Journal of Experimental Psychology:
Human Perception and Performance
2009, Vol. 35, No. 1, 94–107

0096-1523/09/$12.00 DOI: 10.1037/0096-1523.35.1.94

A Visual Short-Term Memory Advantage for Objects of Expertise

Kim M. Curby and Kuba Glazek
Temple University

Isabel Gauthier
Vanderbilt University

Visual short-term memory (VSTM) is limited, especially for complex objects. Its capacity, however, is greater for faces than for other objects; this advantage may stem from the holistic nature of face processing. If the holistic processing explains this advantage, object expertise—which also relies on holistic processing—should endow experts with a VSTM advantage. The authors compared VSTM for cars among car experts and car novices. Car experts, but not car novices, demonstrated a VSTM advantage similar to that for faces; this advantage was orientation specific and was correlated with an individual's level of car expertise. Control experiments ruled out accounts based solely on verbal- or long-term memory representations. These findings suggest that the processing advantages afforded by visual expertise result in domain-specific increases in VSTM capacity, perhaps by allowing experts to maximize the use of an inherently limited VSTM system.

Keywords: faces, objects, expertise, visual short-term memory, holistic processing

Each of our interactions with the world is constrained by bottlenecks of information processing, including how many pieces of visual information we can retain in memory. Is there anything a person can do to increase his or her visual short-term memory (VSTM) capacity? Typically, people are able to retain only three to four objects in VSTM at any given time. One possibility is that VSTM capacity is determined by a fixed number of "slots" (three to four) that can hold one object each (Vogel, Woodman, & Luck, 2001). However, some researchers have pointed out that VSTM capacity may be limited by the complexity or number of features of the items stored (Alvarez & Cavanagh, 2004; Wheeler & Treisman, 2002). One possibility is that visual expertise can help an observer overcome such limitations. Visual experts process highly complex objects within their domain of expertise with relative ease, and they create qualitatively different "holistic" representations (Gauthier, Curran, Curby, & Collins, 2003; Tanaka & Sengco, 1997) that support faster identification judgments (Tanaka, 2001; Tanaka & Taylor, 1991) and that can be searched more efficiently (Tong & Nakayama, 1999). But do such advantages impact VSTM capacity?

Editor's Note. Mary A. Peterson served as the action editor for this article.—GWH

Kim M. Curby and Kuba Glazek, Department of Psychology, Temple University; Isabel Gauthier, Department of Psychology, Vanderbilt University.

This study was supported by National Science Foundation Grants 0091752 and SBE-0542013, National Institutes of Health Grant EY13441, and James S. McDonnell Foundation awards. We thank Gordon Logan, Randolph Blake, Jeff Schall, Dan Levin, Steve Most, and members of the Perceptual Expertise Network for helpful discussions and Ludvik Bukach for help with data collection.

Correspondence concerning this article should be addressed to Kim M. Curby, Department of Psychology, Temple University, 1701 North 13th Street, Philadelphia, PA 19122. E-mail: kim.curby@temple.edu

Recently, we showed that VSTM capacity for upright faces is larger than is that for other categories, such as cars, watches, or even inverted faces (Curby & Gauthier, 2007). One possibility is that humans are innately endowed with a greater memory capacity for upright faces due to the importance of face memory for survival. Alternatively, superior VSTM capacity for upright faces may be a product of human expertise with this category.

Currently, there is limited support for an impact of experience on VSTM capacity. Some evidence suggests that VSTM capacity remains stable from early in development (12 months; Rose, Feldman, & Jankowski, 2001) to adulthood (Luck & Vogel, 1997) and, thus, that it may be relatively inflexible. However, more recent studies have found that children's VSTM capacity for simple colored shapes doubles throughout childhood, from two items at 5 years of age to the adultlike capacity of three to four items by 10 years of age (Cowan et al., 2005; Riggs, McTaggart, Simpson, & Freeman, 2006). Similar increases in capacity with development have been reported for verbal short-term memory (Cowan, Nugent, Elliott, Ponomarev, & Saults, 1999). It is unclear, at the moment, which aspects of cognitive development could be the basis of such dramatic changes.

Other researchers have employed training regimens to investigate whether VSTM capacity is influenced by experience, and they have generally found little or no training effect. For example, in a recent study exploring the influence of domain-specific training on VSTM for novel objects, Chen, Eng, and Jiang (2006) reported that participants who viewed a set of eight random polygons 160 times in the context of a VSTM task were no more accurate in detecting a change in a VSTM array that contained these trained shapes than a change in an array that contained unfamiliar shapes. It is important to note that participants in this study could accurately identify the trained polygons in a two-alternative forced-choice task, thus confirming that they did have representations of these items in long-term memory (LTM). Notably, VSTM performance did improve with practice but did so equally for trained and untrained polygons. Thus, it is unclear whether this change in performance represents a general effect of practice on VSTM performance or a more specific influence of experience that generalizes to new exemplars within the trained category. It is also possible that more extensive training, such as that required for development of perceptual expertise with a category of objects, is required to induce a change in VSTM capacity beyond that attributed more generally to practice.

In a recent neuroimaging study, Moore, Cohen, and Ranganath (2006) adopted a more extensive (10.5-hr) training paradigm to explore the impact of experience on the neural substrates that support VSTM. Training increased activity during both encoding and maintenance of artificial objects in the classic VSTM network, including the bilateral dorsolateral prefrontal, posterior parietal, and occipitotemporal cortices. In contrast, the lateral occipital cortex and the fusiform face area (FFA) showed expertise effects during encoding only. These changes in the functional network supporting VSTM for trained stimuli are consistent with the suggestion that behavioral changes in VSTM capacity may be possible after extensive training. However, VSTM capacity was not measured in this study, so it is difficult to know whether increases in capacity would correlate more with changes in encoding or in maintenance.

Some findings suggest that the superior VSTM documented in chess experts, relative to chess novices, for configurations of chess pieces can be considered evidence that experience can increase the capacity of VSTM (Chase & Simon, 1973). However, this advantage is believed to rely on specific stored representations in long-term memory, rather than a more qualitative change in the way information is stored in VSTM (Chase & Simon, 1973). Support for this LTM account of expert chess memory comes from the finding that intervening short-term memory tasks during the retention interval do not impact memory performance for familiar chess positions (Charness, 1976). Practice appears to increase chess experts' VSTM performance by allowing the "chunking" of information into larger units in long-term memory, which can be accessed through "pointers" stored in short-term memory (Chase & Simon, 1973; Freyhof, Gruber, & Ziegler, 1992; Gobet & Simon, 1998). Therefore, although previous studies have demonstrated a benefit of experience or expertise on VSTM capacity, these advantages have been shown to reflect the utilization of additional resources (e.g., LTM or verbal memory) to supplement VSTM rather than a change to VSTM capacity (Charness, 1976; Chase & Ericsson, 1981).

Here, we are concerned with a different way in which experience may influence VSTM. Specifically, the impact of the perceptual organization of information on VSTM capacity (Delvenne & Bruyer, 2004; Luck & Vogel, 1997; Vogel et al., 2001; Xu, 2002, 2006) may provide a potential avenue for extensive learning to influence VSTM capacity. For example, VSTM capacity is greater for features presented in the form of a unified object than for those presented in isolation (Luck & Vogel, 1997). Thus, in addition to facilitating the recruitment of additional capacity from other systems, as in the case of chess experts, experience may impact VSTM capacity more directly because of a change in the manner in which an item is encoded and/or represented in VSTM. For example, faces and other objects of expertise are processed more holistically than are objects of non-expertise, which are processed in a more feature-based manner (Busey & Vanderkolk, 2005; Gauthier et al., 2003; Gauthier & Tarr, 2002). Classic holistic processing effects typically found with faces, such as sensitivity to inversion or difficulty in selectively attending to an object part presented in the context of a whole object, have been demonstrated among observers trained to become experts with a novel category (Gauthier & Tarr, 2002) and among real-world car experts (Gauthier et al., 2003; but see Robbins & McKone, 2007). Thus, expertise with faces or with chess appears to lead to the processing and/or storage of information in larger units or chunks. However, unlike the memory strategy used by chess experts, which operates over meaningful arrangements of multiple independent, self-contained objects, holistic processing of faces and other expert categories operates within a single object (e.g., a car). This is not unlike other object-based perceptual advantages previously reported in the literature (e.g., Egly, Driver, & Rafal, 1994; Saiki & Hummel, 1998; Xu, 2006). Notably, holistic processing has been related to activity in the FFA (Gauthier & Tarr, 2002; Rotshtein, Geng, Driver, & Dolan, 2007), which is engaged during the encoding of faces (Druzgal & D'Esposito, 2001, 2003) and objects of expertise (Moore et al., 2006). This processing strategy has also been associated with the earliest face-specific electrophysiological potential, the N170, which occurs only 170 ms after the onset of a stimulus (Gauthier et al., 2003). Therefore, it is possible that the VSTM advantage for faces, rather than relying on LTM, stems from differences in face processing at a more perceptual level.

In sum, although there is a documented VSTM advantage for upright faces (Curby & Gauthier, 2007) and the literature has suggested that most observers are experts with upright but not inverted faces, these findings can only indirectly support inferences about the effect of perceptual expertise on VSTM capacity. In the current studies, we specifically set out to test if expertise can increase VSTM capacity by comparing experts and novices with a nonface category: cars. Previous studies have found that cars are processed more holistically among car experts than among car novices (Gauthier et al., 2003) and that a subject's car expertise is related to the amount of activity in response to cars in the part of the visual system most responsive to faces, the FFA (Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier, Curby, Skudlarski, & Epstein, 2005; Xu, 2005). Activity in this area has also been linked with holistic processing (Gauthier & Tarr, 2002; Rotshtein et al., 2007).

Experiment 1

Experiment 1 assessed VSTM capacity for upright and inverted faces and cars among participants with a range of perceptual expertise with cars. Given our hypothesis that holistic processing underlies the VSTM advantage for faces and because inversion disrupts holistic processing (Tanaka & Sengco, 1997), we predicted an expert VSTM advantage for upright, but not inverted, faces and cars. In addition, we varied encoding time up to 4,000 ms to ensure that VSTM was not limited by encoding speed, as complex objects require more time to be encoded in VSTM than do simple objects (Curby & Gauthier, 2007; Eng, Chen, & Jiang, 2005).

Method

Participants. Thirty-six individuals with a range of experience in identifying cars participated for payment. Participants were employees, undergraduate students, or graduate students of Vanderbilt University or members of the surrounding Nashville community. All had normal or corrected-to-normal vision. A self-report measure of participants' car and bird expertise was obtained in the form of a rating on a scale of 1 to 10. Participants were informed that 5 corresponded to average skill at identifying cars or birds, whereas 10 reflected perfect skill in recognizing these categories. An objective measure of car expertise was obtained with a sequential matching task used in previous studies (Gauthier et al., 2003, 2005; Gauthier, Skudlarski, Gore, & Anderson, 2000). In this task, participants were required to make a same/different judgment about different images of cars at the level of model, regardless of year (see Figure 1A in Gauthier et al., 2003). This task can be performed at least to some minimal degree by all participants, regardless of their level of experience with cars, as it does not require knowledge of car names. To provide a baseline of their perceptual skills, we asked participants to perform the same task with birds (i.e., to make a same/different decision at the level of species about different images of passerine birds). A car expertise index was defined as (car d' − bird d'). Participants with a car expertise index ≥1 and a d' for cars ≥2 were classified as experts (Gauthier et al., 2000).

Eighteen participants (11 male) met the criteria for car expertise (age, M = 22.28 years, SD = 4.71; car d' M = 2.55, bird d' M = 0.87), and the remaining 18 participants (10 male) were classified as car novices (age, M = 20.64 years, SD = 2.42; car d' M = 1.34, bird d' M = 0.84).[1] One of the participants classified as an expert had a car d' of 2.24 but a car d' − bird d' of 0.83; he was included in the car expertise group because he also reported having above-average skills at recognizing birds, which likely resulted in the smaller difference between the d' measures for these categories. Car expertise scores from the matching task were generally consistent with subject self-reports; those participants who were classified as novices reported their car recognition skills at an average of 5.77/10. Those who met criteria for car expertise rated their skills, on average, as 8.00/10.

Stimuli. The stimuli were 72 gray scale faces (1.9° × 2.3°) from the Max Planck Institute for Biological Cybernetics in Tübingen, Germany (Troje & Bülthoff, 1996), and 72 gray scale images of cars, gathered from various public websites (2.3° × 1.5°, profile view).

Procedure. For each participant, half the faces and half the cars appeared in the upright trials. The remaining images appeared in the inverted trials. Participants performed a delayed match-to-sample probe recognition task simultaneously with an articulatory suppression task (see Figure 1). The sequence of events in each trial was as follows: Participants were first presented aurally with two digits and a mask, which they overtly rehearsed throughout the trial to prevent verbal rehearsal. Then, the study array, which consisted of one, three, or five faces or cars, either all upright or all inverted and evenly spaced in a circle (6.1° diameter), appeared for 500, 2,500, or 4,000 ms. After a 1,200-ms delay, a face or car probe was presented in one of the locations from the study array. The probe remained until participants indicated with a keypress whether the probe was the same as (50% of trials) or different from the one that had appeared in that location in the study array. Within each trial, to minimize confusion, the probe was never an item that had appeared at a different location in the study array. After a response was made, a screen with two digits appeared and participants were required to state whether the two digits on the screen were the same as those they had been rehearsing throughout the trial.

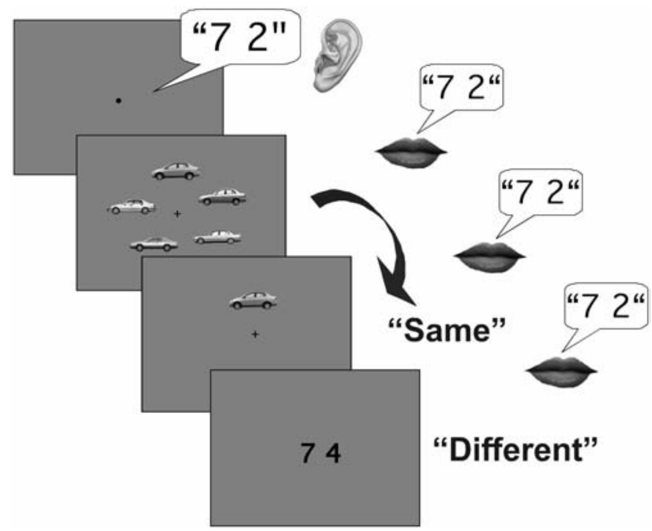

Figure 1. The sequence of events in each trial: First, participants were presented auditorily with two digits and a mask, which they overtly rehearsed throughout the trial to prevent verbal rehearsal. Then, the study array appeared for 500, 2,500, or 4,000 ms. After a 1,200-ms delay, a face or car probe was presented in one of the locations from the study array. The probe remained until participants indicated with a keypress whether the probe was the same as or different from the one that had appeared in that location in the study array. After a response was made, a screen with two digits appeared and participants were required to state whether the two digits on the screen were the same as those they had been rehearsing throughout the trial.

Participants performed a total of 1,152 trials across four different sessions, each consisting of eight alternating blocks of upright and inverted images (36 trials/block, randomized for set size and presentation duration). Two sessions consisted only of face trials, whereas the other two sessions consisted only of car trials. Session order was counterbalanced within and across expertise groups. In sum, there were 288 trials for each of the four categories (upright faces, inverted faces, upright cars, inverted cars). For each category, there were nine conditions (three set sizes × three durations), presented 32 times each.

Analysis. Incorrect articulatory suppression trials (<2%) were discarded. For each participant and condition, the number of objects successfully encoded in VSTM was estimated with Cowan's K, where K = (hit rate + correct rejection rate − 1) × set size (Cowan, 2001). The maximum K (K max) was identified for each duration, regardless of set size. All analyses were performed on the K max values. In addition to analyses of variance (ANOVAs) and regression analyses exploring the relationship between level of car expertise and VSTM capacity, we conducted a series of planned t tests to explore the following predictions based on the proposed

[1] The novice data in Experiment 1 are a subset of the data reported in Experiment 3 in Curby and Gauthier (2007). We limited novice data to ensure that session order was counterbalanced across expert and novice groups.

role of holistic processing in increasing VSTM capacity: (a) the presence of an inversion cost for cars among car experts but not novices, (b) greater VSTM for cars among car experts compared with novices, and (c) greater VSTM capacity for faces than for cars among car novices, with sufficient encoding time.

Results

Both car experts and car novices demonstrated an inversion cost for faces, whereas only car experts experienced such a cost for cars (see Figures 2A and 2B). Car experts also demonstrated greater VSTM for upright cars than did novices when the presentation duration was sufficiently long (≥2,500 ms). Furthermore, although VSTM for faces was not different from that for cars among car experts regardless of presentation duration, car novices demonstrated an advantage for upright faces over upright cars at the longest presentation (4,000 ms). With a 4,000-ms presentation, car expertise was correlated with VSTM for cars but not for faces (see Figures 2C and 2D).

VSTM for faces among car experts and novices. For faces, a 2 (orientation: upright, inverted) × 3 (duration: 500, 2,500, 4,000 ms) × 2 (group: novice, expert) ANOVA on *K* max revealed main effects of orientation, $F(1, 34) = 69.53$, $p \leq .0001$, and duration, $F(2, 68) = 40.24$, $p \leq .0001$, but no main effect ($F < 1$) or interaction involving car expertise (all *p*s > .425). The interaction between orientation and duration failed to reach significance, $F(2, 68) = 1.77$, $p = .178$. In sum, face VSTM was greater for longer presentations and for upright than for inverted faces but was not influenced by car expertise.

VSTM for cars among car experts and novices. For cars, a 2 (orientation: upright, inverted) × 3 (duration: 500, 2,500, 4,000 ms) × 2 (group: novice, expert) ANOVA on *K* max revealed main effects of orientation, $F(1, 34) = 9.12$, $p = .0048$, and duration,

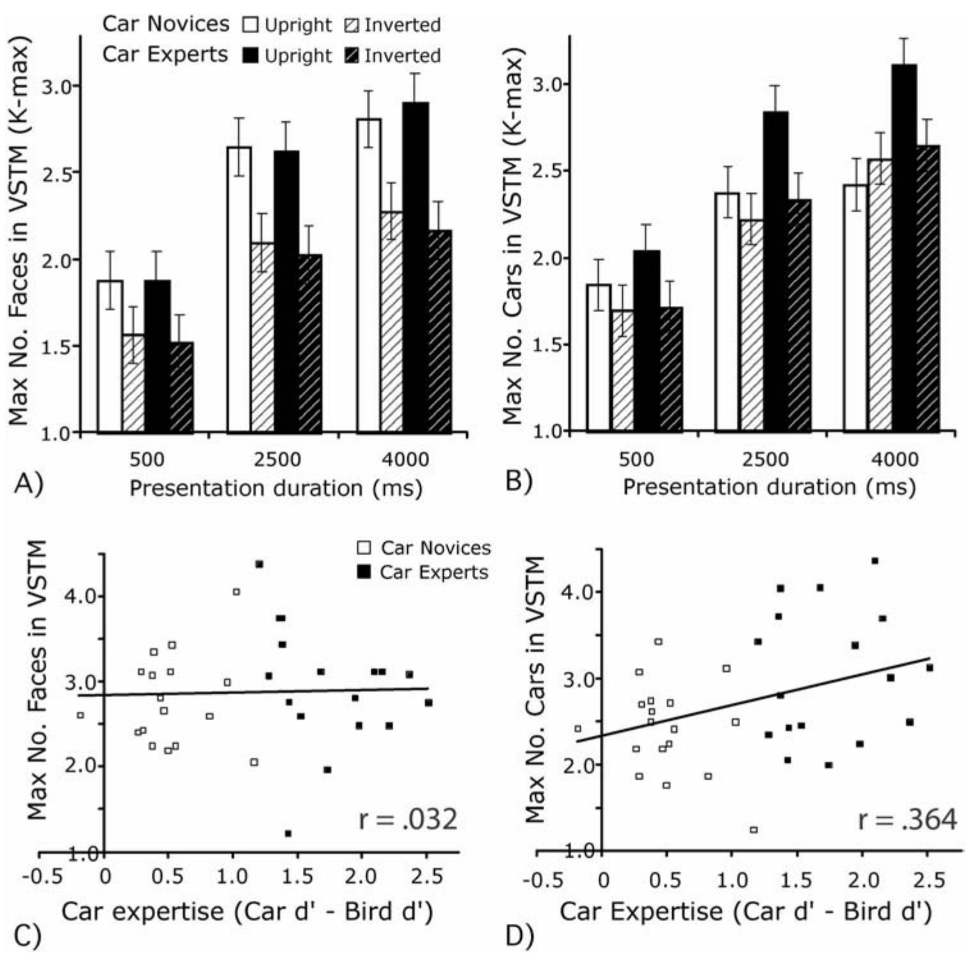

Figure 2. The maximum number of objects (*K* max) in visual short-term memory (VSTM) for 500-, 2,500-, and 4,000-ms presentation durations for (A) upright and inverted faces and (B) upright and inverted cars among participants who were car experts and novices. There was a VSTM advantage for upright cars among car experts that was similar in magnitude to that for upright faces. Car experts, but not novices, showed an inversion effect for cars. Error bars represent pooled standard error values. Scatterplots of individuals' car expertise scores (car $d' -$ bird d') and *K* max when the memory array was presented for 4,000 ms illustrate (C) the absence of a significant correlation between a participant's level of car expertise and VSTM capacity for upright faces, but (D) the presence of a significant correlation with VSTM capacity for upright cars.

$F(1, 34) = 61.93$, $p \leq .0001$, but no interaction between orientation and duration ($F < 1$). Although there was no main effect of car expertise, $F(1, 34) = 1.95$, $p = .172$, there was an interaction between expertise and orientation, $F(1, 34) = 5.61$, $p = .024$. Interactions between duration and expertise and/or orientation failed to reach significance (all $ps \geq .230$). In sum, VSTM was generally greater for longer presentations and, relative to novices, car experts showed superior VSTM for upright, but not inverted, cars.

Planned comparisons. K max for cars was greater for upright compared with inverted orientations for all durations among experts (all $ps < .022$) but not among novices (all $ps > .206$), whereas both groups demonstrated an advantage for upright orientations for faces regardless of duration (experts, all $ps < .010$; novices, all $ps < .033$). In addition, with sufficient presentation duration, K max for upright faces reliably exceeded that for upright cars among car novices, 500 ms, $t < 1$; 2,500 ms, $t(17) = 1.48$, $p = .157$; 4,000 ms, $t(17) = 2.87$, $p = .011$, but not among experts, 500 ms, $t(17) = 1.19$, $p = .251$; 2,500 ms, $t(17) = 1.16$, $p = .263$; 4,000 ms, $t(17) = 1.17$, $p = .257$. A VSTM advantage for cars among car experts compared with novices emerged when the presentation duration was sufficiently long: 500 ms, $t < 1$; 2,500 ms, $t(34) = 1.68$, $p = .103$; 4,000 ms, $t(34) = 3.05$, $p = .004$.

Correlation between car expertise and VSTM for cars and faces. With a 4,000-ms presentation, K max for upright cars was correlated with participants' car expertise index ($r = .364$, $p = .032$). Notably, car expertise indexes were not correlated with VSTM for upright faces ($r = .032$, $p = .885$) or inverted faces ($r = .000$, $p = .951$) or for inverted cars ($r = .036$, $p = .838$).[2]

Discussion

Consistent with the proposed influence of perceptual expertise on VSTM capacity, car experts demonstrated an orientation-dependent VSTM advantage for cars. As with the VSTM advantage for faces, this advantage depended on sufficient encoding time (Curby & Gauthier, 2007). These results suggest that the VSTM advantage for faces is not due to a face-specific mechanism; other objects within a domain of expertise can demonstrate this advantage.

The expert advantage reported here may stem, as we predicted, from differences in perceptual processing, but one alternative is that experts benefited from better knowledge of car names, possibly leading to a contribution from verbal short-term memory (Olsson & Poom, 2005). Experiment 2 explored this hypothesis.

Experiment 2

The articulatory suppression load we used in Experiment 1 may have been insufficient to prevent a contribution from verbal short-term memory; according to Marsh and Hicks (1998),[3] participants can perform a verbal memory task with reasonable accuracy (82%) with an articulatory suppression load equivalent to the two syllables we used in Experiment 1, but their performance drops considerably (54%) when the load is increased to six syllables. Thus, we increased the articulatory suppression load to five to six syllables in Experiment 2. We also used a semantically relevant load to interfere further with any verbal rehearsal strategy; participants were required to rehearse three car model names during car trials and three person names during face trials.

Method

Participants. Thirty-one individuals, whose car expertise was quantified as in Experiment 1, participated for payment. Fourteen participants (11 male) met the criteria for car expertise (age, $M = 21.64$ years, $SD = 2.10$; car d' $M = 2.72$, bird d' $M = 0.93$), and 17 participants (13 male) were classified as car novices (age, $M = 22.41$ years, $SD = 3.02$; car d' $M = 1.27$, bird d' $M = 0.76$).[4]

Stimuli, procedure, and analyses. The stimuli, procedure, and data analysis were as in Experiment 1, but instead of rehearsing digits, participants rehearsed three car models (e.g., "Spectra, Blazer, Accord"; no overlap with models from visual task) or three personal names (e.g., "Leanne, Amy, Cathryn"). These were probed auditorily at the end of each trial.

Results

As in Experiment 1, both car experts and novices demonstrated an inversion cost for faces, but only car experts experienced such a cost for cars (see Figures 3A and 3B). Furthermore, although VSTM for faces was no different from that for cars among car experts regardless of presentation duration, car novices demonstrated an advantage for upright faces over upright cars when the stimulus presentation was sufficiently long (4,000 ms). In addition, car experts demonstrated greater VSTM for upright cars than did novices only when presentation was sufficiently long ($\geq$2,500 ms). With 4,000-ms presentation, car expertise was correlated with VSTM for cars but not for faces (see Figures 3C and 3D).

VSTM for faces among car experts and novices. For faces, a 2 (orientation: upright, inverted) $\times$ 3 (duration: 500, 2,500, 4,000 ms) $\times$ 2 (group: novice, expert) ANOVA on K max revealed main effects of orientation, $F(1, 29) = 32.61$, $p \leq .0001$, and duration, $F(2, 58) = 40.79$, $p \leq .0001$, but no main effect, $F(1, 29) = 2.12$, $p = .156$, or interaction involving car expertise ($Fs < 1$). There was no reliable interaction between orientation and duration ($F < 1$). In sum, VSTM for faces was greater for longer presentations and for upright faces, but car expertise did not impact VSTM for faces.

VSTM for cars among car experts and novices. For cars, a 2 (orientation: upright, inverted) $\times$ 3 (duration: 500, 2,500, 4,000 ms) $\times$ 2 (group: novice, expert) ANOVA on κ max revealed main effects of orientation, $F(1, 29) = 11.24$, $p = .002$, and duration, $F(1, 29) = 33.59$, $p \leq .0001$, but not of car expertise, $F(1, 29) = 2.71$, $p = .110$. The interaction between orientation and duration was not reliable, $F(2, 58) = 1.10$, $p = .340$, but there were interactions between car expertise and orientation, $F(1, 29) = 7.40$, $p = .011$, and car expertise and duration, $F(1, 29) = 3.64$, $p = .032$. Compared with car novices, car experts demonstrated greater VSTM for upright, but not inverted, cars, and they also benefited more from additional encoding time than did novices. The interaction between duration, group, and orientation was not reliable ($F < 1$).

[2] One participant who reported having above-average skills in bird recognition was not included in the regression analyses.

[3] This level of articulatory load is commonly used in VSTM studies and is often assumed to be adequate to prevent verbal contamination of VSTM performance (Luck & Vogel, 1997; Todd & Marois, 2004; Vogel et al., 2001).

[4] Car expertise scores were consistent with average self-report ratings of skill in identifying cars (novices, 6.03/10; experts, 8.43/10).

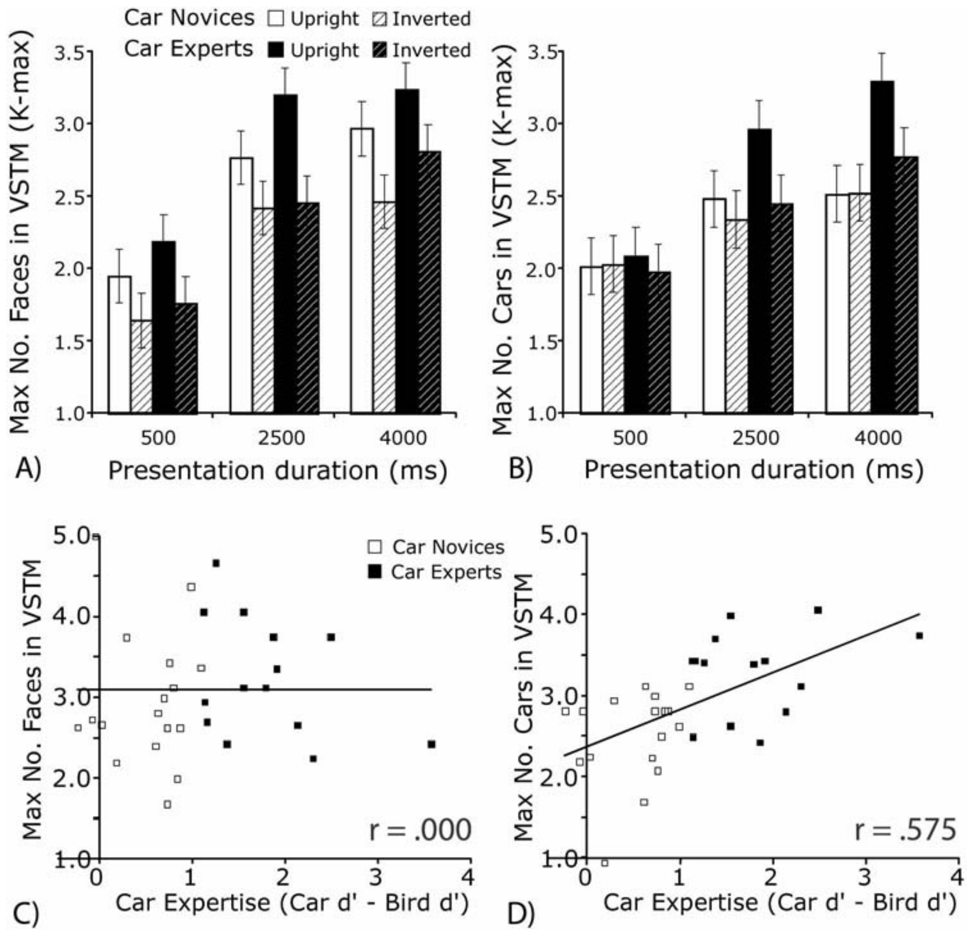

Figure 3. The maximum number of items (*K* max) in visual short-term memory (VSTM) for 500-, 2,500-, and 4,000-ms presentation durations for (A) upright and inverted faces and (B) upright and inverted cars among participants who were car experts or car novices. There was a VSTM advantage for upright cars among car experts that was similar in magnitude to the advantage for upright faces. Car experts, but not novices, showed an inversion effect for cars. Scatterplots of individuals' car expertise scores (car $d' - $ bird d') and the *K* max when the memory array was presented for 4,000 ms illustrate (C) the absence of a significant correlation between a participant's level of car expertise and VSTM capacity for upright faces, but (D) the presence of a significant correlation with VSTM capacity for upright cars.

Planned comparisons. Our predictions were confirmed by planned one-tailed *t* tests: *K* max for cars among car experts showed an advantage for upright over inverted orientations for presentations of 2,500 ms or longer: 500 ms ($t < 1$); 2,500 ms, $t(13) = 3.47$, $p = .002$; 4,000 ms, $t(13) = 2.52$, $p = .013$. In contrast, among car novices, inversion failed to impact VSTM capacity for cars regardless of duration (all ts < 1). VSTM capacity was greater for upright than inverted faces for all durations among both experts (all ps $< .05$) and novices (all ps $< .02$). Furthermore, *K* max for upright faces reliably exceeded that for upright cars among car novices at the longest encoding duration: 500 ms, $t < 1$; 2,500 ms, $t(16) = 1.27$, $p = .112$; 4,000 ms, $t(16) = 2.09$, $p = .027$. However, VSTM for cars and faces did not differ among car experts regardless of duration: 500 ms, $t < 1$; 2,500 ms, $t(13) = 1.25$, $p = .117$; 4,000 ms, $t < 1$. A VSTM advantage for cars among car experts compared with car novices emerged with presentations longer than 500 ms: 500 ms, $t < 1$; 2,500 ms, $t(29) = 2.42$, $p = .011$; 4,000 ms, $t(29) = 3.90$, $p = .0003$.

Correlation between car expertise and maximum VSTM for cars and faces. *K* max for upright cars with a 4,000-ms presentation was correlated with participants' car expertise ($r = .575$, $p = .0007$). Notably, car expertise was not correlated with VSTM for inverted cars ($r = -.027$, $p = .891$)[5] or for upright faces ($r = .000$, $p = .996$) or inverted faces ($r = .184$, $p = .322$).

Discussion

Despite an increase in the articulatory suppression load, the VSTM advantage for faces and other objects of expertise remained

[5] The correlation between inverted car VSTM and car expertise approached significance ($r = .327$, $p = .072$), but this result was carried by two outliers (>2 *SD* above the mean).

intact; car experts, when given sufficient encoding time, demonstrated not only greater VSTM capacity for cars but also a greater inversion cost for these stimuli than did novices. Similarly, novices demonstrated greater VSTM for faces than for cars when given sufficient encoding time. Overall performance was also similar across Experiments 1 and 2. These results suggest that the greater VSTM capacity for cars among car experts does not rely on a contribution from verbal memory. It is possible that the knowledge of a label for a stimulus may change the manner in which it is processed in the VSTM task, regardless of whether the label is explicitly accessed or used to aid recall. However, assuming a common underlying cause for the VSTM advantage demonstrated for faces and for cars among car experts, as suggested by the similar qualitative and quantitative nature of these two effects, the presence of this advantage for unfamiliar faces with no known labels provides evidence against this account (e.g., Experiments 1 and 2; see also Curby & Gauthier, 2007).

Although the expert VSTM advantage does not appear to depend on a contribution from verbal short-term memory, it is possible that experts are better able to recruit or establish representations in LTM to aid VSTM performance. Experiment 3 explored this possibility.

Experiment 3

Among chess experts, LTM has been shown to play an important role in their superior ability to remember meaningful configurations of chess pieces (Chase & Simon, 1973). Chess experts store large chunks of information about the spatial configuration of items in long-term memory and recall them through a simple cue stored in VSTM. Similarly, it is possible that car experts' VSTM advantage may depend on stimulus-specific representations in LTM (Gobet & Simon, 1998).

Visual expertise is clearly an example of long-term learning, and it is reasonable to argue that any task recruiting such expertise must rely on at least some form of LTM. Experiment 3 explored whether the expert VSTM advantage depends on stimulus-specific LTM representations (Ericsson & Kintsch, 1995). Specifically, it is possible that the large stimulus sets (72 items/category) we employed in Experiments 1 and 2 allowed participants to use information in LTM. Each item appeared infrequently (approximately 1:10), and the familiarity of true probes (relative to foils) could serve as a useful cue to aid performance. Furthermore, experts' superior ability to distinguish exemplars might increase the reliability of familiarity cues. In Experiment 3, we used a small stimulus set to increase the frequency of item repetition and thus reduced the usefulness of LTM traces through the buildup of proactive interference. Additionally, we incorporated a manipulation check by changing the stimulus set partway through the experiment: If participants could use LTM traces despite increased proactive interference, their performance should drop when the stimulus set change occurred. Thus, Experiment 3 explored the potential role of stimulus-specific representations in LTM in contributing to the VSTM advantage for objects of expertise.

Method

Participants. Thirty-six individuals, whose car expertise was quantified as in Experiment 1, participated for payment. Of the participants, 18 (13 male) met our criteria for car expertise (age, $M = 25.3$ years, $SD = 4.54$; car d' $M = 2.84$, bird d' $M = 1.02$) and the remaining 18 (8 male) were classified as novices (age, $M = 27.2$ years, $SD = 8.36$; car d' $M = 0.70$, bird d' $M = 0.83$).[6]

Stimuli. The stimuli were 40 gray scale images of faces ($1.9° \times 2.3°$) and 40 profile views of cars ($2.3° \times 1.5°$).

Design, procedure, and analysis. The design and procedure were similar to those used in Experiment 1, except that the set size was fixed at 5 and all images were upright and presented for 500 ms or 4,000 ms. Participants performed six blocks of 36 trials for each category (faces, cars), for a total of 432 trials. Trials for each category were performed in two separate (216-trial) sessions, with order of sessions counterbalanced across expert and novice groups. Twenty faces and 20 cars were randomly selected for each participant. Stimuli presented in the first four blocks for each category were selected from a subset of 10 images. After four blocks of trials, the stimulus set was switched to the remaining 10 images. Each item appeared on average in every second trial.

Results

Neither a switch in stimulus set partway through the experiment nor the smaller size of the stimulus set eliminated the expert VSTM advantage (see Figure 4A). Once again, a short encoding duration (500 ms) yielded relatively low VSTM capacity, which was not qualified by category or expertise effects. In the 4,000-ms stimulus duration condition, only performance for cars among car novices was reduced by the stimulus set switch. VSTM capacity for faces exceeded that for cars among novices, but not among experts, for both pre- and post-switch stimulus sets. In addition, the VSTM advantage for cars among car experts, compared with novices, was apparent in the post-switch, but not the pre-switch, stimulus set condition. Similarly, the correlation between car expertise and VSTM for cars was reliable for the post-switch, but not the pre-switch, condition (see Figures 4B and 4C).

VSTM for faces among car experts and novices. For faces, a 2 (duration: 500 ms, 4,000 ms) $\times$ 2 (stimulus set: pre-switch, post-switch) $\times$ 2 (group: expert, novice) ANOVA revealed a main effect of duration, $F(1, 34) = 135.50$, $p \leq .0001$, but no effect of stimulus set ($F < 1$) or car expertise, $F(1, 34) = 1.00$, $p = .324$. In addition, no interactions between duration and/or expertise and/or stimulus set were reliable (all ps $> .196$). In sum, VSTM for faces was greater for longer presentations, but neither car expertise nor the change in stimulus set affected VSTM for faces.

VSTM for cars among car experts and novices. For cars, a 2 (duration: 500 ms, 4,000 ms) $\times$ 2 (stimulus set: pre-switch, post-switch) $\times$ 2 (group: expert, novice) ANOVA revealed main effects of duration, $F(1, 34) = 73.39$, $p \leq .0001$, and car expertise, $F(1, 34) = 4.30$, $p = .046$, but not of stimulus set ($F < 1$). In addition, there was an interaction between duration and expertise, $F(1, 34) = 8.19$, $p = .007$, with car experts demonstrating a VSTM advantage only for cars for the long presentation. The interaction between duration, expertise, and stimulus set approached reliability, $F(1, 34) = 3.37$, $p = .075$, with the novices' VSTM perfor-

[6] Those participants who met the criteria for expertise on the task rated themselves an average of 8.42/10; those who were classified as novices rated their skills, on average, as 3.58/10.

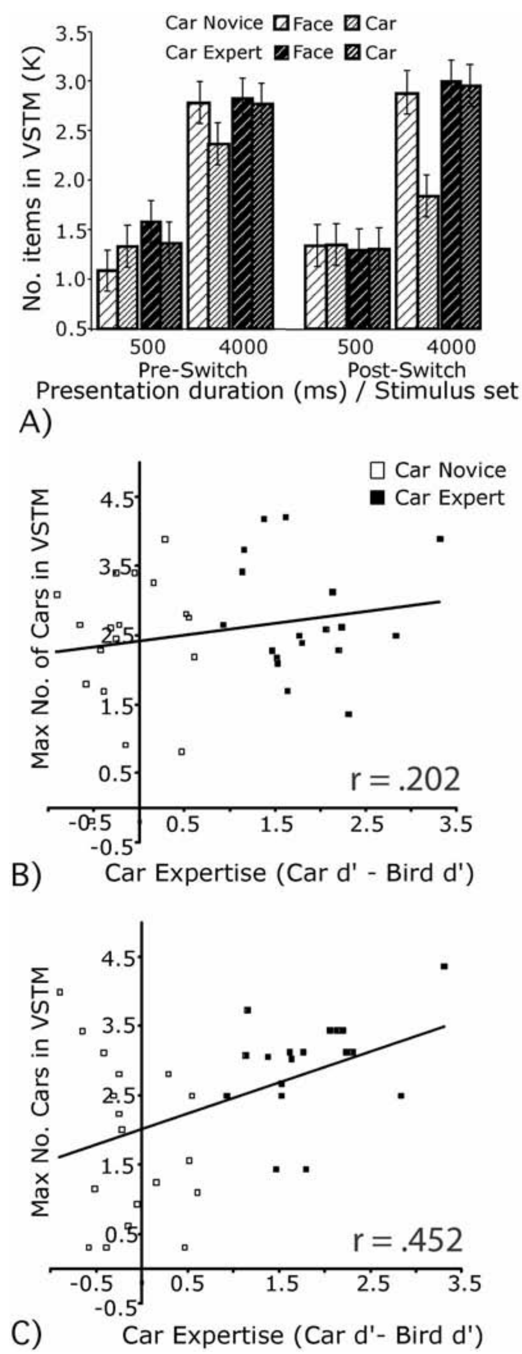

Figure 4. (A) The maximum number of upright faces or cars (*K* max) in visual short-term memory (VSTM) before and after a stimulus set change. Participants had performed 140 trials with the same small (10-item) stimulus set before the stimulus set was switched, and they performed 70 additional trials. Participants were either car experts or car novices, and the memory array was presented for 500 or 4,000 ms. Only novice VSTM performance with cars in the 4,000-ms presentation duration condition was influenced by the change in stimulus set. Scatterplots show individuals' car expertise scores (car $d'-$ bird d') and the *K* max for cars (B) pre- and (C) post-stimulus set switch. There was a significant correlation between a participant's level of car expertise and that participant's VSTM capacity for cars in the post- but not the pre-stimulus switch condition.

mance dropping after the switch in stimulus set but experts' performance showing a slight increase.

Planned comparisons. Planned two-tailed *t* tests explored the effect of the switch in stimulus set on VSTM capacity. Among car experts, capacity was unaffected by the switch, regardless of duration for both cars (ts < 1) and faces, 500 ms, $t(17) = 1.28$, $p = .218$; 4,000 ms, $t < 1$. Among novices, VSTM performance was reduced for cars only at the longest duration, 4,000 ms, $t(17) = 2.16$, $p = .045$; 500 ms, $t < 1$, and face VSTM performance was unaffected by the switch, regardless of duration, 4,000 ms, $t < 1$; 500 ms, $t(17) = 1.21$, $p = .241$.

In addition, planned one-tailed *t* tests revealed that among car novices, VSTM for upright faces exceeded that for upright cars, as long as the presentation was sufficiently long (4,000 ms), regardless of stimulus set: pre-switch, 500 ms, $t(17) = 1.24$, $p = .883$; 4,000 ms, $t(17) = 2.08$, $p = .027$; post-switch, 500 ms, $t < 1$; 4,000 ms, $t(17) = 3.18$, $p = .0027$. Among car experts, VSTM for cars and faces did not differ for either stimulus set regardless of duration: pre-switch, 500 ms, $t(17) = 1.32$, $p = .103$; 4,000 ms, $t < 1$; post-switch, 500 ms, $t < 1$; 4,000 ms, $t < 1$. A VSTM advantage for cars among car experts compared with car novices emerged only in the 4,000-ms presentation condition for the post-switch stimulus set, 500 ms, $t < 1$; 4,000 ms, $t(34) = 3.50$, $p = .0007$. This car advantage in car experts failed to reach significance regardless of duration for the pre-switch stimulus set, despite a trend in the long presentation condition: 500 ms, $t < 1$; 4,000 ms, $t(34) = 1.29$, $p = .103$.

Correlation between car expertise and VSTM for cars and faces. *K* max for cars when the memory array was presented for 4,000 ms reliably correlated with car expertise but only for the post-switch stimulus set (pre-switch, $r = .202$, $p = .236$; post-switch, $r = .452$, $p = .006$). Notably, car expertise did not correlate with VSTM capacity for faces for either stimulus set (pre-switch, $r = .028$, $p = .870$; post-switch, $r = .074$, $p = .667$).

Discussion

Neither the face advantage nor the expertise advantage for cars among car experts was eliminated by the buildup of proactive interference in LTM, which should have increased due to the small stimulus set in Experiment 3. In addition, consistent with the hypothesis that experts were not using stimulus-specific representations in LTM to aid their VSTM performance, no detectable cost to expert VSTM capacity was found when the stimulus set was replaced by an entirely different set of items after a period of learning. These results suggest that the expert VSTM advantage does not depend on access to stimulus-specific representations in LTM.

In contrast with the results from Experiments 1 and 2, in which a larger stimulus set was used, VSTM capacity for cars was not significantly higher in experts than in novices in the pre-switch condition (although there was a trend for such an effect). The use of a small stimulus set in Experiment 3 may have facilitated novices' performance. Novices typically use a feature-based strategy to distinguish items (e.g., relying on the length of the trunk or the angle of the windscreen on a car); thus, the use of a small set of items in Experiment 3 would make such features even more diagnostic and would allow novices to adequately represent and distinguish items in VSTM. The success of such a strategy in novices would render experts' advantage for encoding highly complex objects moot.

The drop in VSTM for cars among car novices after a change in stimulus set could reflect the relative inflexibility of feature-based strategies in which the cars are identified by a single salient feature. After all, the relative usefulness of a feature as a distinguishing characteristic would critically depend on the variability of that feature among the items in the stimulus set. A feature-based strategy could be quite effective for a small stimulus set, consistent with the high-level of performance in the prechange condition among novices. But this strategy would presumably be suboptimal for transfer to a different set of objects, as the same features are unlikely to be diagnostic across stimulus sets. In contrast, the more holistic perceptual strategy, believed to be recruited for faces by both car experts and car novices and for cars among car experts, would transfer equally well to a new stimulus set. Therefore, the drop in car VSTM performance among novices after the stimulus set switch may reflect the different encoding strategies used by novices and experts.

The presence of the expert VSTM advantage under conditions in which there should have been significant proactive interference in LTM provides additional evidence against the reliance of this effect on stimulus-specific representations in LTM. However, it has been suggested that memory experts, such as digit span experts, may be able to overcome the influence of proactive interference by employing one of two strategies (Ericsson & Kintsch, 1995). For example, the most recently stored item can be distinguished on the basis of its temporal context. However, it is unlikely that this temporal information would be sensitive enough to be reliable under conditions such as those in Experiment 3, where items frequently appeared in consecutive trials (a little more than a few seconds apart at times) and subjects performed 144 of such trials, with six faces per trial, within a half hour period. Alternatively, Ericsson and Kintsch (1995) suggested that experts can minimize proactive interference by generating multiple unique meaningful associations for the same chunk of information. Once again, although the digits typically used in Ericsson's studies can be easily encoded as a running time, a zip code, or a birthday, it would be considerably more difficult to implement such reliable alternative encoding strategies for differentiating unfamiliar faces at the individual level. Recall that in Experiment 3, each stimulus repeats up to 80 times; therefore, a large number of different reliable alternative encoding strategies would have to be generated to avoid proactive interference. Thus, the idea that the VSTM expert advantage relies on accessing stimulus representations in LTM finds little or no support in Experiment 3.

Some might suggest that the VSTM advantage among experts may not have been affected by a change in stimulus set because the experts had already established LTM representations for the cars in both the pre- and postchange stimulus sets. In contrast, novices—who were presumably less familiar with the cars—may have been influenced by the additional exposure to the cars in the prechange set, as it may have provided an opportunity to establish item-specific representations in LTM. However, this account would have predicted that VSTM for unfamiliar faces would have incurred a cost due to the change in stimulus set, as participants could not have had preexisting representation of these faces in LTM. In addition, this explanation would have predicted that novices' VSTM for cars, and VSTM for the unfamiliar faces across all participants, would have increased from the first to the second half of the pre-change trials, which was not the case.[7] Therefore, the use of participating, preexisting representations in LTM by experts, whether of faces or of cars, is apparently unable to account for the pattern of results found in Experiment 3. In Experiment 4, we more directly evaluated the potential role of familiarity in contributing to the VSTM advantage for objects of expertise.

Experiment 4

Although the findings from Experiment 3 provide strong evidence that the VSTM advantage for objects of expertise does not depend on stimulus-specific LTM representations acquired during the study, further evidence may be necessary before we can rule out contributions from long-term stimulus familiarity. Specifically, it is possible that experts might still be able to recruit preexisting representations in LTM acquired through their extensive real-world experience. One way to test this hypothesis would be through a comparison of VSTM performance for familiar and unfamiliar cars. However, comparisons with less familiar cars (e.g., foreign models or antique cars) are problematic, as they can entail moving outside the trained perceptual space. For example, expertise with modern cars and the associated holistic processing style adopted do not transfer to antique cars (Phillips, Grovola, Bukach, & Gauthier, 2007). Similarly, faces from an unfamiliar race are not as well recognized and are not processed as holistically as are own-race faces (Tanaka, Kiefer, & Bukach, 2004). However, previous studies that have not manipulated race have shown an inversion effect with both familiar and unfamiliar faces (Scapinello & Yarmey, 1970; Yarmey, 1971). This finding suggests that holistic processing does not depend on familiarity with an exemplar, although it may be necessary for objects to come from an area of perceptual space that is very familiar to the observer.

From a practical standpoint, this aspect of holistic processing makes the manipulation of familiarity more difficult with cars than with faces. That is, the car category is limited by the finite number of car models that have ever been manufactured, and a very experienced car expert might easily notice when presented with an image of a model that does not exist, even though it might fit in principle within the familiar perceptual space. In contrast, face experts cannot aspire to have experienced all possible faces, so there is nothing strange or particular about unfamiliar faces. In Experiment 4, we manipulated the familiarity of faces to see if we could further rule out contributions of preexisting LTM representations to the VSTM advantage observed in Experiments 1–3. Notably, the nature of the VSTM advantage for faces and that for cars among car experts is remarkably similar not only in size but in its orientation specificity and encoding time course. Therefore, the manipulation of face familiarity should have implications for effects of familiarity on VSTM more generally.

Familiar objects, whether they are faces of famous individuals or best selling car models, are typically distinguished both by the

[7] To test for an increase in performance due to experience with the images, we divided trials from the first two thirds of the experiment into two bins. A 2 (category: faces, cars) × 2 (duration: 500 ms, 4,000 ms) × 2 (block: first third, second third) × 2 (group: car expert, car novice) ANOVA found no main effect or interaction with order (all $ps > .229$).

frequency with which they are seen and by the labels or semantic information associated with them. Thus, if such information can facilitate VSTM, we might expect better VSTM for famous faces, not because they are processed more holistically but because there is a contribution from semantically related information. Although it would be reasonable to expect main effects of familiarity and of inversion (because of reduced holistic processing) on VSTM capacity, we predicted that there would not be an interaction between them. Such an interaction would be required to account for the orientation-specific VSTM advantage for cars that emerged among car experts but not among novices. That is, if familiarity is a possible basis of the orientation-specific VSTM advantage for cars among car experts, it should increase VSTM for upright stimuli more than it does for inverted stimuli. In sum, Experiment 4 allowed us to test two important questions: Does familiarity facilitate VSTM performance, and, if so, can it account for the orientation-specific VSTM advantage for objects of expertise?

Method

Participants. Thirty-one individuals (4 male; age, $M = 19.12$ years, $SD = 2.02$) from Temple University participated for course credit. All participants had normal or corrected-to-normal vision. Data from 2 participants were excluded prior to analysis due to poor performance (i.e., the VSTM capacity estimate, kappa, was equal to 0 in at least one condition).

Stimuli. We used a total of 120 images that consisted of three front-on images of each of 40 famous individuals (actors or entertainers; 20 female, 20 male). All images had a neutral facial expression. To maximize the familiarity of the faces, we chose only individuals who appear frequently in the current popular media (e.g., Tom Cruise, Tom Hanks, Julia Roberts). We cropped each image to remove hair and background, converted it to gray scale, and scaled it to equate image height. Each image was divided into three segments that separated the eye, nose, and mouth regions of the face. Forty new "familiar" images of these famous individuals were created by combining a mouth, nose, and eye segment from each of the three images of the same famous person (see Figure 5A, left column). The parts were aligned in such a way that the configural relations between the different regions of the face were as close as possible to those in the original images of the famous individuals. Forty additional "unfamiliar" face images were created by recombining the same face segments in such a way that the different pieces within any one face came from three different famous individuals (see Figure 5A, right column). This manipulation of familiarity was necessary to control the content of the images across the familiar and unfamiliar conditions and thus to eliminate spurious lower level perceptual differences that could potentially impact VSTM. Therefore, what defined familiarity was the conjunction of the different parts of faces. To confirm the validity of our stimulus manipulation, we had 21 of the participants complete a survey, in which they were asked to rate the familiarity of each face on a scale of 0 (*completely unfamiliar*) to 10 (*highest possible familiarity*), after they had completed the VSTM task. The mean familiarity ratings for the familiar (5.60) and unfamiliar (2.46) faces were significantly different, $t(20) = 7.99$, $p \leq .0001$.

Design, procedure, and analysis. The design and procedure were similar to those used in Experiments 1 and 2, except that the set size was fixed at 5 and the VSTM study array was always presented for 4,000 ms. Participants performed 10 blocks of 16 trials. Forty trials were presented in each of the four conditions (i.e., upright famous faces, inverted famous faces, upright unfamiliar faces, and inverted unfamiliar faces). Trials for each orientation were blocked, and participants performed alternating blocks that contained only upright or only inverted stimuli. Twenty famous faces (10 upright, 10 inverted) and 20 unfamiliar faces (10 upright, 10 inverted) appeared throughout the study. Stimuli were selected so that each of the face segments appeared in only one condition (i.e., in the context of an upright famous face, an inverted famous face, an upright unfamiliar face, or an inverted unfamiliar face). In addition, in the unfamiliar face condition, in which the three regions from the same face were separated across different facial images, all the segments from the same original identity always appeared within the same condition (e.g., if Nicole Kidman's eyes appeared in the inverted condition for a participant, Nicole Kidman's nose and mouth appeared, also inverted but across different, unfamiliar face images). The allocation of the stimuli to each of these conditions was counterbalanced across participants.

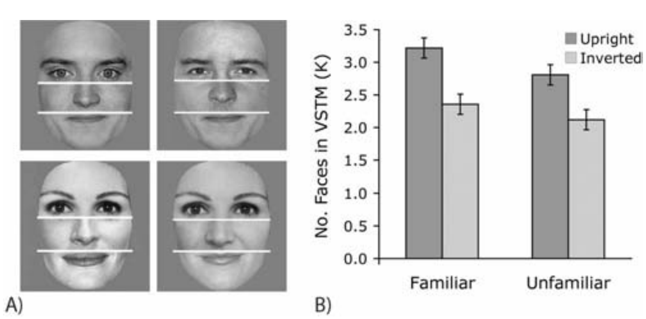

Figure 5. A: Examples of the familiar (left column) and unfamiliar (right column) faces used in Experiment 4. Note that the unfamiliar faces in the right column share a feature with the familiar faces in the left column (i.e., that of Elijah Wood and Julia Roberts). B: Visual short-term memory (VSTM) performance in Experiment 4. VSTM was greater for familiar compared with unfamiliar faces regardless of orientation, with both types of stimuli showing a similar drop in performance with inversion.

Results

A 2 (orientation: upright, inverted) × 2 (familiarity: familiar, unfamiliar) ANOVA revealed main effects of orientation, $F(1, 28) = 27.97$, $p \leq .0001$, and familiarity, $F(1, 28) = 6.63$, $p = .016$, but no interaction between these two variables, $F(1, 28) = 0.63$, $p = .433$. In sum, despite a general increase in VSTM for familiar faces over unfamiliar faces, the VSTM advantage for upright over inverted faces was similar for both familiar and unfamiliar faces (see Figure 5B).

Discussion

The fact that VSTM performance was impacted by the familiarity of face stimuli in Experiment 4 suggests that familiarity could play a role in the larger VSTM capacity for cars among car experts compared with car novices. However, arguing against a familiarity account of the VSTM advantages for objects of expertise is the generalization of this familiarity advantage to both upright and inverted stimuli. It is interesting, in hindsight, that a similar small familiarity advantage for cars among car experts (relative to car novices), irrespective of orientation, appears to have been present as a trend in the other studies reported in this article (e.g., $p = .172$ in Experiment 1; $p = .11$ in Experiment 2). Thus, although familiarity may play a general role in increasing VSTM capacity for both upright and inverted cars among car experts, it cannot account for the orientation-specific VSTM advantage for objects of expertise. Therefore, our findings in Experiment 4 suggest that familiarity on its own cannot account for the expert advantage reported in Experiments 1–3.

Previous studies that explored the effect of familiarity on the face inversion effect have reported results similar to our own. For example, Scapinello and Yarmey (1970) reported a lack of an interaction between inversion and familiarity in their study of face recognition. Their study, however, used faces that became familiar through training within the lab. To explore whether these findings extend to familiarity gained in the real world, Yarmey (1971) used faces of famous individuals in the familiar condition in a follow-up study that replicated the findings. The follow-up results were remarkably similar to those reported in Scapinello and Yarmey (1970); although Yarmey (1971) suggested that verbal labels typically associated with familiar faces may increase recognition performance more for upright than for inverted faces, no such interaction between familiarity and inversion was present in the data.

The lack of an impact of familiarity on the orientation-specific nature of the VSTM advantage for objects of expertise is consistent with findings from electrophysiological studies. The N170 electrophysiological potential has been robustly linked with the structural (holistic) encoding of faces and other objects of expertise (Gauthier et al., 2003; Rossion et al., 2000). Critically, this potential is modulated by inversion but not by familiarity (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Bentin & Deouell, 2000; Eimer, 2000; Jemel et al., 2003; Rossion et al., 1999, 2000; but see Caharel et al., 2002; Marzi & Viggiano, 2007). Thus, the failure of our familiarity manipulation to impact the orientation-specific VSTM advantage for cars among car experts is consistent with a holistic processing locus for this advantage. The electrophysiological literature also provides evidence regarding the potential locus of the small VSTM advantage for familiar over unfamiliar faces, irrespective of orientation: A later electrophysiological potential, the P250, is modulated by familiarity and shows greater amplitude for familiar than unfamiliar faces (Begleiter, Porjesz, & Wang, 1995; Pfütze, Sommer, & Schweinberger, 2002; Schweinberger, Pfutze, & Sommer, 1995; Schweinberger, Pickering, & Jentzsch, 2002; Tanaka, Curran, Porterfield, & Collins, 2006). Notably, it has been suggested that this potential is the earliest component associated with a stored perceptual representation in long-term memory (Pfütze et al., 2002; Schweinberger et al., 2002). Intriguingly, a recent finding suggests that although this later P250 component is modulated by familiarity, it is less sensitive to inversion (Marzi & Viggiano, 2007). Thus, further studies might look to the P250 for additional insight into the impact of familiarity on VSTM.

General Discussion

The VSTM advantage for cars among car experts is remarkably similar to that demonstrated for faces; this advantage requires sufficient encoding time, is orientation specific, and is similar in magnitude to the VSTM advantage for faces. These findings are consistent with a general perceptual expertise account of the VSTM advantage for faces. The fact that this advantage was not eliminated by the introduction of a verbal memory load previously demonstrated to impact verbal short-term memory performance suggests that it does not rely on verbal short-term memory. Nor was it eliminated by the use of a small stimulus set, which increased the potential for proactive interference on LTM recall, or by a surprise switch in stimulus set, which probed for any advantages due to stimulus-specific representations in LTM. Finally, a role of real-world familiarity was ruled out despite evidence that familiarity could produce a general boost to VSTM performance, irrespective of object orientation. This expert advantage is robust and does not seem to depend on the direct recruitment of additional capacity from other memory systems, such as verbal or long-term memory.

We suggest that the mechanism underlying this expert VSTM advantage likely involves holistic processing, which is common to the processing of faces and of cars among car experts. The correlation between VSTM for cars and sensitivity on an established measure of car expertise is consistent with such an account: This car expertise index is correlated with measures of holistic processing of cars, and it is correlated with the N170 electrophysiological potential (Gauthier et al., 2003), which is modulated by inversion but not by familiarity. The orientation-specific nature of this advantage is also consistent with a contribution from holistic processing mechanisms: The inversion effect for faces is thought to result from reduced access to configural information critical for holistic processing (Collishaw & Hole, 2002; Kemp, McManus, & Pigott, 1990; Leder & Bruce, 2000; Leder, Candrian, Huber, & Bruce, 2001; Murray, Yong, & Rhodes, 2000; Rhodes, Brake, & Atkinson, 1993; Searcy & Bartlett, 1996; Tanaka & Sengco, 1997; Thompson, 1980). Thus, many aspects of the VSTM advantage for objects from one's domain of expertise suggest that it may be driven by holistic encoding strategies recruited by visual experts, as in the case of expert face perception.

The robustness of holistic processing effects for nonface objects of expertise has been questioned recently (McKone, Kanwisher, & Duchaine, 2007; Robbins & McKone, 2007), so the current study provided an important test of the relationship between inversion effects and expertise. Contrary to suggestions that the inversion effect is specific to face processing, the results of Experiments 1 and 2 show robust inversion effects for cars among car experts of the same magnitude as those found for faces.[8] This was not the case among novices. Thus, the orientation specificity of car experts' VSTM advantage for cars is consistent with a holistic processing account of this advantage as well as of perceptual expertise more generally.

Holistic encoding may benefit VSTM capacity by providing a tighter binding of information in object representations. Feature-based theories of VSTM suggest that VSTM is limited by both the

[8] *T* tests comparing the size of the cost to performance due to inversion for faces and cars in Experiment 1 found a difference in the size of this cost among novices, $F(1, 17) = 5.287$, $p = .034$, but not experts ($F < 1$). This pattern was found in the data from Experiment 2: novices, $F(1, 16) = 13.848$, $p = .002$; experts, $F < 1$.

capacity of independent feature stores and the capacity of attentional mechanisms required to maintain the binding between features (Delvenne & Bruyer, 2004; Wheeler & Treisman, 2002). Holistic processing may integrate spatially separate features into the same feature unit, thereby using fewer feature slots and reducing the burden on attentional resources. Similarly, consistent with object-based theories of VSTM capacity, holistic encoding may allow experts to incorporate more features into the unified object representations suggested to serve as the units of VSTM (Scolari, Vogel, & Awh, 2007). This capacity could be especially beneficial for discriminating highly similar exemplars of complex objects, such as cars. Therefore, holistic processing may allow experts to maximize the use of an inherently limited VSTM system.

At a finer scale, VSTM is influenced by the organization of features within objects; this influence provides another avenue for holistic encoding to impact VSTM performance. VSTM for features is improved when the features come from the same part, rather than different parts, of an object (Xu, 2002). For example, color and orientation information are best encoded when they are from the same part of an object, less well encoded when they are from different parts of the same object, and least well encoded when they are from spatially separated objects (Xu, 2002). In the context of the holistic processing strategy recruited by visual experts for objects within their domain of expertise, features from what would be considered different parts of an object by novices may be encoded and represented as being from the same part by experts. Thus, this strategy would allow experts to take advantage of the part benefit for feature encoding (Xu, 2002). More specifically, the more integrated nature of representations of faces, and of cars among car experts, may underlie the VSTM advantage demonstrated for objects of expertise. Therefore, the influence of object-based hierarchical feature encoding on VSTM capacity provides another avenue whereby differences in the perceived relationship between features in holistic versus feature-based object representations may impact VSTM capacity. Future work should test these hypotheses more directly.

Recent neuroimaging studies provide further insight into the system underlying VSTM and, thus, possible loci for the expertise effect on VSTM capacity (Song & Jiang, 2006; Xu & Chun, 2006). Such studies provide evidence of dissociable roles for the different nodes in the system supporting VSTM for objects, and they report a neural correlate of the effect of complexity on VSTM capacity. A number of core areas spanning frontal, parietal, occipital, and temporal lobes have been implicated in VSTM (Desimone, 1996; Druzgal & D'Esposito, 2001, 2003; Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002; Todd & Marois, 2004). Xu and Chun found that although activity patterns in the inferior intraparietal sulcus (IPS) suggest that this node of the VSTM system has a fixed capacity of about four objects (regardless of object complexity), activity patterns in the superior IPS and the lateral occipital complex suggest that the capacity of these areas is variable, depending on the complexity of the objects stored. Thus, it appears that a complexity-induced bottleneck in the superior IPS and lateral occipital complex leads to the observed lower VSTM capacity for complex objects. Xu and Chun suggested that the inferior intraparietal sulcus is responsible for maintaining spatial attention over a fixed number of objects, whereas the superior parietal sulcus and the lateral occipital complex are important for encoding and maintaining the specific object representations.

The involvement of object form processing areas in the effect of complexity on VSTM capacity (Song & Jiang, 2006; Xu & Chun, 2006) is consistent with the suggestion that the locus of the expert VSTM effect may be in the nearby fusiform region, namely, the FFA, which has been implicated in the perceptual processing of objects of expertise (Gauthier et al., 2000, 2005; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Xu, 2005). Previous studies have implicated the FFA in VSTM for faces (Druzgal & D'Esposito, 2001, 2003). In addition, the level of activity in the FFA after expertise training with a novel category of objects is correlated with behavioral measures of holistic processing (Gauthier et. al., 1999). Thus, the recruitment of the FFA for objects of expertise and the resulting holistic processing strategy may allow experts to better encode complex visual information and may potentially reduce the perceived complexity of objects of expertise. This strategy may limit the susceptibility of expert VSTM capacity to object complexity. If true, this finding would suggest that activation in the superior IPS and/or occipital/temporal areas responsible for encoding information in VSTM reflects the perceived, rather than the physical, complexity of the objects stored in VSTM. Future studies will explore this prediction.

The findings reported in this study provide an alternative to the recent claim that the "true" capacity of VSTM, free from contamination by LTM, verbal memory, or contexual information, is limited to one object (Olsson & Poom, 2005). Olsson and Poom found that, with 500 ms of encoding time, participants had a VSTM capacity for intracategorical geometric shapes (e.g., ovals with varying aspect ratios) of only a single item. On the basis of this finding, they suggested that performance in previous studies reporting a VSTM capacity of three to four objects (e.g., Luck & Vogel, 1997; Vogel et al., 2001) was facilitated by categorical structures in LTM. Specifically, they suggested that such a benefit arises from the use of stimuli that cross category boundaries (e.g., a red and a yellow square cross a color boundary). In the studies reported here, the faces were unfamiliar, had no obvious labels, and belonged to a single category. Therefore, according to Olsson and Poom, observers should have had a capacity of only a single face under such conditions. One possible reason for this inconsistency is the limited encoding time in the Olsson and Poom study; our findings and those of Eng et al. (2005) and Curby and Gauthier (2007) suggest that VSTM capacity for complex objects is underestimated with 500 ms of encoding time because of perceptual encoding limitations. It is possible that capacity for the geometrical objects used by Olsson and Poom could reach that reported for cars among novices, for instance, if participants were given enough encoding time.

In conclusion, we provide evidence that VSTM capacity for complex objects is not hardwired and instead can be influenced by one's experience. More specifically, we suggest that the nature of representations stored in VSTM allows visual experts to maximize the storage efficiency of an otherwise inherently limited system. Thus, extensive experience with a category of objects, such as that required to produce the qualitative shift in encoding strategy seen among perceptual experts, leads to greater VSTM than would be expected on the basis of the complexity of the objects stored. Holistic processing appears to benefit VSTM capacity; it remains an empirical question as to whether other types of encoding strategies can benefit as well.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106–111.

Begleiter, H., Porjesz, B., & Wang, W. Y. (1995). Event-related brain potentials differentiate priming and recognition to familiar and unfamiliar faces. *Electroencephalography and Clinical Neurophysiology, 94,* 41–49.

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

Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology, 17*(1–3), 35–54.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research, 45*(4), 431–448.

Caharel, S., Poiroux, S., Bernard, C., Thibaut, F., LaLonda, R., & Rebai, M. (2002). ERPs associated with familiarity and degree of familiarity during face recognition. *International Journal of Neuroscience, 112,* 1531–1544.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 641–653.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81.

Chen, D., Eng, H. Y., & Jiang, Y. (2006). Visual working memory for trained and novel polygons. *Visual Cognition, 12*(1), 37–54.

Collishaw, S. M., & Hole, G. J. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception, 31*(3), 287–296.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–185.

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42–100.

Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development, 70*(5), 1082–1097.

Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin and Review, 14*(4), 620–628.

Delvenne, J.-F., & Bruyer, R. (2004). Does short-term memory store bound features? *Visual Cognition, 11*(1), 1–27.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, USA, 93,* 13494–13499.

Druzgal, T. J., & D'Esposito, M. (2001). Activity in fusiform face area modulated as a function of working memory load. *Cognitive Brain Research, 10,* 355–364.

Druzgal, T. J., & D'Esposito, M. (2003). Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *Journal of Cognitive Neuroscience, 15*(6), 771–784.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123,* 161–177.

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology, 111*(4), 694–705.

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin and Review, 12*(6), 1127–1133.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Freyhof, H., Gruber, H., & Ziegler, A. (1992). Expertise and hierarchical knowledge representation in chess. *Psychological Research, 54*(1), 32–37.

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 and Behavioral Neuroscience, 5*(2), 222–234.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6*(4), 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*(2), 191–197.

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*(6), 568–573.

Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory, 6*(3), 225–255.

Jemel, B., Schuller, A., Cheref-Kahn, Y., Goffaux, V., Crommelinck, M., & Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *NeuroReport, 14*(16), 2035–2039.

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception, 19*(4), 531–543.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology, A, 53*(2), 513–536.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception, 30*(1), 73–83.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Marsh, R. L., & Hicks, J. L. (1998). Event-based prospective memory and executive control of working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 336–349.

Marzi, T., & Viggiano, M. P. (2007). Interplay between familiarity and orientation in face processing: An ERP study. *International Journal of Psychophysiology, 65*(3), 182–192.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11*(1), 8–15.

Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *Journal of Neuroscience, 26*(43), 11187–11196.

Murray, J., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science, 11*(6), 492–496.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences, USA, 102*(24), 8776–8780.

Pessoa, L., Gutierrez, E., Bandettini, P., & Ungerleider, L. (2002). Neural correlates of visual working memory: FMRI amplitude predicts task performance. *Neuron, 35*(5), 975–987.

Pfütze, E. M., Sommer, W., & Schweinberger, S. R. (2002). Age-related slowing in face and name recognition: Evidence from event-related brain potentials. *Psychology of Aging, 17*(1), 140–160.

Phillips, W. S., Grovola, M., Bukach, C., & Gauthier, I. (2007, November).

Limits of expertise. Poster session presented at the 15th Annual Object Perception, Attention, and Memory Conference, Long Beach, CA.
Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition, 47*(1), 25–57.
Riggs, K. J., McTaggart, J., Simpson, A., & Freeman, R. P. J. (2006). Changes in the capacity of visual working memory in 5- to 10-year-olds. *Journal of Experimental Child Psychology, 95*(1), 18–26.
Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition, 103*(1), 34–79.
Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Visual short-term memory in the first year of life: Capacity and recency effects. *Developmental Psychology, 37,* 539–549.
Rossion, B., Delvenne, J.-F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., & Guerit, J.-M. (1999). Spatio-temporal localization of the face inversion effect: An event-related potentials study. *Biological Psychology, 50,* 173–189.
Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (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*(1), 69–72.
Rotshtein, P., Geng, J. J., Driver, J., & Dolan, R. J. (2007). Role of features and second-order spatial relations in face discrimination, face recognition, and individual face skills: Behavioral and functional magnetic resonance imaging data. *Journal of Cognitive Neuroscience, 19*(9), 1435–1452.
Saiki, J., & Hummel, J. E. (1998). Connectedness and part–relation integration in shape category learning. *Memory and Cognition, 26*(6), 1138–1156.
Scapinello, K., & Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science, 21*(6), 329–331.
Schweinberger, S. R., Pfutze, E. M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 722–736.
Schweinberger, S. R., Pickering, E. C., & Jentzsch, A. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research, 14*(3), 398–409.
Scolari, M., Vogel, E. K., & Awh, E. (2007). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review, 15*(1), 215–222.
Searcy, J., & Bartlett, J. C. (1996). Inversion and processing of component and spatial–relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 904–915.
Song, J. H., & Jiang, Y. (2006). Visual working memory for simple and complex features: An fMRI study. *NeuroImage, 30*(3), 963–972.
Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130,* 534–543.
Tanaka, J. W., Curran, T., Porterfield, A. L., & Collins, D. (2006). Activation of preexisting and acquired face representations: The N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience, 18*(9), 1488–1497.
Tanaka, J. W., Kiefer, M., & Bukach, C. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93,* B1–B9.
Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*(5), 583–592.
Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457–482.
Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception, 9*(4), 483–484.
Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428,* 751–754.
Tong, F., & Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1016–1035.
Troje, N., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*(12), 1761–1771.
Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114.
Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131,* 48–64.
Xu, Y. (2002). Encoding color and shape from different parts of an object in visual short-term memory. *Perception and Psychophysics, 64*(8), 1260–1280.
Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex, 15,* 1234–1242.
Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Perception and Psychophysics, 68*(5), 815–828.
Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440,* 91–95.
Yarmey, A. D. (1971). Recognition memory for familiar "public" faces: Effects of orientation and delay. *Psychonomic Science, 24*(6), 286–288.

Received April 24, 2007
Revision received February 25, 2008
Accepted March 19, 2008 ■