17 Does the Human Brain Process Objects of Expertise Like Faces? A Review of the Evidence

Elinor McKone and Nancy Kanwisher

In the quest for homologies between monkey and human cognition, domain-specific mechanisms are excellent candidates: highly specialized processors that operate on specific kinds of information, that develop early, and that are likely to be evolutionarily conserved. Here we explore one of the strongest contenders for a domain-specific processor, the human system for the perceptual analysis of faces.

Substantial evidence supports the domain specificity of face processing in humans. First, behavioral data indicate a different style of cognitive processing for faces (configural or holistic) than for other objects (feature- or part-based); for example, effects of orientation inversion are much more severe for faces than for other objects (e.g., Yin, 1969), and parts appear to be particularly strongly integrated into wholes in upright faces (e.g., Tanaka & Farah, 1993). Second, brain injury can produce selective deficits in face recognition while leaving object discrimination intact (McNeil & Warrington, 1993; Sergent & Signoret, 1992; de Renzi, 1986; Wada & Yamamoto, 2001) and vice versa (Moscovitch, Winocur, & Behrmann, 1997). Third, brain imaging (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997) and event-related potential (ERP) research (Bentin et al., 1996) have found neural responses that are highly specific to faces (see also Desimone, Albrignt, Gross, & Bruce, 1984). These findings support the "face specificity" hypothesis, according to which the cognitive and neural mechanisms underlying face processing are selectively engaged in perceptually processing faces per se and play little if any role in the perceptual analysis of nonface stimuli.

However, the face specificity hypothesis has become the subject of considerable debate. In the alternative "expertise" hypothesis, it is argued that mechanisms for face processing are not engaged only by faces, but are also applied in expert within-class discrimination of nonfaces (Diamond & Carey, 1986; Gauthier & Tarr, 1997). The idea here is that faces all share a basic "first-order" configuration, that is, the same parts are arranged in the same fixed layout (e.g., two eyes above a central nose), but recognition requires discrimination of individuals (*Bill* vs. *John*) who differ only in "second-order" deviations from this shared basic structure. In contrast, in most object recognition situations only between-class discrimination is required (*dog* vs. *bird*), for which parts and their first-order configurations are sufficient for identification. According to the expertise hypothesis, then, face recognition seems special only because adult humans have had substantial experience discriminating individual faces, but almost no practice at making similar within-class discriminations about objects. This view predicts that, for those rare people who have developed expertise in within-class object discrimination—such as dog-show judges capable of discriminating one Scotch terrier from another (Diamond and Carey, 1986)—the same mechanisms will be engaged in the processing of faces and of objects-of-expertise.

The debate between the face specificity and expertise hypotheses has now been active for some years. The aim of the present chapter is to provide a critical review of the current empirical evidence on the issue. We focus here on the two key predictions of the expertise hypothesis that have received most attention: that experts with nonface objects (1) should show behavioral signatures of face processing when processing nonface objects of expertise, and (2) should engage putatively face-specific neural mechanisms when perceptually processing objects of expertise. As our review of the literature shows, there is currently little convincing evidence for these predictions, and considerable evidence against them. Thus current evidence favors the face specificity hypothesis, although several key experiments remain to be conducted. We do not address here the important but distinct question of the developmental origins of face processing mechanisms, that is, the relative roles of genetic and experiential factors in their development.

Behavioral Evidence

How Faces Are Special in Human Behavior: Configural Processing

In behavioral terms, the special style of cognitive processing that occurs for faces is variously referred to as configural, holistic, or relational processing. There is little agreement on the exact nature of this type of processing (Maurer, Le Grand, & Mondloch, 2002), but the general idea is that either there are interactions between multiple parts over a broad region of the face ("configural," cf. Rhodes, 1988), or there is no decomposition into parts at all ("holistic"; Tanaka & Sengco, 1997), beyond perhaps simple lines and edges of early vision (Moscovitch, Winocur, & Behrmann, 1997). It is also presumed that configural processing includes coding of detailed distance information, such as the spacing between the eyes (e.g., "relational," Diamond & Carey, 1986). For present purposes, we employ the term *configural* and do not distinguish between alternative definitions. Configural processing is usually contrasted with "part-based" processing, which is presumed to involve some sort of decomposition at obvious part boundaries, plus some coding of basic relations among these parts (above, below, to-the-side-of, etc).

Initial behavioral evidence for special processing of faces came from the *disproportionate inversion effects* in recognition memory. All objects are remembered worse inverted than upright but inversion effects are much larger for faces (25 percentage point decrement) than for a wide range of other object classes (2–10 points)(Yin, 1969; Diamond & Carey, 1986; Scapinello & Yarmey, 1970). This finding occurs despite the memory-task requirement for within-class discrimination in all cases.

Other methods directly demonstrate configural processing of upright faces. In the *composite effect* (Young, Hellawell, & Hay, 1987), it is difficult to name one half of a face formed from a composite of two different individuals when the two halves are aligned (indicating automatic perceptual integration into a new whole), in comparison to a control condition in which the two halves are offset (figure 17.1). In the *part-whole effect* (Tanaka & Farah, 1993), memory for a face part (Bill's nose) is much poorer in isolation (Bill's nose vs. John's nose) than in the context of the original whole face (Bill's nose in Bill's face vs. John's nose in Bill's face). In the *whole-vs-configurally-transformed-whole* version of the part-whole paradigm (Tanaka & Sengco, 1997), memory for a part is worse in the context of a configurally altered version of the original face (Bill's face with the eyes shifted apart slightly) than in the original face. In other *relational alteration paradigms*, perceived bizarreness, perceived distinctiveness, and memory are substantially affected by altering distances between natural face parts (shifting eyes apart, or the mouth down; Bartlett & Searcy, 1993; Leder & Bruce, 1998; Le Grand, Mondloch, Maurer, & Brent, 2001; Rhodes, Brake, & Atkinson, 1993).

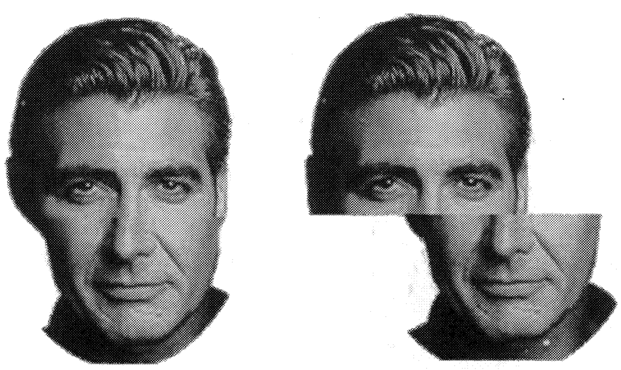

Figure 17.1
The composite effect of Young et al. (1987). To name the top half as George Clooney (or the bottom half as Harrison Ford) takes approximately 200 ms longer in the aligned version on the left than in the misaligned version on the right. Given the two conditions are matched for simple response competition (the to-be-ignored half is always a different individual), this effect must reflect perceptual integration of the two halves into a new whole in the aligned condition. The composite effect does not occur when the stimuli are shown inverted.

In all the studies cited above, configural effects occurred for upright faces but were absent or nearly absent for inverted faces. To the extent that the paradigms have been applied to objects, nonexperts have also demonstrated little or no indication of configural processing (houses, dogs, car fronts, and biological cells) in the part-whole paradigm (Tanaka & Sengco, 1997, Tanaka et al., 1996, cited in Tanaka and Gauthier 1997).

Behavioral Findings: Laboratory-Trained Experts

One approach to expertise has been to test laboratory-trained subjects, using an artificial class of objects called "greebles" (figure 17.2). Participants are trained to identify many greebles over several sessions (8–10 hr) until reaction times (RTs) for individual identity decisions are as fast as those for "family" and "gender" decisions. We have general theoretical concerns about experiments involving greebles. Any *failure* to find evidence of facelike special processing with greebles is not conclusive, given the relatively small amounts of practice and the rather weak criterion for expertise (cf. 31 years to show face-sized inversion effects in behavioral studies; see Diamond & Carey, 1986, below). Equally, however, any *positive evidence* of "face-specific" processing for greebles can be inconclusive. This is because greebles have a high degree of structural similarity to faces and/or face-body combinations (figure 17.2). Thus small greeble expertise effects could reflect a specialized face-processing mechanism learning to stretch its definition of a face, rather than a generic expertise effect that could occur for any object class. However, because others are taking the results of greeble studies as relevant to the debate, we will consider them in detail here.

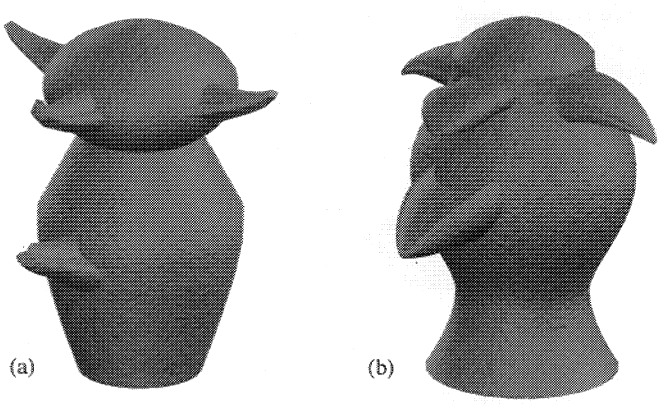

Figure 17.2
Greebles have two horizontally arranged parts above two centrally arranged parts, replicating the T-shaped configuration of eyes, nose and mouth. Perceptually, most appear to have heads (*a*), and in some cases the whole greeble looks like a face (*b*). (Images provided courtesy of Michael J. Tarr, Brown University.)

Several studies have examined behavioral performance in subjects trained with Greebles. The authors of those studies have summarized their findings as having "compared a wide range of putatively face-specific behavioral effects across Greeble novices and Greeble experts and . . . consistently obtained face-like patterns of performance with experts but not novices" (Tarr, 2003), or as "suggest[ing] that subjects shifted from feature-based to more configural processing as they became greeble experts" (Gauthier & Tarr, 1997; Gauthier et al., 1999). Gauthier and Tarr have thus argued strongly for the expertise hypothesis.

Our own review of the greeble papers, however, indicates that the evidence is in fact very weak. The essential problem is that Gauthier and Tarr have focused on a few close-to-significant positive findings, and have ignored equally important results that were either null or in the reverse-to-predicted direction. To explain this, it is necessary to go into some detail and describe the results in full.[1]

In the *standard part-whole paradigm*, Gauthier and Tarr (1997) found a small whole-over-part advantage for both novices (6 percentage points) and experts (11 percentage points).The effect reached significance only for experts but, at the same time, there was no interaction involving expert vs. novice status. In Gauthier, Williams, Tarr, and Tanaka (1998), experts showed a significant whole-over-part advantage for one of three different parts tested, but so did novices. Experts showed trends in the reverse-to-predicted direction for the other two parts. Averaged across all parts, the mean part-whole difference trended in the wrong direction for an expertise effect (7 percentage points for novices, vs. 0 percentage points for experts). In Gauthier and Tarr (2002), the size of the part-whole difference also failed to increase with expertise across five training sessions (d′ difference in session 1 = .76, session 5 = .68).

Comparing *part-in-original-whole* with *part-in-configurally-transformed-whole*, Gauthier and Tarr (1997) found a significant difference for experts on reaction time (although there was no effect at all on accuracy), with a close-to-significant interaction involving expert-vs.-novice status. While these results were suggestive of an expertise effect, further studies failed to confirm the finding. In Gauthier, Williams, Tarr, and Tanaka (1998) only one part showed a close-to-significant effect, while the other two parts showed trends in the reverse-to-predicted direction. Averaged across all parts, the mean effect was 0 percentage points for novices and 0 percentage points for experts (cf., 5–11 points for faces; Tanaka and Sengco, 1997). Gauthier and Tarr (2002) again tested all three parts, on both accuracy and RT. In experts, there was one significant effect in the predicted direction (RT for the quiff), but there was also another significant effect in the reverse-to-predicted direction (accuracy for the dunth). Averaged over all parts, the size of the effect in d′ accuracy was 0.69 in session 1 (novices) and 0.64 in session 5 (experts).

In the *composite test*, Gauthier, Williams, Tarr, and Tanaka (1998) found that, for composites made of same-family halves, there was a nonsignificant trend in the predicted direction (i.e., aligned more difficult than unaligned) on reaction time in greeble experts, and no effect on accuracy. For composites made of different-family halves, trends were in the reverse-to-predicted direction on both accuracy and reaction time. In Gauthier and Tarr (2002), after five practice sessions, there was no indication that experts had developed a composite effect. On reaction times, the aligned vs. unaligned difference was 12 ms (in the predicted direction, but with SEMs for each condition = 35 ms) and the difference on d′ accuracy was .31 in the reverse-to-predicted direction. Statistics showed a close-to-significant interaction with degree of training on reaction times (there was no effect on accuracy), but this reflected a peculiar finding that the composite effect started close to zero, went in the reverse-to-predicted direction in sessions 2 and 3, and then returned to zero.

In interpreting these findings, Gauthier and Tarr (2002) accept that there is *no part-whole effect* for greeble experts. Rather than take this as evidence against configural processing, however, they choose to define the standard part-whole test as providing a general test of contextual influences rather than a test of configural processing. We see no theoretical basis for this idea. We agree that most objects show some advantage for parts in the context of wholes rather than in isolation (and the same is true for words; i.e., "the word superiority effect" [Reicher, 1969; Wheeler, 1970]), but this ignores the fact that the part-whole effect remains very much greater for faces than it is for objects. Gauthier and Tarr do accept the *whole-vs.-configurally-transformed-whole* test and the *composite paradigm* as tests of configural processing. However, these measures do not show evidence of configural processing in greeble experts, as discussed above.

We also note that most greeble studies have not explicitly tested for *disproportionate inversion effects* in experts. Two experiments including inverted greebles do not suggest any unusually large (i.e., face-sized) inversion effects. In part identification in the context of the whole studied stimulus, greeble experts showed only a 4-percentage-point decrement with inversion (87 percent correct upright, 83 percent inverted [Gauthier & Tarr, 1997]); this compares with an 18-percentage-point decrement previously reported for faces (87 percent upright, 69 percent inverted [Tanaka & Sengco, 1997]). Second, in naming whole greebles that were learned upright, greeble experts showed a misorientation effect (60, 120, and 180 degree image-plane rotations were combined) that trended in the reverse-to-predicted direction for an expertise effect (RT difference between upright and misoriented = 180 ms for experts, 231 ms for novices [Gauthier, Williams, Tarr, & Tanaka, 1998]).[2] In a direct test, Rossion et al. (2002) found a larger inversion effect after training with greebles (46 ms) than before (25 ms), although the effect remained smaller than for faces (75 ms).

Behavioral Findings: Real-World Experts

As mentioned above, lab training studies provide only around 10 hours of experience, which is very different from the lifetime of experience all of us have with faces. In another approach to investigating expertise which gets a little closer to the situation with faces, several studies have tested real world experts, such as dog show judges with 5, 10, 20, or more years of experience in individuating exemplars of dogs within their particular breed-of-expertise.

In an early indication that configural processing might develop for objects-of-expertise, Diamond and Carey (1986) found that dog show judges (with 31 years experience) showed face-sized inversion effects on memory for their breed of expertise. When a wider range of breeds were used, the inversion effect for dogs remained smaller than that for faces, and the inversion × expertise interaction did not reach significance. These results were taken to indicate that facelike processing of objects (a) was possible and (b) took many years of extensive experience to develop. We note, however, that memory in general is better when appropriate preexisting knowledge can be applied to the to-be-learned material (e.g., soccer experts show better memory for text describing a soccer game than do novices; Schneider, Korkel, & Weinert, 1989). Thus, even though Diamond and Carey's dog experts were clearly applying some sort of perceptual knowledge better to upright dogs than to inverted dogs, it is not necessarily the case that this knowledge was configural in nature. However, a recent study using a perceptual task (sequential matching) found face-sized inversion effects for discrimination of body configuration (Reed et al., 2003), which could reflect either widespread expertise for bodies, or the engagement of special cortical mechanisms for perceiving bodies (Downing et al., 2001; see also Tsao et al., 2003). Finally, the current evidence on perceptual inversion effects for cars in car experts is ambiguous: although both Gauthier et al. (2000) and Xu et al. (2002) found slightly larger differences in d′ for sequential matching of upright versus inverted cars in car experts than control subjects (d′ = 2.4 upright versus 1.6 inverted for experts, 1.4 upright versus 0.8 inverted for nonexpert controls, interaction not significant for Gauthier et al.; 2.9 upright versus 2.0 inverted for experts, 1.2 upright versus 0.7 inverted for controls, interaction significant for Xu et al.), scaling differences make the interpretation of these data difficult, and indeed ratios of upright to inverted d′ go in the opposite direction in both studies (larger inversion effects for nonexperts than experts). It would clearly be useful to resolve these many ambiguities by measuring perceptual inversion effects in real-world experts when scaling effects are eliminated.

Finally, the one study that included a direct behavioral test of configural processing on real-world experts[3] found that objects-of-expertise are *not* processed configurally. Tanaka et al. (1996; cited in Tanaka & Gauthier, 1997) used the part-whole effect and found no influence of expertise (a minimum of 5 years, and mostly more than 10 years) for cars, biological cells, and Rottweiler dogs. Specifically, for each object class, a small whole-part advantage was of a similar size in experts (mean = 8.0 percentage points) and novices (8.3 percentage points) and both groups showed a far smaller effect for objects than for faces (23 percentage points).

Summary of Behavioral Studies

Currently, there are no behavioral data that unequivocally demonstrate "facelike" configural processing for objects-of-expertise. The failure to find such effects in laboratory-trained greeble experts does not seriously damage the expertise hypothesis, in that 10 hrs of training might not be expected to produce sufficient expertise. Yet, even in real-world experts, the one study that tested directly for configural processing failed to find any configural effect in the part-whole paradigm (Tanaka et al., 1996, cited in Tanaka & Gauthier, 1997). This leaves large inversion effects on memory tasks as the only result suggestive of an expertise effect (and, as discussed earlier, this could result from properties of long-term memory rather than properties of object processing). Most important, even clear behavioral evidence that objects of expertise are processed like faces would leave open the important question of whether this similarity reflects engagement of the very same mechanisms by faces and by objects of expertise, or the engagement of distinct mechanisms with similar functional properties. Neural measures are well suited to distinguish these alternatives.

Neural Evidence

How Faces Are Special in Neural Processing: The FFA, N170, and M170

Several face-selective neural responses have been reported based on research using fMRI, ERPs, and MEG in ordinary people—that is, those not expert in any particular object domain. Most relevant to the expertise debate are the fusiform face area (FFA) in fMRI studies, and the N170 and M170 responses measured with event-related potentials (ERPs) and magnetoencephalography (MEG), respectively.

The FFA is a cortical region in the fusiform gyrus that responds twice (or more) as strongly in fMRI when subjects view faces than when they view any other class of visual stimuli yet tested, even when the object task is more difficult than the face task, and even when all members of the object class (e.g., hands) share a basic configuration (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Tong et al., 2000). Although the FFA shows only weak face inversion effects (Kanwisher, Tong, & Nakayama, 1998), the magnitude of the fMRI signal from this region is correlated on a trial-by-trial basis with successful identification of faces (Grill-Spector et al., 2004). Evidence that this region is not only activated during face recognition but is also critically involved in face recognition comes from the study of a patient with a very small lesion in this region who suffered profound prosopagnosia but no detectable deficit in object recognition (Wada & Yamamoto, 2001; see also Barton, Press, Keenan, & O'Connor, 2002; Puce, Allison, & McCarthy, 1999; Mundel et al., 2003)

The *N170* is an ERP component occurring about 170 ms after stimulus onset over occipitotemporal sensors (Bentin et al., 1996; Jeffreys, 1996). Its amplitude at many scalp sensors is larger for faces than for a wide range of other objects (Carmel & Bentin, 2002). A face-specific inversion effect has been reported for the N170: for faces but not other objects, the N170 is delayed by 10 ms for the inverted orientation compared to upright (Rossion et al., 1999; Rossion et al., 2000). Similar properties have been reported for the face-selective "M170" response recorded with MEG (Liu, Higuchi, Marantz, & Kanwisher, 2000). It is not yet clear whether the N170 and M170 arise from the same cortical source, or whether the source for either is the FFA.

Neural Findings: Laboratory-Trained Experts

Gauthier et al. (1999) scanned subjects looking at faces and greebles. They found that activation for upright minus inverted greebles in the FFA region increased throughout greeble training. While Gauthier et al. interpreted their data as evidence for an expertise effect in the FFA, there are several problems with this conclusion. First, rather than measure the percent signal change from baseline for each stimulus type, they reported only the difference between upright and inverted orientations; this tells us nothing about the crucial question of the magnitude of response to upright greebles and upright faces after training. (This important distinction is easily lost; for example, Bentin and Carmel (2002) wrongly describe the Gauthier et al. result as "greebles recruited the FFA to nearly the same degree as faces do"). Second, the fMRI inversion effect is an odd choice as a marker of faceline processing because it is found only very weakly (Kanwisher, Tong, & Nakayama, 1998) or not at all (Haxby et al., 1999) for faces in the FFA. Third, the "FFA" was defined as a large square region of interest, over a cm on a side, a method that virtually guarantees the inclusion of voxels neighboring but not in the FFA. Finally, the "activation" was defined as the sum across the 64 voxels in this ROI of *t* values resulting from a comparison of upright to inverted responses within each voxel (after excluding all *t* values less than 0.1). This truncated "sum-of-ts" measure (see also Gauthier & Tarr, 2002) confounds an increase in signal change for upright stimuli after training with a reduction in variance. These problems leave the results difficult to interpret.

In an N170 study, Rossion and colleagues (2002) examined whether greeble training induced a facelike pattern of inversion effects. In terms of amplitude, greeble experts showed the predicted direction of effect in the left hemisphere (inverted minus upright greebles = 0.49 microvolts), but there was a similar-sized effect in the reverse-to-predicted direction in the right hemisphere (–0.50 microvolts). With respect to latency, there was a significant interaction of orientation (upright versus inverted) by session (pretraining versus posttraining) for greebles but not for faces. However, the effect for greebles (the latency delay for inverted versus upright stimuli) after training was significant only in the left hemisphere (5 msecs), not in the right (1.6 msecs; cf. 10 msecs for faces). Because no evidence was presented that the N170 recorded in this experiment was face selective, this study does not address the critical question of whether the same neural mechanisms are engaged in processing faces and objects of expertise. Indeed, given that behavioral, fMRI, and ERP markers for face processing are all right lateralized, the left-lateralized expertise effects in this study would seem if anything to provide evidence for a dissociation, not an association, between expertise and face processing.

Neural Findings: Real-World Experts

Words are probably the only stimulus class for which we have perceptual expertise that approaches our expertise for faces. Words produce very weak FFA responses (Puce et al., 1996), demonstrating that perceptual expertise alone is insufficient to strongly engage the FFA. However, words and letters do not share a first order configuration, and it remains possible that expertise has an effect only where this is the case.

Gauthier et al. (2000) reported greater FFA activation for cars and birds than control objects in car and bird experts (19 years experience), respectively. This result has been confirmed in an event related design (Xu & Kanwisher, 2001). Note, however, that the expertise effect is small and percent signal increase from fixation remains twice as large for faces as for cars in car experts in both studies. Further, although Gauthier et al. emphasize as their strongest finding the correlation across subjects between behavioral expertise for cars/birds and the FFA response to cars/birds, this was found only for fMRI data collected during performance of a location discrimination task (on objects of expertise); it is not clear why they found no such correlation between behavioral expertise and fMRI responses to objects of expertise during a task requiring discrimination of objects of expertise. In a third study, Rhodes et al. (2004) scanned lepidoptera experts and found little overlap between the regions in the fusiform gyrus that were activated by lepidoptera (versus objects) and those activated by faces (versus objects); they concluded that distinct neural populations are tuned to the two object classes. Finally, Grill-Spector, Knouf, & Kanwisher (2004) found a clear trial-by-trial correlation between success at individual face recognition and the magnitude of the rFFA response to faces in car experts, but no trial-by-trial correlation between subordinate-level identification of cars in the same subjects. These data suggest that the weak engagement of the rFFA by objects of expertise is not causally related to expert object identification.

In terms of the N170, Tanaka and Curran (2001) reported a higher N170 response for bird than dog stimuli in bird experts, and vice versa for dog experts (all had at least 10 years of experience, most had 20 years). However, the scalp location where these expertise effects were found was different from the usual site of the face-selective N170, and no evidence was presented that these sites produce face-selective N170s. Thus, there is no reason to think that the expertise effects on the N170 reported in this study reflect the same neural source as the face-selective N170. Another general problem was that the dogs and birds used in the Tanaka and Curran (2001) study had *faces*, and the N170 could reflect a response to these faces (enhanced by expertise) rather than generic expertise (Bentin & Carmel, 2002).

Another study of effects of expertise for cars on the N170 (Gauthier et al., 2003) reports a number of complex high-order interactions that are argued to be consistent with the expertise hypothesis, but fails report the results of the more straightforward prediction that the response to cars relative to control stimuli should be higher for experts than novices in face-selective sensors. Such an effect may be present in figure 3a of that paper, but this effect appears to be present only in the left hemisphere, not the right. Further, the paper mentions that different scalp distributions were found for the responses to cars and faces, suggesting a dissociation, not an association, between the processing of the two stimulus classes.

The only magnetoelectrophysiological study that actually tested the key prediction of the Expertise hypothesis—that the same neural response should show both face selectivity and enhanced responses for objects of expertise—decisively refuted the hypothesis. In particular, Xu, Liu, and Kanwisher (2004) found that the face-selective M170 was no greater for cars (relative to objects) in car experts than in control subjects; they also found that whereas the M170 amplitude was correlated trial-by-trial with successful face recognition, it was not correlated with successful car recognition.

Summary of Neural Processing Studies

Currently, the only real evidence of an expertise effect on either the FFA, N170, or M170 comes from the small but significant increase in the FFA response to objects-of-expertise in real world experts. According to the expertise hypothesis, the fact that car experts still show twice the FFA response to faces as to cars would be attributed to car experts being less experienced with cars than with faces. While this is one explanation, there are others. For example, because attention enhances fMRI activation (Wojciulik, Kanwisher, & Driver, 1998), the expertise effect could reflect greater attentional engagement by experts than novices. Even a task requirement to attend to identity (Gauthier, Skudlarski, Gore, & Anderson, 2000), or the use of an event related rather than blocked design (Xu & Kanwisher, 2001), may not completely override the greater interest in stimuli from the expert domain. The fact that ERP studies apparently show effects of expertise lasting hundreds of milliseconds after the N170 (e.g., figure 2 in Tanaka & Curran, 2001; figure 3a in Gauthier et al., 2003) is consistent with an attentional account.

Conclusions

We have examined in detail findings cited as support for the claim that facelike special processing emerges with experience for objects-of-expertise. We have shown that the empirical evidence *for* the expertise hypothesis is currently not strong, being essentially limited to a large inversion effect on memory for dog experts (which could be attributed to memory rather than object recognition processes) and a larger-than-other-objects (but much smaller than faces) FFA activation effect in car and bird experts (which could be attributed to attentional differences). We have also reviewed evidence *against* the expertise hypothesis, namely (1) the lack of any configural processing in the behavioral greeble studies and, more strongly, (2) the lack of any effect of real-world expertise in the part-whole paradigm, as well as (3) the lack of any effect of expertise on the face-selective M170, and (4) the lack of a correlation across subjects between rFFA activation during identification of objects of expertise and degree of expertise (Gauthier et al., 2000) as well as the lack of a correlation across trials between rFFA activation and successful car identification in car experts (Grill-Spector et al., 2004). Although all of the evidence discussed so far tests the hypothesized identity between face processing and expertise by considering the case of nonface expertise, the syndrome of congenital prosopagnosia enables us to consider the converse case of nonexpert face processing. Although the existence of apparently normal FFAs in these subjects (Hasson et al., 2003; see also Rossion et al., 2003), who have apparently never been experts at face processing at any point in their lives, indicates that the FFA is not sufficient for face recognition, it also shows that expertise is not necessary for FFA activation.[4] (In contrast, patient evidence suggests that the FFA is necessary for face recognition; Wada & Yamamoto 2001 and Barton et al., 2002.) Finally, perhaps the strongest evidence against the expertise hypothesis comes from the double dissociation in the neuropsychology literature in which some patients are impaired at face recognition but not at nonface identification of objects of expertise (Sergent & Signoret, 1992), while others show the opposite pattern (Moscovitch, Winocur, & Behrmann, 1997).

While we suspect that the alternative face specificity hypothesis will turn out to be correct one, we note that this conclusion is at present somewhat open. Quite a few studies of expertise have been published, but many have used laboratory-trained greeble experts, and relatively few have tested real-world experts with many years of experience. Moreover, key behavioral markers for "face-specific" processing are yet to be been tested in real-world experts. This includes the composite effect, and relational versus local alteration paradigms (where relational alterations to faces are sometimes more strongly affected by inversion than local part alterations in many studies; e.g., Leder & Bruce, 1998). Should these paradigms turn out to show evidence of configural processing for objects-of-expertise, the face specificity hypothesis would need to be reconsidered. Similarly, any clear evidence that the FFA itself (as opposed to nearby cortical regions) is necessary for expert identification of nonfaces would seriously challenge the face specificity hypothesis.

The research reviewed above suffers from two important shortcomings. First, even real-world experts are never as expert in any other stimulus domain as they are with

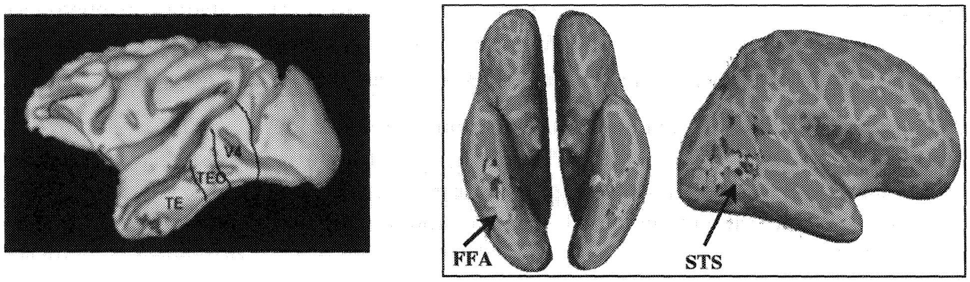

Figure 17.3
Regions responding more strongly in fMRI during viewing of faces than objects in a macaque (*left*, adapted from Tsao et al., 2003) and in an inflated human brain (*right*). It is unclear whether the macaque face area is homologous to the fusiform face area or the face-selective region in the STS region. See plate 20 for color version.

faces, so expertise effects that are much weaker than the corresponding effects for faces can generally be excused as reflecting insufficient expertise with the nonface domain. Second, because disruption methods are very limited in humans, it is difficult to test the *necessity* of face-selective cortical regions for processing objects of expertise. A possible new avenue for addressing both of these problems comes from the recent report (Tsao et al., 2003) of face areas in macaques (figure 17.3, plate 20), candidate homologs to the human FFA. To further investigate the homology across species, it will be important to test whether monkeys show the same behavioral signatures of configural processing of faces that have been so extensively demonstrated in humans. If it turns out that monkeys and humans process faces in similar ways, then further research in monkeys will make possible strong tests of the expertise hypothesis that avoid the shortcomings of the human literature. First, disruption methods not available in human research could be used to test the necessity of face-selective cortex for processing objects of expertise in monkeys. Second, by controlling perceptual experience from birth, it should be possible for the first time to conduct strong behavioral and neural tests of the expertise hypothesis in animals whose perceptual experience with faces and with other stimuli is matched.

Acknowledgments

Preparation of this article supported by Australian Research Council–Discovery grant DP0208630 to EM and NEI grant EY13455 to NK. We thank Marlene Behrmann, Brad Duchaine, Jia Liu, Morris Moscovitch, Rachel Robbins, and Galit Yovel for comments.

Notes

1. Note that, in some cases in this section, values have been estimated from figures.

2. The misorientation effect did disappear more slowly with practice in experts than in novices, but this is not a test for a "face-specific" effect: in normal nonexpert object recognition, misorientation effects for objects with a highly overlearned upright bias (letters, chairs, penguins, etc.) take 3–30 times through the stimulus set to disappear (Jolicoeur, 1985; McKone & Grenfell, 1999).

3. Gauthier, Curran, Curby, and Collins (2003) claimed that holistic processing was greater in car experts than in controls. They used a method bearing some similarity to the composite paradigm, but did not make the usual comparison between aligned and misaligned stimuli (which are matched for response competition from the other half; see figure 17.2). Instead, they defined "holistic processing" as the performance decrement when the to-be-ignored half suggested a response inconsistent, as opposed to consistent, with that required to the target half. This definition as merely the inability to ignore a notionally irrelevant component of the stimulus display is not what is usually meant by "holistic processing": indeed, under this definition, the Stroop effect (i.e., the difficulty of ignoring the word "red" when trying to name the color of the ink it is printed in) would incorrectly be interpreted as showing that color and word identity are processed together "holistically."

4. We thank Rebecca Saxe for pointing out the relevance of congenital prosopagnosia to the expertise hypothesis.

References

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*, 281–316.

Barton, J. J., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology, 58*, 71–78.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of human face perceptions in humans. *Journal of Cognitive Neuroscience, 8*, 551–565.

Bentin, S., & Carmel, D. (2002). Accounts for the N170 face-effect: A reply to Rossion, Curran and Gauthier. *Cognition, 85*, 197–202.

Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: Factors influencing distinct processing of faces. *Cognition, 83*, 1–29.

de Haan, M., Humphreys, K., & Johnson, M. H. (2002). Developing a brain specialized for face perception: A converging methods approach. *Developmental Psychobiology, 40*, 200–212.

de Renzi, E. (1986). Current issues in prosopagnosia. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. W. Young (Eds.), *Aspects of Face Processing* (pp. 243–252). Martinus Nijhiff: Dordrecht.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. J. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neurosciences*, *4*, 2051–2062.

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

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473.

Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology*, *17*, 117–123.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience*, *6*, 428–432.

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). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*, 1673–1682.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 431–446.

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

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, *38*, 2401–2428.

Grill-Spector, K. (2003). The functional organization of the ventral visual pathway and its relationship to object recognition. In N. Kanwisher & J. Duncan, (Eds.), *Attention & Performance XX: Functional Neuroimaging of Visual Cognition*. London: Oxford University Press.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The Fusiform face area subserves face detection and face identification, not generic within-category identification. *Nature Neuroscience*, *7*, 555–562.

Hasson, U., Avidan, G., Deouell, L. Y., Bentin, S., & Malach, R. (2003). Face-selective activation in a congenital prosopagnosic subject. *Journal of Cognitive Neuroscience*, *15*, 419–431.

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*, 189–99.

Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. *Visual Cognition*, *3*, 1–38.

Jolicoeur, P. (1985). The time to name disoriented objects. *Memory and Cognition*, *13*, 289–303.

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., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68*, B1–B11.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature, 412*, 786.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology, 51A*, 449–473.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5*, 910–916.

Liu, J., Higuchi, M., Marantz, A., & Kanwisher, N. (2000). Magnetophysiological studies of face and non-face processing. *NeuroReport*, 11, 337–341.

Maurer, D., Le Grand, R., & Mondloch, J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255–260.

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

McKone, E., & Grenfell, T. (1999). Orientation invariance in naming rotated objects: Individual differences and repetition priming. *Perception and Psychophysics, 61*, 1590–1603.

McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: A face-specific disorder. *Quarterly Journal of Experimental Psychology, 46A*, 1–10.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLEARN: A two process theory of infant face recognition. *Psychological Review, 98*, 164–181.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9*, 555–604.

Mundel, T., Milton, J. G., Dimitrov, A., Wilson, H. W., Pelizzari, C., Uftring, S., et al. (2003). Transient inability to distinguish between faces: electrophysiologic studies. *Journal of Clinical Neurophysiology, 20*, 102–110.

Nederhouser, M., Mangini, M., Biederman, I., & Okada, K. (2002). Matching face-like objects is invariant to differences in direction of contrast. Paper presented at the 10th Annual Workshop on Object Perception and Memory, Kansas City, MO.

Puce, A., Allison, T., Asgari, M., Gore, J., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letter strings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience, 16*, 5205–5215.

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.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*, 275–280.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17*, 43–63.

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

Rhodes, G., Byatt, G., Michie, P. T., & Puce, A. (2004). Is the fusiform face area specialized for faces, individuation, or expert individuation? *Journal of Cognitive Neuroscience, 16*, 189–203.

Rossion, B., Delvenne, J. F., Debatisse, D., Goffaux, V., Bruyer, R., Grommelinck, M., et al. (1999). Spatio-temporal localization of the face inversion effect: an event related potentials study. *Biological Psychology, 50*, 173–189.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science, 13*, 250–257.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., et al. (2000). The N170 occipito-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, 11*, 69–74.

Scapinello, K. F., & Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science, 21*, 329–330.

Schneider, W. X., Korkel, J., & Weinert, F. (1989). Domain specific knowledge and memory performance: A comparison of high and low aptitude children. *Journal of Educational Psychology, 81*, 301–312.

Sergent, J., & Signoret, J. L. (1992). Varieties of functional deficits in prosopagnosia. *Cerebral Cortex, 2*, 375–388.

Subramanian, S., & Biederman, I. (1997). Does contrast reversal affect object identification? *Investigative Ophthalmological and Visual Sciences, 38*, 998.

Tanaka, J., & Gauthier, I. (1997). Expertise in object and face recognition. *The Psychology of Learning and Motivation, 36*, 83–125.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science, 12*, 43–47.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46A*, 225–245.

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, L. (1996). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition, 5*, 479–496.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition, 25*, 583–592.

Tarr, M. J. (2003). Visual object recognition: Can a single mechanism suffice? In M. Peterson & G. Rhodes (Eds.), *Perception of Faces, Objects and Scenes: Analytic and Holistic Processes* (pp. 172–211). Oxford University Press: London.

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences, 7*, 23–30.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology, 17*, 257–279.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, *6*, 989–995.

Wada, Y., & Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. *Journal of Neurological and Neurosurgical Psychiatry, 71*, 254–257.

Wheeler, D. D. (1970). Processes of word recognition. *Cognitive Psychology, 1*, 59–85.

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: A fMRI study. *Journal of Neurophysiology, 79*, 1574–1478.

Xu, Y., & Kanwisher, N. (2001). What is the magnitude and theoretical significance of the FFA response to expert stimuli? Paper presented at the Society for Neuroscience, San Diego, CA.

Xu, Y., Liu, J., & Kanwisher, N. (2004). The M170 is selective for faces, not for expertise. *Neuropsychologia*, in press.

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

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*, 747–759.

18 Representation of Object Images by Combinations of Visual Features in the Macaque Inferotemporal Cortex

Manabu Tanifuji, Kazushige Tsunoda, and Yukako Yamane

The primate ventral visual pathway plays a role in extracting features from object images that are necessary for object recognition. Inferior temporal (IT) cortex in monkeys is the final stage of the ventral visual pathway, and visual responses of neurons in this area have been extensively investigated to find these features extracted from object images. Area TE in IT cortex has been of particular interest, since this area is the final purely visual area in the ventral visual pathway (Desimone et al., 1984; Tanaka, Saito, Fukado, & Moriya, 1991; Kobatake & Tanaka, 1994). As described in detail below, these investigations suggest that an object image is not represented by a single feature but by a set of features. Thus, it is difficult to explore the sets of features necessary for object recognition by conventional single cellular recordings because these techniques only provide responses of a small number of neurons at the same time. The combination study of a optical imaging technique and single cellular recordings overcomes this difficulty and reveals how combination of features could specifically represent object images in area TE.

Structural Description and View-Based Object Representation

Object representation in our brain is also an issue in computational and psychological studies of object vision. One proposal in these fields is that objects are represented by parts and spatial relationships among parts. Multiple groups have investigated possible models based on this proposal, referred to as "structural description" (Biederman, 1987; Marr & Nishihara, 1978). According to this proposal, an object image is first decomposed into parts and then the combination of these parts and their spatial relationships is used for recognition. Thus, object recognition could be view independent unless object parts are occluded. It should be emphasized that representation of the spatial relationships as well as that of parts themselves is essential for recognition in these studies. On the other hand, another group of studies has proposed that recognition is based on object views (Poggio & Edelman, 1990; Tarr & Bulthoff, 1998; Ullman, 1998; Riesenhuber & Poggio, 1999). In latter models, a view of an object could