FFA). Other techniques with better temporal resolution have the potential to resolve this issue.

In the case of reading, the representational levels involved are so varied (orthographic, phonemic, semantic, etc.) that the problem may be even more complex than for faces. In terms of the early visual stages of reading, prior studies have described brain regions more engaged by letter strings than faces or textures (Kim et al., 1999; Puce et al., 1996). Our results build onto these results and provide evidence for one of the earliest stages involved in reading words (Peressotti et al., 1995; Evett & Humphreys, 1994), that of accessing letter identity despite variations in font. Consistent with our thinking about face recognition, we do not believe that letter recognition requires a module specifically dedicated to a given category. For faces and letters, it is our hypothesis that the critical factors that lead to automatic processing are one's experience with the stimuli in question and particular recognition goals. In the end, both face and letter recognition are complex problems most likely mediated by a variety of neural substrates. To focus solely on the FFA (or its equivalent for letters) would be to belie the sophisticated computational network that is undoubtedly present in the primate brain.

METHODS

Subjects

Twenty young and healthy adult subjects participated in the experiments, approved by the Human Subjects Committee at the Yale School of Medicine.

Stimuli

Eighty grayscale pictures of faces, 40 males and 40 females, were cropped in an oval fashion to remove hair

and facial contour. Faces were organized randomly in 10 groups of eight same-sex faces and averaged together in a binary fashion (see Figure 1) using Morph software (Gryphon) using 41 anchors on each face. This produced an "average" face for each set of eight faces. All faces, original and average, were subsequently filtered as two-tone images. A similar method was used for images of 10 capital letters (A, B, E, G, K, M, R, S, T, Z) in eight different fonts (see Figure 1), with the exception that the original letters were already two-tone and a smaller number of anchors was used for letters than faces (from 10 to 20, depending on the letter). All final stimuli were included in a square area of about 10° by 10° of visual angle, and where shown in different positions on the screen within a square area of about 18° by 18° of visual angle.

fMRI Design

Each subject performed four runs, each lasting 216 sec. A run included twelve 18-sec epochs, three for each of four conditions in a counterbalanced order. The conditions were: Different faces, Repeated average face, Different fonts, and Repeated average font. An epoch consisted of 16 trials, each beginning with a fixation cross (145 msec) and followed by an object in a random position and shown for 980 msec. In Repeated epochs, the same average object was shown for 16 consecutive trials, while in Different epochs, eight stimuli were shown twice each, in a random order. Subjects performed a 1-back location judgment task during all conditions, monitoring for stimuli appearing consecutively in the same location regardless of identity (go–no go, 10% repetitions).

ROI Selection

Up to six ROIs (OFA, FFA, LA in each hemisphere) were defined in each subject, using the t maps for the comparison between all epochs showing faces and all epochs showing letters. An arbitrary threshold of $t = 1$ was used because it led to FFAs that were comparable in size with prior studies (Gauthier et al., 1999a, Gauthier et al., 2000; Kanwisher et al., 1997). The location, average size, and the number of subjects in which we were able to define each ROI are shown in Table 1.

fMRI Imaging Parameters and Analyses

Scanning was carried out on the 1.5 T GE Signa scanner equipped with resonant gradients (Advanced NMR, Wilmington, MA) using echo planar imaging (gradient echo single shot sequence, 144 images per slice, FOV = 40 × 20 cm, matrix = 128 × 64, NEX = 1, TR = 1500 msec, TE = 60 msec, flip angle = 60 msec). Six contiguous 7-mm–thick axial-oblique slices aligned along the longitudinal extent of the fusiform gyrus covered most of the temporal lobe. Voxels were, thus, $3.125 \times 3.125 \times 7 \text{ mm}^3$ in size.

Prior to statistical analysis, images were motion corrected for three translation directions and three possible rotations using the SPM-96 software (Wellcome Department of Cognitive Neurology, London, UK). Maps of t values and percent signal change, both corrected for a linear drift in the signal were created. Maps were spatially smoothed using a Gaussian filter with a full-width half-maximum value of 2 voxels. Twelve images were acquired during each epoch. After taking into account a hemodynamic lag of 3 sec, we discarded the first three images in each epoch in order to confidently remove effects of the previous condition (thus, the first image of the first block that was analyzed took place 7.5 sec after the beginning of the stimulation). For the analyses on the habituation effect, the remaining nine images per block were put into four time blocks (2, 2, 2, 3) and the percent signal change was calculated in each condition for each time block relative to the average of all nine images. For the Different versus Repeated analyses, the percent signal change was calculated using all nine images per epoch.

Note

1. Note that a similar method can be used to investigate more complex computational stages, for instance, whether there is an area that is invariant to rotations in depth for a single face, which would be evidence for the structural encoding stage proposed by Bruce and Young (1986).

REFERENCES

Aguirre, G. K., Singh, R., & D'Esposito, M. (1999). Stimulus inversion and the response of face object-sensitive cortical areas. *NeuroReport, 10,* 189–194.

Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Face recognition in human extrastriate cortex. *Journal of Neurophysiology, 71,* 821–825.

Archambault, A., O'Donnell, C., & Schyns, P. G. (1999). Blind to object changes: When learning the same object at different levels of categorization modifies its perception. *Psychological Science, 10,* 249–255.

Bentin, S., Deouell, L. Y., & Soroker, N. (1999). Selective visual streaming in face recognition: Evidence from developmental prosopagnosia. *NeuroReport, 10,* 823–827.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal Psychology, 77,* 305–327.

Courtney, S. M., Ungerleider, L. G., Keil K., & Haxby, J. V.

(1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex, 6,* 39–49.

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

Dubois, S., Rossion, B., Schiltz, C., Bodart, J. M., Michel, C., Bruyer, R., & Crommelinck, M. (1999). Effect of familiarity on the processing of human faces. *Neuroimage, 9,* 278–289.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron, 23,* 115–125.

Epstein, R., Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Evett, L. J., & Humphreys, G. W. (1994). The use of abstract graphemic information in lexical access. *The Quarterly Journal of Psychology, 33A,* 325–350.

Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P., & Gore, J. C. (1997). Levels of categorization in visual object recognition studied with functional MRI. *Current Biology, 7,* 645–651.

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

Gauthier, I., & Tarr, M. J. (1997). Orientation priming of novel shapes in the context of viewpoint-dependent recognition. *Perception, 26,* 51–73.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999a). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience, 2,* 568–573.

Gauthier, I., Tarr, M. J., Moylan, J., Anderson, A. W., & Gore, J. C. (1999b). The functionally-defined "face area" is engaged by subordinate-level recognition. *Cognitive Neuropsychology,* in press.

George, N., Dolan, R. J., Fink, G. R., Baylis, G. C., Russell, C., & Driver, J. (1999). Contrast polarity and face recognition in the human fusiform gyrus. *Nature Neuroscience, 2,* 574–580.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital cortex. *Neuron, 24,* 187–203.

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B. H., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human Brain Mapping, 7,* 29–37.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex: A PET-RCBF study of selective attention to faces and locations. *Journal of Neuroscience, 14,* 6336–6353.

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron, 22,* 189–199.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Science, U.S.A., 96,* 9379–9384.

Kanwisher, N., Chun, M. M., McDermott, J., & Ledden, P. J. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research, 5,* 55–67.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *NeuroReport, 10,* 183–187.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effects of face inversion on the human fusiform face area. *Cognition, B68,* 1–11.

Kim, J. J., Andreasen, N. C., O'Leary, D. S., Wiser, A. K., Boles Ponto, L. L., Watkins, G. L., & Hichwa, R. D. (1999). Direct comparison of the neural substrates of recognition memory for words and faces. *Brain, 122,* 1069–1083.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., Tootell, R. B. (1995). Object related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Science, U.S.A., 92,* 8135–8138.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9,* 605–610.

Peressotti, F., Job, R., Rumiati, R., & Nicoletti, R. (1995). Levels of representation in word processing. *Visual Cognition, 2,* 421–450.

Puce, A., Allison, T., Asgari, M, Gore, J. C., & McCarthy, G. (1996). Face-sensitive regions in extrastriate cortex studied by functional MRI. *Neurophysiology, 74,* 1192–1199.

Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cerebral Cortex, 9,* 445–458.

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

Rhodes, G., & McLean, I. G. (1990). Distinctiveness and expertise effects with homogeneous stimuli: Towards a model of configural coding. *Perception, 19,* 773–794.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R, Linotte, S., & Crommelinck, M. (in press). The N170 occipi-to-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *NeuroReport.*

Sanocki, T. (1987). Visual knowledge underlying letter perception: Font-specific, schematic tuning. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 267–278.

Sanocki, T. (1988). Font regularity constraints on the process of letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 472–480.

Sergent, J., Otha, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. *Brain, 115,* 15–36.

Tanaka, K., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457–482.

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Modulation of activity in the fusiform face area by covert attention: An fMRI study. *Journal of Neurophysiology, 79,* 1574–1578.

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