

# Limits of generalization between categories and implications for theories of category specificity

CINDY M. BUKACH AND W. STEWART PHILLIPS  
*University of Richmond, Richmond, Virginia*

AND

ISABEL GAUTHIER  
*Vanderbilt University, Nashville, Tennessee*

Both domain-specific and expertise accounts of category specialization assume that generalization occurs within a domain but not between domains. Yet it is often difficult to define the boundaries and critical features of object domains. Differences in how categories are defined make it difficult to adjudicate between accounts of category specificity and may lead to contradictory results. For example, evidence for whether car experts recruit the fusiform face area is mixed, and this inconsistency may be due to the inclusion of antique cars in one of those previous studies (e.g., Grill-Spector, Knouf, & Kanwisher, 2004). The present study tested the generalization of expertise from modern to antique cars and found that modern-car experts showed expert discrimination and holistic processing of modern cars but not of antique cars. These findings suggest that the neural specialization underlying perceptual expertise is highly specific and may not generalize to distinct subclasses, even when they share some degree of perceptual and conceptual features.

By definition, expertise is domain specific. We would not expect a sommelier (wine expert) to know much about perfumery. We would not be surprised if a sommelier who can detect a taste of vanilla in a Chardonnay could not detect the same fragrance in a complex perfume like Shalimar, which, according to Guerlain, includes notes of lemon, bergamot, jasmine, may rose, iris, incense, opopanax, tonka bean, and vanilla. In this case, it may be intuitive that wines and perfumes form two separate domains, in great part as a result of functional differences. A sommelier may fail to engage wine-related expert strategies in the context of perfume identification. Yet perceptual expertise is also defined in part by generalization of expert skills to novel exemplars within a domain. For example, we would expect a sommelier trained with Australian wines to be able to apply his or her skills to a new Australian wine never before encountered. Indeed, empirical studies of expertise have demonstrated that an expert in a given domain can learn new items in his or her domain faster than can novices (Gauthier, Williams, Tarr, & Tanaka, 1998) and can outperform novices even on tasks that were not part of the training experience (Gauthier & Tarr, 2002).

Because expertise is characterized by both specificity between domains and generalization within a domain, understanding limits of generalization is particularly important for developing an account of how the brain becomes specialized for recognizing particular classes of objects. Specialization within the visual recognition system has

been demonstrated for many homogeneous categories, including faces, cars, birds, and novel objects (Diamond & Carey, 1986; Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997; Gauthier, Behrmann, & Tarr, 1999; Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier & Tarr, 1997; Wong, Palmeri, Rogers, Gore, & Gauthier, 2009). Differences between theoretical accounts for the development of such category specificity have centered on the relationship between recognition of faces and objects and the role of experience in the development of specialization within the visual recognition system. On the one hand, a domain-specificity account focuses on the idea that face recognition is carried out by a modular system that is dedicated to faces and is separate from a more generalized object-recognition system (Kanwisher, 2000). On the other hand, a perceptual-expertise account suggests that specialization for visual recognition of any homogeneous class of objects, including faces, is the result of an interaction between experience, task demands, and neural biases (Bukach, Gauthier, & Tarr, 2006; Gauthier & Tarr, 2002). Understanding domain boundaries and the limits of generalization between object categories is critical to both approaches. Differences in how categories are defined make it difficult to weigh the evidence in favor of one or the other theory of category specificity and may lead to contradictory results. For example, fMRI evidence for whether modern-car experts recruit the fusiform face area (FFA) when they look at cars is mixed,

---

C. M. Bukach, [cbukach@richmond.edu](mailto:cbukach@richmond.edu)

---

and the inconsistency may be due to a difference in how category boundaries were defined: Failures to replicate car-expertise effects included both antique and modern cars (Grill-Spector, Knouf, & Kanwisher, 2004), whereas studies that obtained an expertise effect used only modern cars (Gauthier, Curby, Skudlarski, & Epstein, 2005; Gauthier et al., 2000; Xu, 2005).

Unfortunately, domain boundaries and the circumstances under which expertise should generalize are not always easy to predict, particularly when there is substantial overlap between perceptual and semantic features of two domains or two subclasses within a domain, as there is, for instance, for two species of birds or for two classes of cars. In this study, we considered generalization of expertise between two categories of objects that share semantic and visual features (modern cars and antique cars) and the implications of limits of generalization to theories of category specificity.

### **Importance of Generalization to Theories of Perceptual Expertise**

Generalization of skilled performance has played a central role in studies of perceptual expertise and in fact is a salient difference between perceptual expertise and perceptual learning (Tanaka, Curran, & Sheinberg, 2005). In laboratory studies of perceptual expertise, subjects are first trained to criterion with one set of objects and then assessed for transfer of expertise using novel exemplars from the same class. Not only can experts learn to identify new exemplars more rapidly than do novices, they also demonstrate other behavioral and neurological markers of expertise that are typically associated with face recognition, including inversion effects, holistic and relational processing, recruitment of the FFA, and increased amplitude of the N170 (Gauthier & Tarr, 2002; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Gauthier et al., 1998; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002; Wong, Palmeri, & Gauthier, 2009; Wong, Palmeri, Rogers, et al., 2009; Xu, 2005). This methodology for testing expertise via generalization is based on theories of object representation that account for generalization to novel exemplars as a function of the similarity between specific experienced views (Edelman & Bülthoff, 1992; Tarr & Pinker, 1990).

Surprisingly few expertise studies have examined category boundaries or the limits of generalization in a systematic way. Some work on expertise training suggests that categories are rather narrowly defined and that transfer is best for novel exemplars whose diagnostic features are most similar to the trained set. For example, in one study of expertise training using novel objects called *Greebles* (Gauthier et al., 1998), experts showed a distinct advantage over novices for learning to identify new exemplars from the trained families immediately following training. When retested 8–13 weeks later, however, experts were no better than novices for *Greebles* whose diagnostic features differed from the trained set, and expertise did not generalize to *Greebles* from new families that were more homogeneous than the trained families. Although set distinctiveness was confounded with time

since training, these results converge with those from a study that examined generalization of learning following training of different bird species (Tanaka et al., 2005). Subjects who learned to individuate species of wading birds or owls demonstrated transfer of learning to new exemplars from within the trained species, a lesser degree of transfer to new exemplars from new species within the same family, and no transfer between bird families (e.g., subjects trained to recognize great blue herons showed a robust advantage for new instances of great blue herons, a small advantage for instances of black-crowned night herons, and no advantage for owls). This pattern of generalization was explained as a result of selective tuning of perceptions of color, shape, and texture cues that were common across the relevant species. Interestingly, this study also found that transfer between species within a family occurred only for wading birds and not for owls. There is no theoretical explanation for this difference between bird species, and it highlights the difficulty of making predictions regarding generalization between object categories and the need to test category boundaries empirically.

### **Importance of Generalization to Theories of Domain Specificity**

Generalization is also important for domain-specific theories of object recognition. This theoretical approach conceives of the mind as a set of modules that are specified in terms of the content of the information processed rather than the underlying processes themselves (Fodor, 1983). One assumption is that domains and their boundaries can be clearly specified in terms of their essential properties, with the expectation that such content-specified modules would not generalize to exemplars outside of the domain for which the module was specialized.

Domain specificity for faces is perhaps the most well-cited example of a content-specified module. In fact, the difference between content and process has been used to contrast the domain-specific account with the perceptual-expertise account of face recognition (Kanwisher, 2000). According to the domain-specific theory of face recognition, activity in neural regions specialized for faces is explained in terms of the overlap in content between faces and the category of interest. An a priori theory of what a face is would help to predict the degree to which different classes of stimuli will activate face-specific regions.

Yet even faces have been difficult to define as a simple cohesive category. For instance, some objects with a configuration of features similar to faces are not treated as such (Gauthier, Behrmann, & Tarr, 1999), whereas mere blobs can be processed like faces in the right context (Cox, Meyers, & Sinha, 2004). The human face recognition system appears to be specialized for species-specific information, since adults show poor ability to recognize faces from other species (Dufour, Pascalis, & Petit, 2006; Pascalis, de Haan, & Nelson, 2002). Importantly, the critical morphological features of own-species faces have yet to be specified, and, moreover, research has shown that these boundaries are flexible: The ability to recognize other-species faces can improve with early experience (Pascalis

et al., 2005). From a domain-specific perspective, the degree to which experience can expand domain boundaries should depend on how well the exemplars share the critical features that define the module's domain. Yet it is unclear how far experience can expand the boundaries of own-species recognition and whether face-specific modules can process even nonface stimuli.

Given the difficulty inherent in defining category boundaries, researchers must be cautious when making predictions regarding generalization, particularly when contrasting theoretical explanations that depend on whether or not effects generalize between categories. As a case in point, consider one study that failed to replicate expertise effects for car experts. Grill-Spector et al. (2004) found no expertise-related activation of the FFA for car experts, contrary to findings from previous expertise studies (Gauthier et al., 2005; Gauthier et al., 2000; Xu, 2005). Neither set of studies explicitly defined the category of cars; in fact, most researchers would assume that people have a general concept of a car that is fairly universal. However, a critical difference between the two studies was the range of car models that were included in the scanner. Whereas Gauthier and colleagues (Gauthier et al., 2005; Gauthier et al., 2000) used sedans from the 1990s to assess both expertise and FFA activation, Grill-Spector et al. used sedans from the 1990s to assess level of expertise but used two different subclasses of cars—jeeps or primarily antique cars—during the imaging task.<sup>1</sup> It is possible that antique cars represent a subclass of cars that are distinct from modern cars, and that expertise does not generalize from modern to antique cars, thereby limiting the possibility of finding FFA activation.

The purpose of the present study was to test whether expertise generalizes between antique and modern cars, in an effort to reconcile these contradictory results and to consider the implications of limits of generalization between subclasses to theories of category specificity more generally. Although modern and antique cars share a similar configuration of parts (four wheels, two or four doors, a front and rear windshield, etc.), there have been significant changes to the shape of car bodies and to car parts over time; thus, antique cars may represent a structurally distinct subclass relative to modern cars. Car expertise is a domain that is frequently used in research on extant expertise, because experts are easily found and because good quality images of cars are available (Carmel & Bentin, 2002; Curby & Gauthier, 2009; Curby, Glazek, & Gauthier, 2009; Gauthier & Curby, 2005; Gauthier, Curran, Curby, & Collins, 2003; Gauthier et al., 2000; Hershler & Hochstein, 2009; McGugin & Gauthier, 2010; Rossion, Collins, Goffaux, & Curran, 2007; Scott, Tanaka, Sheinberg, & Curran, 2008; Williams, Willenbockel, & Gauthier, 2009; Xu, 2005; Xu, Liu, & Kanwisher, 2005). It is therefore important to characterize the specificity of expert skills in this domain.

In Experiment 1, we tested modern-car experts and novices for discrimination of modern and antique cars using the standard expertise-assessment task. In Experiment 2, we tested modern-car experts and novices for holistic processing of modern and antique cars. We chose

to examine this particular behavioral marker of expertise because holistic processing is correlated with activity in the FFA (Gauthier & Tarr, 2002). On the basis of the structural differences between modern and antique cars, as well as the failure to find expertise-related activity in the FFA (Grill-Spector et al., 2004), we hypothesized that modern-car experts would show expert discrimination and holistic processing of modern cars but not of antique cars.

## EXPERIMENT 1 Expert Discrimination

The discrimination task assesses how well individuals can make subordinate-level matching judgments of exemplars from various classes of objects, and it has been used extensively to classify experts in studies of real-world expertise (Gauthier et al., 2000; Grill-Spector et al., 2004). Expertise for a particular class of objects is defined here as a sensitivity score greater than  $2 d'$ . Recently, McGugin and Gauthier (2010) found that this task is more sensitive to expertise if an additional criterion of  $d'$  less than 1 on a nonexpert object class is used as an inclusion criterion for the study, to ensure that expert performance is not due to general discrimination ability. We used this additional criterion here for both experts and novices, with birds as the nonexpert category.

The purpose of the discrimination task was twofold. First, we used performance to classify participants as modern-car experts or as novices, as described above. Second, we compared discrimination for modern and antique cars to determine whether expert discrimination would generalize from modern to antique cars. On the basis of Grill-Spector et al. (2004), we hypothesized that modern-car experts would not reach criterion for antique cars.

### Method

**Participants.** Modern-car experts were recruited through advertisements targeting individuals with above-average ability to recognize modern cars. Experts self-rated their ability to recognize modern cars on a scale from 1 to 10 (1 = *little ability*, 5 = *average ability*, 10 = *superior ability*). A similar scale was used for self-ratings on the ability to recognize antique cars. Only participants who met the criteria for modern-car expertise described above were included in the modern expert group, resulting in a total of 9 modern-car experts (mean self-rating was 8.17 for modern cars and 4.94 for antique cars; mean age = 22.2 years). An age-matched control group was recruited through advertisements targeting individuals with average ability to recognize modern cars. Twelve age-matched controls met criterion for inclusion in the study (mean self-rating was 3.92 for modern cars and 1.42 for antique cars; mean age = 20.42 years). Because young novices have exposure to modern cars but may have little or no prior exposure to antique cars, it is difficult to establish baseline performance for the two categories. To overcome this difficulty, we also recruited a group of older novices between the ages of 50 and 70, who we assumed had been exposed to 1950s cars when they were younger and when these cars were commonly on the road. Twelve older novices met criterion for inclusion in the study (mean self-rating was 5.25 for modern cars and 3.33 for antique cars; mean age = 55.92 years).

**Materials.** On hundred sixteen grayscale images from each of the following categories were used: modern cars (1997–2003), antique cars (1950s), and birds (passerines). A mask consisting of a scrambled image was also used for each category. All images were presented on a gray background and were sized to fit within a cen-

trally positioned  $3.5 \times 3.5$  in. area on the computer screen. See Figure 1 for sample stimuli.

**Procedure.** Subjects matched sequentially presented images of cars and birds on the basis of their model or species. A matching trial consisted of car exemplars from the same model but different years, or from the same bird species but from different views. Each trial began with a central fixation point for 500 msec, followed by an image lasting 1,000 msec, and a mask for 500 msec, concluding with the second image until a response or a maximum of 5,000 msec. Subjects responded with a keypress. Categories were shown in different blocks, and order of blocks was randomized across participants. Each block contained 112 trials in random order, with an additional 4 practice trials.

## Results and Discussion

A one-way ANOVA on sensitivity for the control category of birds showed that all groups performed equally well [ $F(2,30) = 0.84, p = .44$ ]. Performance for birds will therefore not be discussed further. Mean sensitivity for modern and antique cars is plotted in Figure 2. As the 95% confidence intervals around the point estimates show, modern-car experts did not reach criterion for antique-car expertise. In fact, an examination of individual data shows that no modern-car expert reached expertise criterion for antique cars. Data were submitted to a  $2 \times 3$  mixed factorial ANOVA for further analysis, with category (modern cars, antique cars) as the within-subject factor and group (modern-car experts, age-matched novices, older novices) as the between-subjects factor. There was a main effect of group [ $F(2,30) = 14.58, p < .001, \eta^2 = .49$ ]. There was also a main effect of category [ $F(1,30) = 148.31, p <$

$.001, \eta^2 = .83$ ]. Importantly, there was a significant interaction between group and category [ $F(2,30) = 58.01, p < .001, \eta^2 = .80$ ]. Tukey–Kramer post hoc tests with  $\alpha = .05$  showed that both modern-car experts and age-matched controls discriminated modern cars better than they did antique cars. Importantly, however, older novices who had previous exposure to both types of cars showed equal ability to discriminate the two car classes, indicating that antique-car comparisons were not merely more difficult than those for modern cars. Although modern-car experts showed a strong advantage for modern cars relative to the two control groups, modern-car experts did not differ from novices in either control group in their ability to discriminate antique cars. These results provide convincing evidence that expert discrimination of modern cars does not generalize to discrimination of antique cars.

## EXPERIMENT 2 Holistic Processing

Holistic processing is the tendency to process multiple features as a whole, and this effect can be measured by the inability to selectively attend to a single part (Maurer, Le Grand, & Mondloch, 2002). Faces are processed more holistically than are other objects, and holistic processing also is a marker of expertise for nonface objects (Gauthier et al., 2003; Gauthier & Tarr, 2002). In the present study, holistic processing was assessed in the composite task (Young, Hellawell, & Hay, 1987). Composite stimuli

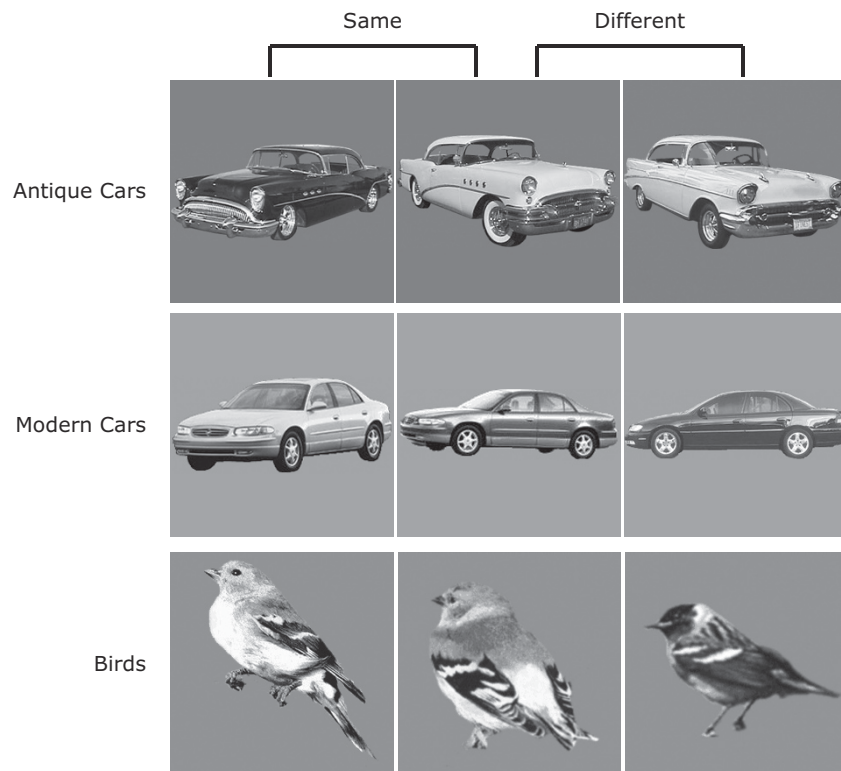
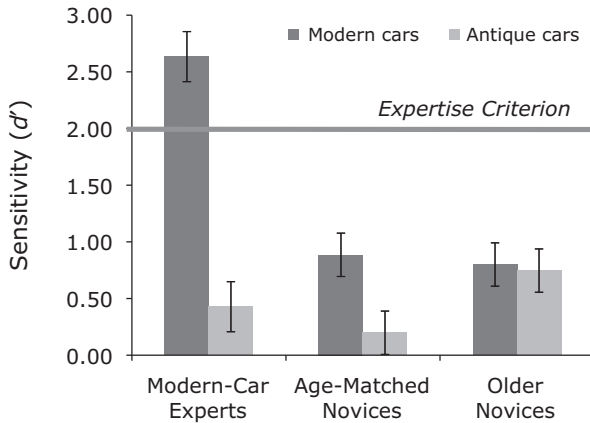


Figure 1. Sample stimuli used in the expertise task.





**Figure 2.** Mean discrimination sensitivity of modern and antique cars in the expertise task. Error bars are 95% confidence intervals based on the  $MS_e$  from the omnibus ANOVA. The line shows the cutoff for the expertise criterion.

are made from the top and bottom of two different exemplars, and the task requires participants to decide whether the cued parts (either top or bottom) of two sequentially presented composites match. On congruent trials, both cued and noncued parts change, or both remain the same. On incongruent trials, one part changes while the other part remains the same. Holistic processing is measured by the degree to which accuracy is reduced and/or response times are slower for incongruent relative to congruent trials (congruency effect). The congruency effect is reduced when parts are misaligned at test (alignment effect), presumably because misalignment disrupts holistic processing. To control for response bias, we used the full version of the composite task, in which both same and different trials are analyzed and contribute to the congruency measure (see Cheung, Richler, Palmeri, & Gauthier, 2008; Gauthier & Bukach, 2007; Richler, Cheung, Wong, & Gauthier, 2009, for a discussion and empirical evidence supporting the importance of including both same and different trials in the composite task). Holistic processing is associated with activity in the right FFA (Gauthier & Tarr, 2002); therefore, on the basis of the failure of modern-car experts to recruit the FFA when car stimuli included antique cars (Grill-Spector et al., 2004) and on the basis of the results of Experiment 1 in the present study, we expected modern-car experts to show greater holistic processing of modern cars than of antique cars. That is, we predicted that experts, but not novices, would show a greater congruency effect on aligned trials for modern cars than for antique cars. In addition, we predicted that modern-car experts, but not novices, would show a greater alignment effect for modern cars than for antique cars.

## Method

**Participants.** The same 33 subjects participated in the composite task.

**Materials.** Composite stimuli were constructed from gray-scale images of 20 antique and 20 modern cars, sized to a width of 256 pixels. Forty aligned composites for each car class were created

by rearranging the top and bottom of two different cars. A solid black line separated the top and bottom. On misaligned trials, the noncued part was moved 60 pixels to the right or to the left. Two cues were created consisting of a black bracket oriented to encompass either the top or the bottom of the car. A  $256 \times 256$  pixel texture mask was also created. See Figure 3 for sample stimuli.

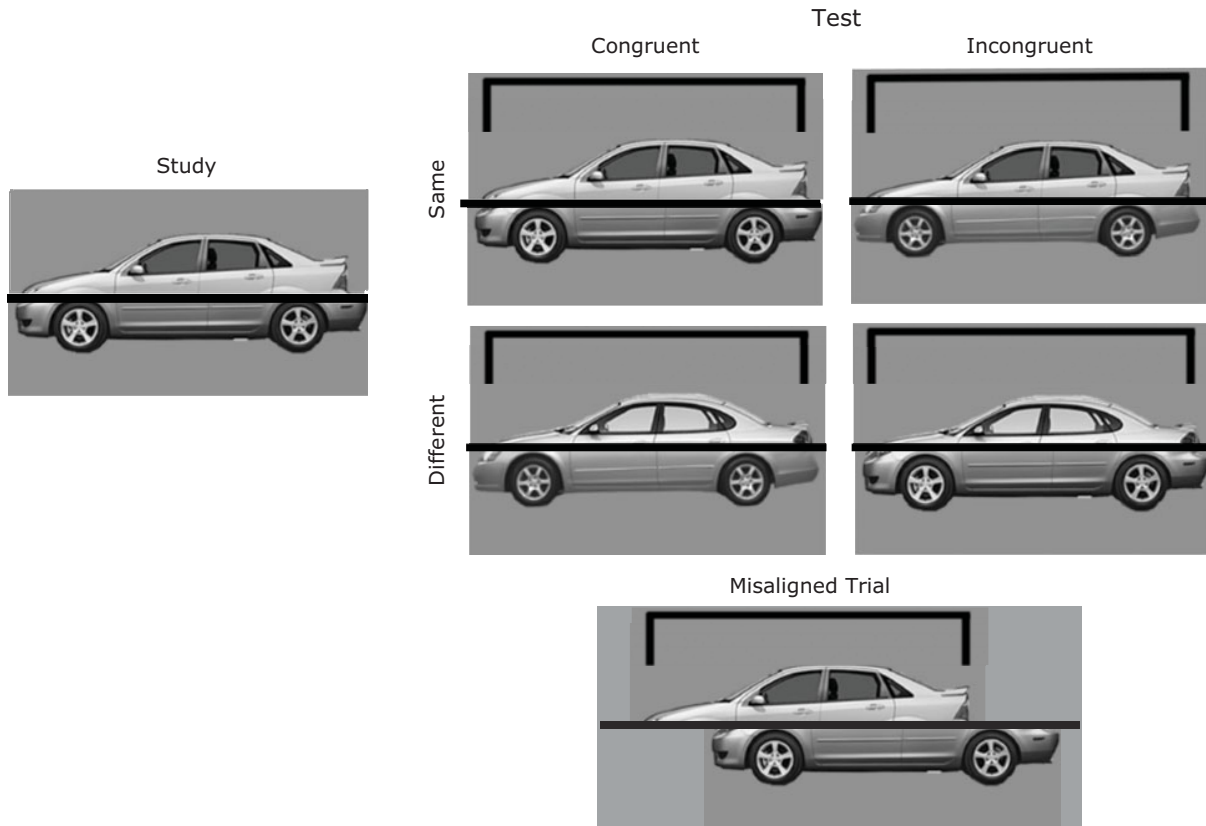
**Procedure.** Each trial began with an initial fixation point for 500 msec, followed by an aligned car composite. The image lasted for 1,500 msec, followed by a mask that flashed for 400 msec and then a cue for 300 msec at the top or bottom of the screen, indicating the part of the car that the subject was asked to match. While the cue remained on the screen, a second composite appeared. This composite was either aligned or misaligned. The subject responded with a keypress (1 = *same*, 2 = *different*). A response deadline of 4,000 msec was used. There were 640 trials, divided into eight blocks of 80 trials each. Trials were blocked by car type to avoid context effects (Richler, Bukach, & Gauthier, 2009), with cue (top or bottom), alignment (aligned or misaligned), congruency (congruent or incongruent), and correct response (same or different) randomized within each block. Order of blocks alternated by car type and was counterbalanced across subjects. Participants were given the opportunity to take breaks at the end of each block. The task started with 16 practice trials, 8 for each car type.

## Results and Discussion

Holistic processing was assessed by calculating the congruency effect for sensitivity (congruent  $d'$  – incongruent  $d'$ ) and response times (incongruent – congruent). Larger congruency effects for aligned trials indicate more holistic processing. Sensitivity and response times were submitted to a  $2 \times 2 \times 2$  mixed factorial ANOVA, with group (expert vs. novice) as a between-subjects factor and car type (modern vs. antique) and alignment (aligned vs. misaligned) as within-subject factors. We first present the results of the omnibus ANOVAs before presenting planned contrasts to test our specific hypotheses.

**Sensitivity ANOVA.** Mean congruency effects for sensitivity are plotted in Figure 4. There was a main effect of alignment [ $F(1,30) = 52.57, p < .001, \eta^2 = .64$ ], with a greater congruency effect for aligned cars ( $M = .86$ ) than for misaligned cars ( $M = .16$ ). There was also an interaction between car type and group [ $F(2,30) = 4.67, p = .017, \eta^2 = .24$ ]. Post hoc tests using simple effects revealed that modern-car experts (but not novices) showed a greater congruency effect for modern cars than for antique cars ( $p < .05$ ). No other effects reached significance.

**Response-time ANOVA.** Outliers smaller than 300 msec and greater than 4,000 msec were removed from the data (0.46% of data removed). Mean congruency effects for response times are plotted in Figure 5. There was a main effect of alignment [ $F(1,30) = 30.04, p < .001, \eta^2 = .50$ ], with a greater congruency effect for aligned cars ( $M = 57.72$ ) than for misaligned cars ( $M = -16.51$ ). This main effect was qualified by a three-way interaction between group, car and alignment [ $F(2,30) = 3.80, p = .034, \eta^2 = .20$ ]. Post hoc tests using simple effects revealed that only experts showed a car  $\times$  alignment interaction ( $p = .006$ ). This interaction was further investigated using simple main effects. These analyses revealed that modern-car experts showed a greater congruency effect for aligned modern cars than for misaligned modern cars ( $M = 118$  and  $-49$  msec, respectively;  $p =$



**Figure 3.** Example trial structure of the composite task, showing study stimulus and four possible test conditions for aligned stimuli. These four test conditions were also misaligned on half of the trials. One sample misaligned stimulus is shown. Car bottoms were also cued on half of the trials.

.002) but no effect of alignment for antique cars ( $M = -8$  and 5 msec for aligned vs. misaligned antique cars, respectively;  $p = .751$ ).

**Hypothesis 1: Holistic processing for modern versus antique cars.** Our first hypothesis was that modern-car experts, but not novices, would show greater holistic processing for modern cars than for antique cars. This hypothesis was investigated using planned comparisons on aligned trials for sensitivity and response-time congruency measures. Effects in sensitivity were consistent with our hypothesis: Experts showed a greater congruency effect on aligned trials for modern cars than for antique cars [ $t(30) = 4.26, p < .001$ ]. Age-matched novices did not show a difference between modern and antique cars ( $p > .5$ ), whereas older novices showed a greater congruency effect for antique cars than for modern cars [ $t(30) = 2.59, p < .05$ ]. Planned comparisons on congruency effects in response times showed a similar pattern: Modern-car experts showed a greater congruency effect for modern cars than for antique cars [ $t(30) = 4.12, p < .001$ ], but neither age-matched novices nor older novices showed a difference between modern and antique cars ( $ps > .05$ ).

**Hypothesis 2: Effect of alignment for modern versus antique cars.** Holistic processing is stronger when

parts are aligned than when they were misaligned. We therefore predicted that modern-car experts, but not novices, would show a greater alignment effect for modern cars than for antique cars. Planned comparisons on sensitivity measures did not reveal any differences in effects of alignment between modern and antique cars for any group. Consistent with our hypothesis and prior work (Wong, Palmeri, & Gauthier, 2009), however, planned comparisons on response times revealed that modern-car experts showed a greater effect of alignment for modern cars than for antique cars [ $t(30) = 2.93, p < .01$ ]. Alignment effects did not differ between modern and antique cars for either novice group ( $ps > .05$ ).

The pattern of results across both sensitivity and response-time measures supports the prediction that expertise-related holistic processing for modern cars does not generalize to antique cars. The lack of holistic processing for antique cars among modern-car experts cannot be due to the stimulus properties of antique cars alone, because there was a main effect of congruency across modern and antique cars for both novice groups, and in fact older novices who had previous exposure to antique cars showed a greater congruency effect for antique cars than for modern cars. We can therefore conclude that holistic

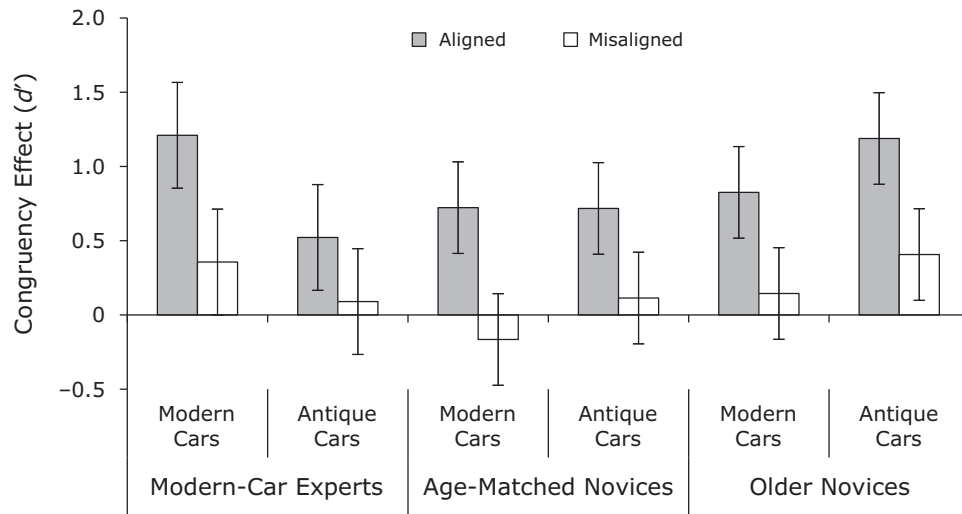


Figure 4. Mean congruency effects for sensitivity in the composite task. Error bars represent the 95% confidence intervals based on the  $MS_e$  for the three-way interaction from the omnibus ANOVA.

processing as a marker of expertise for cars is specific to its subclass and does not generalize from modern to antique cars.

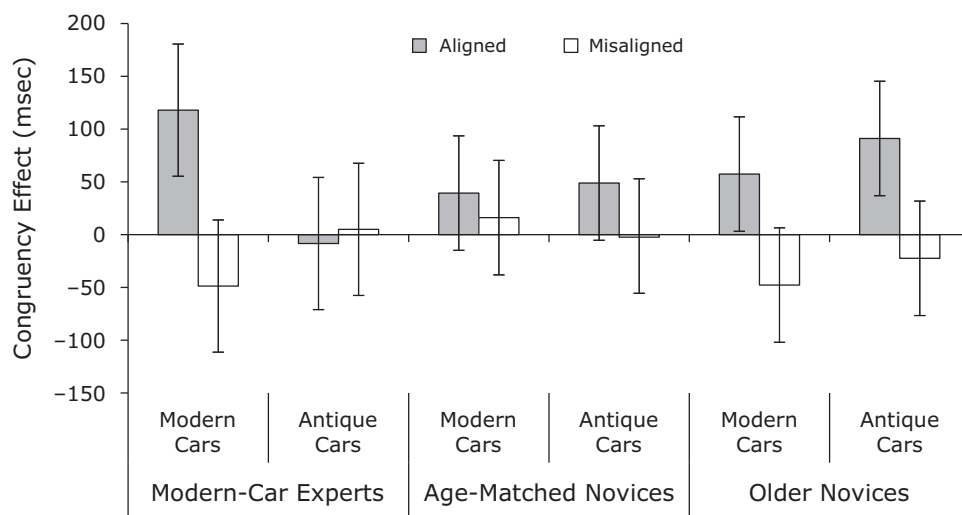
## GENERAL DISCUSSION

We examined the specificity of expert discrimination and holistic processing for modern and antique cars, two subclasses of stimuli that share perceptual and conceptual features. Across two experiments, we found evidence that expertise is highly specific and does not generalize from modern to antique cars. Modern-car experts who reached criterion for modern cars in the discrimination task did not reach expertise criterion for antique cars, and in fact they performed no differently from novices for antique cars. Furthermore, modern-car experts, but not novices, showed more holistic processing of modern cars than for antique cars, as evidenced by both congruency and alignment effects in the composite task. These results suggest that perceptual expertise produces a finely tuned perceptual system that is specific to the particular diagnostic features with which individuals have experience.

Our results are consistent with previous expertise studies that have examined limits of generalization between related subclasses of Greebles (Gauthier et al., 1998) and birds (Tanaka et al., 2005). In both of these studies, expertise failed to generalize to novel exemplars when there were subtle perceptual differences between subclasses. Our findings are also consistent with studies of specificity within the domain of faces that found little or no generalization of face-recognition ability between subclasses of faces. For example, a robust advantage for own-race faces has been demonstrated across a variety of paradigms. This *other-race effect* appears to impact almost all domain-specific markers of face processing: Recognition of other-race faces involves less relational and holistic processing, less activation of the FFA, a delayed N170, and a reduced

inversion effect in both the N170 and behavior (Gajewski, Schlegel, & Stoerig, 2008; Golby, Gabrieli, Chiao, & Eberhardt, 2001; Hancock & Rhodes, 2008; Ito & Urland, 2003, 2005; Michel, Rossion, Han, Chung, & Caldara, 2006; Rhodes, Brake, Taylor, & Tan, 1989; Rhodes, Hayward, & Winkler, 2006; Stahl, Wiese, & Schweinberger, 2008; Tanaka, Kiefer, & Bukach, 2004; Wiese, Stahl, & Schweinberger, 2009). Although the other-race effect can be influenced by nonperceptual factors such as in-group versus out-group status and mood (Bernstein, Young, & Hugenberg, 2007; Johnson & Fredrickson, 2005; Shriver, Young, Hugenberg, Bernstein, & Lanter, 2008), several studies have highlighted the importance of experience in mediating the other-race effect (Michel et al., 2006; Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen, 2005; Tanaka et al., 2004), indicating a perceptual basis for the effect that is experience dependent. These experience-dependent limitations of generalization between subclasses within both face and nonface domains have implications for theories of category specificity and of the organization and function of the object-recognition system more generally.

One implication of our study is the importance of setting appropriate category boundaries when contrasting theoretical accounts of category specialization. The lack of generalization between modern and antique cars offers a resolution to conflicting studies regarding expertise-related activation of the FFA in response to cars. Studies that found expertise-related activation tested modern-car experts with modern cars (Gauthier et al., 2000), whereas studies that failed to replicate this effect included antique cars in the imaging session (Grill-Spector et al., 2004). On the basis of the present study, we would predict that modern-car experts will show little expertise-related FFA activation to antique cars. Rather than providing evidence against an expertise account of activity in the FFA, this failure to recruit the FFA can be interpreted as providing



**Figure 5.** Mean congruency effects for response times in the composite task. Error bars represent the 95% confidence intervals based on the  $MS_e$  for the three-way interaction from the omnibus ANOVA.

further support for the specificity of modern-car expertise. It is possible that a similar explanation also applies to other failures to replicate expertise effects in imaging studies. For example, Rhodes, Byatt, Michie, and Puce (2004) did not find expertise-related FFA activation for Lepidoptera experts when using species of Lepidoptera that were unfamiliar to the experts. Likewise, Yue, Tjan, and Biederman (2006) failed to obtain expertise effects in participants who had been trained with a category of novel objects, but the objects used at test were not only unfamiliar but also arguably qualitatively different in their structure. Although our alternative explanations of these null results in the imaging domain remain speculative because they are based on behavioral data, the lack of generalization between two very similar subclasses in our study demonstrates that category boundaries are not always intuitively obvious, and it highlights the need for caution when choosing stimuli or interpreting null effects in expertise studies. Although researchers often want to demonstrate that effects transfer to novel exemplars, neural effects of expertise should perhaps not be expected when behavior suggests a lack of generalization.

At first glance, it may appear paradoxical that brain regions specialized for faces can also be specialized for such structurally divergent categories as cars or birds and yet exhibit such limitations in generalization between subclasses of faces, cars, and birds. Clearly, principles based on content or structural similarity alone cannot explain how a region such as the FFA could show both such specificity and such diversity. A process-based account that includes additional factors such as task demands and experience can more readily explain such a pattern of results. According to an expertise account, the FFA may process both faces and cars, not because these two categories are structurally similar to one another but because they are both homogeneous object classes that engage holistic processing after intensive experience involving individu-

ation. Although expertise generalizes to novel exemplars within a category (presumably on the basis of structural similarity), we demonstrate here that this generalization is very limited because of the fine-tuning of the perceptual system that results from this experience. This fine-tuning can be conceptualized as a “stretching” of the structural dimensions that were relevant during perceptual training, resulting in improved discrimination for novel exemplars only to the degree that they share these relevant dimensions. Thus, the FFA may be recruited for both modern cars and own-race faces as a result of the engagement of holistic processing, at the same time showing little activation for antique cars and other-race faces because of fine-tuning within each class.

Our study suggests more broadly that common principles may apply to limits of generalization within face and nonface domains. Thus, understanding the limits of generalization within one domain may help us to understand the conditions under which expertise will transfer within another domain, improving predictions of transfer and ultimately improving performance. For example, the study of bird training suggested that individuation experience is necessary for the development of perceptual expertise (Tanaka et al., 2005). This principle has been applied to the face domain to improve recognition between subclasses of faces. For example, Lebrecht, Pierce, Tarr, and Tanaka (2009) recently demonstrated that individuation training, and not mere exposure, is necessary to reduce the other-race effect. Interestingly, although training modern-car experts with antique cars should increase their performance and use of holistic processing for antique cars, it is unclear whether there would be any generalization across car subclasses. In other words, would a modern-car expert acquire antique-car expertise differently than would a bird expert? There are currently no data speaking directly to this issue.

One possible area for future investigation is to better understand how structural differences impact generaliza-



tion. For example, what are the critical structural features that differ between modern and antique cars? Some studies have begun asking similar questions with respect to other-race faces (Bar-Haim, Saidel, & Yovel, 2009; Blair, Judd, Sadler, & Jenkins, 2002; Hills & Lewis, 2006). Currently, however, there is no metric that would allow accurate predictions of expertise transfer between perceptually similar subclasses from either an expertise or a domain-specific account. Further investigation of the features that determine when expertise generalizes between subclasses of both face and nonface objects should provide a set of general principles on which such predictions could be based. For example, generalization could be limited to exemplars that vary on dimensions that are diagnostic for the expert category or to experienced ranges of values on these diagnostic dimensions. It is also possible that greater variation in trained exemplars could result in a more flexible system of representation and greater generalization. At present, these hypotheses are highly speculative and require systematic empirical investigation.

The results of the present study demonstrate that perceptual expertise is highly specific and that at least some failures to replicate expertise effects may be due to incorrect setting of category boundaries. In particular, we argue that behavioral evidence of expert performance is necessary to interpret the brain activity of putative experts and that the behavioral and neural tasks should use objects that are sampled from the same region of perceptual space.

#### AUTHOR NOTE

This research was supported by the James S. McDonnell Foundation and Grant 2 R01 EY013441-06A2 to I.G., by Center Grant P30-EY008126, and by a faculty research grant to C.M.B. and an undergraduate research award to W.S.P., both from the University of Richmond. The authors thank Ludvik Bukach and Kathryn Roberts for assistance in collecting data. Correspondence concerning this article should be sent to C. M. Bukach, Department of Psychology, 212 Richmond Hall, 28 Westhampton Way, University of Richmond, VA 23173 (e-mail: cbukach@richmond.edu).

#### REFERENCES

- BAR-HAIM, Y., SAIDEL, T., & YOVEL, G. (2009). The role of skin colour in face recognition. *Perception*, **38**, 145-148. doi:10.1068/p6307
- BERNSTEIN, M. J., YOUNG, S. G., & HUGENBERG, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, **18**, 706-712. doi:10.1111/j.1467-9280.2007.01964.x
- BLAIR, I. V., JUDD, C. M., SADLER, M. S., & JENKINS, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality & Social Psychology*, **83**, 5-25. doi:10.1037/0022-3514.83.1.5
- BUKACH, C. M., GAUTHIER, I., & TARR, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences*, **10**, 159-166. doi:10.1016/j.tics.2006.02.004
- CARMEL, D., & BENTIN, S. (2002). Domain specificity versus expertise: Factors influencing distinct processing of faces. *Cognition*, **83**, 1-29. doi:10.1016/S0010-0277(01)00162-7
- CHEUNG, O. S., RICHLER, J. J., PALMERI, T. J., & GAUTHIER, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception & Performance*, **34**, 1327-1336. doi:10.1037/a0011752
- COX, D., MEYERS, E., & SINHA, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science*, **304**, 115-117. doi:10.1126/science.1093110
- CURBY, K. M., & GAUTHIER, I. (2009). The temporal advantage for individuating objects of expertise: Perceptual expertise is an early riser. *Journal of Vision*, **9**(6, Art. 7), 1-13. doi:10.1167/9.6.7
- CURBY, K. M., GLAZEK, K., & GAUTHIER, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception & Performance*, **35**, 94-107. doi:10.1037/0096-1523.35.1.94
- DIAMOND, R., & CAREY, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, **115**, 107-117. doi:10.1037/0096-3445.115.2.107
- DUFOR, V., PASCALIS, O., & PETIT, O. (2006). Face processing limitation to own species in primates: A comparative study in brown capuchins, Tonkean macaques and humans. *Behavioural Processes*, **73**, 107-113. doi:10.1016/j.beproc.2006.04.006
- EDELMAN, S., & BÜLTHOFF, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, **32**, 2385-2400. doi:10.1016/0042-6989(92)90102-O
- FODOR, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- GAJEWSKI, P. D., SCHLEGEL, K., & STOERIG, P. (2008). Effects of human race and face inversion on the N170: A cross-race study. *Journal of Psychophysiology*, **22**, 157-165. doi:10.1027/0269-8803.22.4.157
- GAUTHIER, I., ANDERSON, A. W., TARR, M. J., SKUDLARSKI, P., & GORE, J. C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Current Biology*, **7**, 645-651. doi:10.1016/S0960-9822(06)00291-0
- GAUTHIER, I., BEHRMANN, M., & TARR, M. J. (1999). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience*, **11**, 349-370. doi:10.1162/089892999563472
- GAUTHIER, I., & BUKACH, C. (2007). Should we reject the expertise hypothesis? *Cognition*, **103**, 322-330. doi:10.1016/j.cognition.2006.05.003
- GAUTHIER, I., & CURBY, K. M. (2005). A perceptual traffic jam on highway N170: Interference between face and car expertise. *Current Directions in Psychological Science*, **14**, 30-33. doi:10.1111/j.0963-7214.2005.00329.x
- GAUTHIER, I., CURBY, K. M., SKUDLARSKI, P., & EPSTEIN, R. A. (2005). Individual differences in FFA activity suggest independent processing at different spatial scales. *Cognitive, Affective, & Behavioral Neuroscience*, **5**, 222-234. doi:10.3758/CABN.5.2.222
- GAUTHIER, I., CURRAN, T., CURBY, K. M., & COLLINS, D. (2003). Perceptual interference supports a nonmodular account of face processing. *Nature Neuroscience*, **6**, 428-432. doi:10.1038/nn1029
- 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. doi:10.1038/72140
- GAUTHIER, I., & TARR, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, **37**, 1673-1682. doi:10.1016/S0042-6989(96)00286-6
- GAUTHIER, I., & TARR, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 431-446. doi:10.1037/0096-1523.28.2.431
- 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. doi:10.1038/9224
- 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. doi:10.1016/S0042-6989(97)00442-2
- GOLBY, A. J., GABRIELI, J. D. E., CHIAO, J. Y., & EBERHARDT, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience*, **4**, 845-850. doi:10.1038/90565
- GRILL-SPECTOR, K., KNOUF, N., & KANWISHER, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, **7**, 555-562. doi:10.1038/nn1224
- HANCOCK, K. J., & RHODES, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, **99**, 45-56. doi:10.1348/000712607X199981
- HERSHLER, O., & HOCHSTEIN, S. (2009). The importance of being expert: Top-down attentional control in visual search with photographs. *Attention, Perception, & Psychophysics*, **71**, 1478-1486. doi:10.3758/APP.71.7.1478

- HILLS, P. J., & LEWIS, M. B. (2006). Reducing the own-race bias in face recognition by shifting attention. *Quarterly Journal of Experimental Psychology*, *59*, 996-1002. doi:10.1080/17470210600654750
- ITO, T. A., & URLAND, G. R. (2003). Race and gender on the brain: Electro-cortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality & Social Psychology*, *85*, 616-626. doi:10.1037/0022-3514.85.4.616
- ITO, T. A., & URLAND, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*, 21-36. doi:10.3758/CABN.5.1.21
- JOHNSON, K. J., & FREDRICKSON, B. L. (2005). "We all look the same to me": Positive emotions eliminate the own-race in face recognition. *Psychological Science*, *16*, 875-881. doi:10.1111/j.1467-9280.2005.01631.x
- KANWISHER, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*, 759-763. doi:10.1038/77664
- LEBRECHT, S., PIERCE, L. J., TARR, M. J., & TANAKA, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PLoS ONE*, *4*, e4215. doi:10.1371/journal.pone.0004215
- MAURER, D., LE GRAND, R., & MONDLOCH, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*, 255-260. doi:10.1016/S1364-6613(02)01903-4
- MCGUGIN, R. W., & GAUTHIER, I. (2010). Perceptual expertise with objects predicts another hallmark of face perception. *Journal of Vision*, *10*(4, Art. 15), 1-12. doi:10.1167/10.4.15
- MICHEL, C., ROSSION, B., HAN, J., CHUNG, C.-S., & CALDARA, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, *17*, 608-615. doi:10.1111/j.1467-9280.2006.01752.x
- PASCALIS, O., DE HAAN, M., & NELSON, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*, 1321-1323. doi:10.1126/science.1070223
- PASCALIS, O., SCOTT, L. S., KELLY, D. J., SHANNON, R. W., NICHOLSON, E., COLEMAN, M., & NELSON, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences*, *102*, 5297-5300. doi:10.1073/pnas.0406627102
- RHODES, G., BRAKE, S., TAYLOR, K., & TAN, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, *80*, 313-331.
- 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. doi:10.1162/089892904322984508
- RHODES, G., HAYWARD, W. G., & WINKLER, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review*, *13*, 499-505.
- RICHLER, J. J., BUKACH, C. M., & GAUTHIER, I. (2009). Context influences holistic processing of nonface objects in the composite task. *Attention, Perception, & Psychophysics*, *71*, 530-540. doi:10.3758/APP.71.3.530
- RICHLER, J. J., CHEUNG, O. S., WONG, A. C.-N., & GAUTHIER, I. (2009). Does response interference contribute to face composite effects? *Psychonomic Bulletin & Review*, *16*, 258-263. doi:10.3758/PBR.16.2.258
- ROSSION, B., COLLINS, D., GOFFAUX, V., & CURRAN, T. (2007). Long-term expertise with artificial objects increases visual competition with early face categorization processes. *Journal of Cognitive Neuroscience*, *19*, 543-555. doi:10.1162/jocn.2007.19.3.543
- 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. doi:10.1111/1467-9280.00446
- SANGRIGOLI, S., PALLIER, C., ARGENTI, A.-M., VENTUREYRA, V. A. G., & DE SCHONEN, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*, 440-444.
- SCOTT, L. S., TANAKA, J. W., SHEINBERG, D. L., & CURRAN, T. (2008). The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research*, *1210*, 204-215. doi:10.1016/j.brainres.2008.02.054
- SHRIVER, E. R., YOUNG, S. G., HUGENBERG, K., BERNSTEIN, M. J., & LANTER, J. R. (2008). Class, race, and the face: Social context modulates the cross-race effect in face recognition. *Personality & Social Psychology Bulletin*, *34*, 260-274. doi:10.1177/0146167207310455
- STAHL, J., WIESE, H., & SCHWEINBERGER, S. R. (2008). Expertise and own-race bias in face processing: An event-related potential study. *NeuroReport*, *19*, 583-587. doi:10.1097/WNR.0b013e3282f97b4d
- TANAKA, J. W., CURRAN, T., & SHEINBERG, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, *16*, 145-151. doi:10.1111/j.0956-7976.2005.00795.x
- TANAKA, J. W., KIEFER, M., & BUKACH, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, *93*, B1-B9. doi:10.1016/j.cognition.2003.09.011
- TARR, M. J., & PINKER, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, *1*, 253-256. doi:10.1111/j.1467-9280.1990.tb00209.x
- WIESE, H., STAHL, J., & SCHWEINBERGER, S. R. (2009). Configural processing of other-race faces is delayed but not decreased. *Biological Psychology*, *81*, 103-109. doi:10.1016/j.biopsycho.2009.03.002
- WILLIAMS, N. R., WILLENBOCKEL, V., & GAUTHIER, I. (2009). Sensitivity to spatial frequency and orientation content is not specific to face perception. *Vision Research*, *49*, 2353-2362. doi:10.1016/j.visres.2009.06.019
- WONG, A. C.-N., PALMERI, T. J., & GAUTHIER, I. (2009). Conditions for facelike expertise with objects: Becoming a Ziggerin expert—but which type? *Psychological Science*, *20*, 1108-1117. doi:10.1111/j.1467-9280.2009.02430.x
- WONG, A. C.-N., PALMERI, T. J., ROGERS, B. P., GORE, J. C., & GAUTHIER, I. (2009). Beyond shape: How you learn about objects affects how they are represented in visual cortex. *PLoS ONE*, *4*, e8405.
- XU, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, *15*, 1234-1242. doi:10.1093/cercor/bhi006
- XU, Y., LIU, J., & KANWISHER, N. (2005). The M170 is selective for faces, not for expertise. *Neuropsychologia*, *43*, 588-597. doi:10.1016/j.neuropsychologia.2004.07.016
- YOUNG, A. W., HELLAWELL, D., & HAY, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747-759. doi:10.1068/p160747
- YUE, X., TJAN, B. S., & BIEDERMAN, I. (2006). What makes faces special? *Vision Research*, *46*, 3802-3811. doi:10.1016/j.visres.2006.06.017

## NOTE

1. We thank K. Grill-Spector for kindly sending us the stimuli used in her study, which allowed us to compare the types of cars used in the imaging study and those used in previous imaging studies.

(Manuscript received December 14, 2009;  
revision accepted for publication April 2, 2010.)