PERCEPTUAL EXPERTISE AFFECTS VISUAL SHORT-TERM MEMORY AND THE TIME-COURSE OF PERCEPTUAL ENCODING

By

Kim M. Curby

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in
Psychology
May, 2006
Nashville, Tennessee

Approved:
Professor Isabel Gauthier
Professor Randolph Blake
Professor Daniel Levin
Professor Gordon Logan
Professor Jeffrey Schall

ACKNOWLEDGMENTS

I am extremely grateful to each of the members of my Dissertation Committee, not only for providing invaluable guidance on the work reported in this dissertation, but also for their enthusiasm and interest in my work. I would especially like to thank my advisor and committee chair, Isabel Gauthier. I feel very lucky to have had such an inspirational group of researchers as my mentors.

I would also like to thank my fellow lab mates, Corrie Camalier, and the members of the Perceptual Expertise Network for many insightful discussions. I am also grateful to Marci Flanery and Paul Dux for their invaluable suggestions and comments on earlier versions of this dissertation.

On a personal level, I give many thanks to my family, especially my parents, for their support and encouragement. Last but not least, there was no person more important to me in the pursuit of this dissertation than Steven Most. His unfailing support and encouragement provided me with endless inspiration, both personally and academically.

TABLE OF CONTENTS

Page

Chapter

Appendix

LIST OF TABLES

LIST OF FIGURES

PREFACE

OVERVIEW OF THE DISSERTATION

This dissertation aims to augment our understanding of the determinants of visual short-term memory (VSTM) capacity for complex objects. Much of the work exploring VSTM capacity has used very simple objects, such as colored dots (Luck & Vogel, 1997; Todd & Marois, 2004) and it is unclear how well such findings extend to VSTM for more complex items, such as faces or other real-world objects such as cars. A number of studies have described the influence of stimulus-based features—such as perceptual complexity or object structure—on VSTM, but a crucial insight, as yet unexplored by this literature, is that different classes of complex stimuli tend to be processed in qualitatively different ways. In this dissertation I specifically address the influence of processing biases on visual short-term memory capacity (VSTM) for faces and objects. I build an important bridge between the VSTM literature and the vast literature demonstrating that upright faces and other objects of expertise recruit fundamentally different (holistic) processing strategies than do inverted faces or objects for which one does not possess expertise with.

This dissertation begins (CHAPTER I) with a review of our current knowledge of the determinants of VSTM capacity, specifically VSTM for objects. I then describe the current understanding of face and object processing, with an emphasis on the implications of the development of perceptual expertise and the shift to a more holistic processing strategy. Following this review of the literature, CHAPTER II reports a series of three experiments that compare VSTM for holistically processed faces and more featurally processed non-face objects and inverted faces.

After establishing a clear VSTM advantage for upright faces when sufficient encoding time was allowed in CHAPTER II, CHAPTER III reports a series of four experiments that tests whether this advantage is specific to faces or whether it extends to non-face expert categories that have also been shown to recruit a holistic processing strategy, thus providing a clear test of the proposed influence of encoding strategy on VSTM capacity. An effect of expertise on VSTM capacity for cars was demonstrated and once again this effect depended on sufficient encoding time. Additional experiments replicated the VSTM advantage for objects of expertise and also ruled out alternative accounts based on verbal working memory, long-term memory and even differences in eye-movement strategies.

CHAPTER IV reports experiments that were inspired by the dependency of the expert VSTM advantage on sufficiently long encoding times. Specifically, CHAPTER IV addresses the question as to whether holistic processing results in a greater temporal burden to perceptual processing than that for more feature-based processing strategies. Recognition performance for upright faces and inverted faces, and cars among car experts and car novices, is compared under different temporal constraints. Estimates of the rate of processing and onset of perceptual processing are compared for holistically processed items (upright faces, and cars among car experts) and less holistically processed items (inverted faces and cars, and upright cars among car experts). The results of these studies suggest that experts' experience a 'head-start' in the form of an earlier onset of perceptual performance, although there may be no difference in the rate of increase in performance with encoding time relative to that for novices.

I conclude my dissertation with CHAPTER V in which I discuss the specific implications and limitations of a role of experience and/or processing biases on our understanding of VSTM capacity and perceptual expertise. Here I also speculate about the broader implications of these findings and future avenues that could provide an empirical answer to these speculations.

CHAPTER I

VISUAL SHORT-TERM MEMORY, HOLISTIC FACE REPRESENTATIONS AND PERCEPTUAL EXPERTISE

Shelves are stocked with endless books offering advice on how to increase one's memory capacity. Their popularity is not surprising given how painfully obvious the limitations of short-term memory are in our daily lives. Even a mundane task like attaching new electronic equipment to a television or following a street map forces us to look continuously back and forth between the diagram and the task at hand, revealing the limited amount of visual information we can keep in memory at any one time. However, people who are experienced with maps, even if they have never followed a particular path before, appear to need only a few glances to extract and retain all the information needed to perform the same task. Despite its centrality to everyday life, the precise factors that constrain or increase short-term memory are still relatively unknown. Some suggest that visual short-term memory (VSTM) has a fixed number of 'slots', each capable of temporarily storing one object (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). Others argue that VSTM capacity is influenced by the complexity of the items stored (Alvarez & Cavanagh, 2004). An interesting, yet hitherto neglected, question is whether VSTM is fixed solely by stimulus factors such as perceptual complexity or object number, or if instead it can be influenced by the processing strategy used to encode an object, as suggested by an experienced map reader's efficiency at reading maps. This question is particularly important given evidence that experience can lead to a qualitative change in the perceptual processing of visual information. This chapter will first review the state of current knowledge about VSTM, factors that influence its capacity, and how learning might influence VSTM capacity. The specialized holistic processing strategy used for faces will then be discussed with an emphasis on this processing strategy as a result of the development of visual expertise more generally. This chapter will conclude by raising important questions about the influence of learning and encoding strategy on VSTM—questions that form the basis of this dissertation.

Working memory and visual short-term memory

Distinctions between working memory and short-term memory are often lost in the literature with the two terms frequently used interchangeably. The term *short-term memory* was originally introduced as a contrast to long-term memory (LTM), describing the temporary storage of information as opposed to the more permanent storage of information (James, 1890). Decades later, the term *working memory* was introduced to describe a system that encompassed not only mechanisms for the temporary storage of information, but also executive or control mechanisms for the manipulation of this information during cognitive tasks (Atkinson & Shiffrin, 1971; Baddeley, 1986; Baddeley & Hitch, 1974). Short-term memory came to be considered one component of the working memory system.

Early models of short-term memory assumed a unitary nature, with no distinction between the storage of different types or modalities of information (Atkinson & Shiffrin, 1971). However, the lack of interference between concurrent performance of visual and verbal memory tasks, in contrast to the substantial interference between concurrent performance of two verbal or visual tasks, prompted Baddeley and Hitch to propose a multi-component model of short-term memory (Baddeley, 1986; Baddeley & Hitch, 1974). This model proposed separate modality-specific stores or subsystems for verbal (phonological loop) and visual (visual-spatial sketchpad) information that were controlled by an active attentional control system referred to as the central executive. With each passing decade, more and more evidence accumulates for a distinction between verbal and visual information storage, and now few question the existence of this distinction.

This dissertation focuses on the storage of objects in the visual subsystem of short-term memory, akin to the passive visual cache proposed as an extension to Baddeley's model of working memory (Logie, 1995). This dissertation uses the term "visual short-term memory" (VSTM) to refer to a limited capacity sub-system or mechanism for the temporary representation and maintenance of visual information in the absence of external input. Similarly, VSTM *capacity* is defined as the upper limit of visual information that can be successfully maintained and retrieved from this visual subsystem. Specifically, VSTM for objects is reported in terms of an estimate of the number of objects that can be successfully maintained and later recognized from VSTM stores. It is important to note that VSTM capacity is a function of the particular characteristics of a task, for example the presentation time of the items to be encoded in VSTM can limit capacity (Eng, Chen, & Jiang, in press).

The capacity of VSTM may be overestimated in some classic VSTM paradigms, such as match-to-sample or change detection, since VSTM capacity estimates can be inflated by contributions from other storage systems such as verbal or long-term memory (Ericsson & Chase, 1982; Ericsson, Chase, & Faloon, 1980; Ericsson & Kintsch, 1995; Olsson & Poom, 2005). Following Olson, Jiang and Moore (2005), throughout this dissertation I will refer to these inflated estimates as 'functional' VSTM capacity.

VSTM can be readily distinguished from other forms of visual memory. Unlike iconic memory, which has a temporal span of approximately half a second (Phillips, 1974; Sperling, 1960), and LTM, which can have a seemingly infinite span, VSTM lasts for seconds (Pashler, 1988). In addition, in contrast with LTM or iconic memory, VSTM for objects has an extremely limited capacity with typically only 3-4 items stored at any one time (Luck & Vogel, 1997; Vogel et al., 2001).

At a neurophysiological level, VSTM is believed to be supported by a network of areas distributed across prefrontal, parietal, and temporal/occipital cortex (Smith & Jonides, 1997, 1998; Wager & Smith, 2003). Recently, the role of temporal cortex in VSTM for objects has become a topic of increasing interest. Perceptual, category-selective, areas in the temporal lobe appear necessary for *encoding* objects into VSTM (Druzgal & D'Esposito, 2001). There is a growing body of evidence suggesting that these areas also play an important role in *storing* object information in VSTM. Consistent with this proposed role of perceptual areas in VSTM storage, the face-selective temporal area (also known as the FFA) remains activated during delay periods in VSTM tasks with faces (Courtney, Ungerleider, Keil, & Haxby, 1997; Druzgal & D'Esposito, 2003; Ranganath, Cohen, Dam, & D'Esposito, 2004). FFA activation is also influenced by face memory load during both the encoding and maintenance stages in a VSTM task (Druzgal & D'Esposito, 2001; Druzgal & D'Esposito, 2003; Postle, Druzgal, & D'Esposito, 2003). Some researchers even suggest that the FFA serves as a domain-specific storage space for VSTM (Druzgal & D'Esposito, 2003; Postle et al., 2003; Ranganath, DeGutis, & D'Esposito, 2004). Similar to the FFA, the scene-selective parahippocampal area (PPA) appears to play a role in maintaining scenes in VSTM (Ranganath, DeGutis et al., 2004). Therefore, the perceptual areas recruited at encoding also appear to be involved in storing information in VSTM.

The determinants of visual short-term memory capacity

Object number and featural organization

Much research has targeted the question of what determines the capacity of VSTM. Factors such as perceptual organization can greatly increase the capacity of VSTM, such as whether visual features are attributed to the same or different objects (Luck & Vogel, 1997). For example, although observers can only reliably retain information about four colors or four orientations of simple stimuli, they can retain both the color and orientation of the same number of objects (Luck & Vogel, 1997; Vogel et al., 2001). This result extends to objects defined by conjunctions of up to four different features, thereby allowing observers to retain 16 features providing they are only distributed across four objects. These results provide support for object-based accounts of VSTM, in which VSTM capacity is determined by the perceptual organization of the information to be stored.

At a finer scale, VSTM is also influenced by the organization of features within objects; VSTM for features is improved when the features come from the same part of an object compared to when the features come from different parts of an object (Xu, 2002a). 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, 2002a). These findings provide evidence of the influence of an object-based hierarchical feature coding on VSTM capacity, suggesting that the perceived relationship between features within objects may also have consequences for VSTM.

Not only does the organization of features influence VSTM, but the dimension from which they belong does as well. Wheeler and Triesman (2002) reported that predictions of a strong object-based account of VSTM capacity break down when the features, within or across objects, are from the same dimension, such as color or shape. A strong object-based account would predict that VSTM capacity would only be sensitive to the number of objects, irrespective of the number or organization of features within the objects. Instead, they found that features from the same dimension compete for capacity, while features from different dimensions can be stored in parallel (Delvenne & Bruyer, 2004; Wheeler & Treisman, 2002). They suggest that binding in VSTM can occur for features within an object, but that this requires capacity-limited attention. Consistent with previous work, the viability of such binding depends on the perceptual organization of the features. However, more recent work suggests that features such as shape and texture can be bound together in VSTM and stored just was well as a single feature, providing the features share the same coherent boundary and thus share the same spatial location (e.g., a textured square; Delvenne & Bruyer, 2004). Therefore, feature-based accounts of VSTM suggest that capacity is determined by the perceptual organization of the features to be stored, the capacity of independent feature stores, and also the capacity of attentional mechanisms required to bind the features together into objects.

Object complexity and encoding limitations

It has recently been suggested that VSTM is not only influenced by the organization of features into objects, but also by the complexity of the resulting objects. More specifically, although VSTM capacity for objects consisting of a number of simple features can appear to be strongly object-based, VSTM for more complex stimuli does not fit with a 'strong' object-based account (Alvarez & Cavanagh, 2004; Olson & Jiang, 2002). Contrary to the object-based account, it has been shown that VSTM capacity for object categories that vary in complexity, from simple colored squares to complex line drawings of familiar objects, is highly correlated with the information load for each category (information load was operationalized as the rate with which items of a particular category could be searched; Alvarez & Cavanagh, 2004). Therefore, VSTM capacity is at least partly determined by the nature of the items being stored, with more complex objects requiring more 'space' in VSTM.

Notably, Alvarez & Cavanagh's (2004) measure of information load, visual search rate, is influenced by the same factors of object-based hierarchical encoding that have been demonstrated for VSTM. For example, visual search rate for conjunctive targets was faster for features from the same part of an object compared to features from different parts of an object (Xu, 2002b). Therefore, the information load of an object is not simply the sum of the relevant features composing that particular object, as would be suggested by a strong version of the feature-based theory of VSTM capacity. Instead, it also appears to be influenced by the perceptual organization of diagnostic features (Delvenne & Bruyer, 2004).

Recent work has revealed that the relationship between information load and VSTM capacity is particularly strong when perceptual encoding time is limited (Eng et al., in press).

Complex objects with greater information loads, such as faces, require more time to be encoded into VSTM than simpler objects like letters or colored squares. When the presentation duration of the memory array was limited (500 ms), such as it was in Alvarez and Cavanagh's (2004) study, this resulted in a disadvantage for objects with greater complexity. However, when encoding limitations were taken into consideration by providing longer presentation durations, the relationship between information load and capacity was reduced. This finding suggests that studies of VSTM capacity must also consider the role of task factors such as the presentation duration of the memory array. Such factors may create bottlenecks in perceptual encoding or consolidation that impact capacity independently of any storage or maintenance limitations in VSTM. Thus, VSTM capacity is influenced both by the complexity of the items to be stored and, under some conditions, encoding limitations.

Learning and familiarity

Most studies of VSTM have focused on the influence of task- or stimulus-based factors, such as object structure or complexity, with little regard to how learning or experience might also impact VSTM. VSTM capacity estimates of 3-4 items remain stable from early in development (12 months; Rose, Feldman, & Jankowski, 2001) to adulthood (Luck & Vogel, 1997) suggesting that VSTM capacity may be relatively inflexible and immune to the effects of learning. However, other studies from the verbal working memory literature suggest that task-specific experience and training can have a profound impact on the 'functional' capacity of short-term memory (Chase & Ericsson, 1981). For example, training can improve digit span memory from 7 items to around 80 items. This increase in short-term is strategy dependent, for example one participant in this training study reported using a strategy in which the digits were encoded as running times. This participant's superior short-term memory capacity was so strongly reliant on this strategy that when the digit list included groups of digits that did not make sense as running times (e.g. "498", or "4 minutes and 98 seconds") his performance on such lists approached his pre-training digit span (Chase & Ericsson, 1981). Therefore, task-specific experience can impact the *functional* capacity of VSTM.

There are a number of ways in which learning and experience might impact VSTM capacity. Most intuitively, learning may establish LTM traces that can supplement VSTM. In this case, learning would increase 'functional' estimates of VSTM capacity but not the actual capacity of the VSTM system. Indeed, a recent study exploring the influence of domain-specific training on VSTM for novel objects reported that participants who viewed a set of eight random polygons 160 times in the context of a VSTM task were no more accurate detecting a change in a VSTM array containing these trained shapes compared to unfamiliar ones (Chen, Eng, & Jiang, in press). Importantly, participants in this study could accurately identify the trained polygons in a two-alternative-forced-choice task, confirming that they did in fact have representations of these items in LTM. However, VSTM performance improved for both trained and untrained polygons, suggesting there is a more general practice effect on VSTM performance rather than a direct influence of LTM. Consistent with this finding, other studies have only found evidence of, at best, a limited role of experience in influencing VSTM capacity (Olson & Jiang, in press).

The inconsistency between the dramatic increase in verbal memory with training (Chase & Ericsson, 1981) and the failure to find an increase in VSTM with training (Chen et al., in press) could be interpreted as a fundamental difference between verbal and visual short-term memory systems. However, it might also reflect the different levels of training between the two studies; 50 hours or more versus less than an hour, respectively. Limited laboratory-based training cannot compare with the degree of experience we have for many object categories in our daily environment. Therefore, conclusions drawn by using such brief laboratory-based training likely underestimate the potential for experience to impact VSTM.

Consistent with the idea that extensive training is required before learning will impact VSTM, students who had never played chess before also required approximately 50 hours of memory practice to demonstrate superior memory for meaningful chess configurations (Ericsson & Oliver, 1989). The absence of a VSTM advantage for unfamiliar chess configurations among real-word chess experts suggests that this advantage relies on stored representations in long-term memory, rather than a more qualitative change to the way information is stored in VSTM (Chase & Simon, 1973). Further support for this suggestion is the failure of intervening short-term memory tasks during the retention interval to impact memory performance for familiar chess positions (Charness, 1976). Practice appears to increase chess experts' VSTM by "chunking" information into larger units in long-term memory and storing pointers to these chunks in VSTM (Chase & Simon, 1973; Freyhof, Gruber, & Ziegler, 1992; Gobet & Simon, 1998). These "chunks" have arbitrary sub-parts and properties, can be used as a processing unit, and—importantly—can be retrieved by a single act of recognition (Gobet & Simon, 1998). These findings are consistent with reports that participants who demonstrated dramatic increases in verbal memory relied on chunking digits together and matching them to existing chunks in LTM, such as semantically meaningful dates or running times (Chase & Ericsson, 1981; Ericsson & Chase, 1982).

Recent work suggests another possibility for how learning might impact VSTM; knowledge of category structures increases our VSTM capacity for items that cross category boundaries (Olsson & Poom, 2005). For instance, different colors and shapes that cross category boundaries are commonly used in VSTM paradigms (e.g., Todd & Marois, 2004; Vogel & Machizawa, 2004; Vogel et al., 2001). Olsson and Poom (2005) have suggested that previous VSTM estimates from such studies are inflated due to support from LTM in the form of this knowledge of category differences. In addition, since items separated by category boundaries are easily labeled (e.g. ‘red’ or ‘blue’) their recall may also be facilitated by contributions from verbal short-term memory. Notably, when these factors were controlled, VSTM capacity for intra-categorical novel items made of geometrical shapes was found to be limited to one object. Therefore, learning can impact the functional capacity of VSTM in a number of different ways, allowing it to not only utilize existing representations and categorical structures in LTM, but also to recode familiar items verbally and thus recruit verbal short-term memory stores.

Although the above-mentioned studies demonstrate a benefit of learning, the advantages they report appear to reflect the utilization of additional resources, such as LTM or verbal memory, to supplement VSTM rather than a true change to VSTM capacity. However, the impact of the perceptual organization of information on VSTM capacity (Delvenne & Bruyer, 2004; Luck & Vogel, 1997; Vogel et al., 2001; Xu, 2002a) may provide a potential avenue for extensive learning to more directly influence VSTM capacity. One possibility is that experience may impact VSTM by influencing what and how information is stored rather than simply allowing for the recruitment of additional capacity from other systems. For example, a single fixation (200 ms) is sufficient for highly experienced radiologists to detect and identify, with 70% accuracy, major pathological features in x-ray images (Kundel & Nodine, 1975) suggesting that experience can impact the relative weighting of different perceptual features. In addition, expert radiologists' have poorer memory for briefly presented (500 ms) normal x-ray images, but better memory for abnormal x-rays, relative to radiology residents or novices (Myles-Worsley, Johnston, & Simons, 1988). Therefore, experience and the development of visual expertise have been demonstrated to have important consequences for both perception of and memory for items from one’s domain of expertise.

Visual search studies with upright and inverted faces or digits also provide indirect evidence for the hypothesis that extensive experience impacts VSTM more directly; despite equivalent perceptual complexity, search rate is faster for upright (familiar orientation) than inverted faces (unfamiliar orientation) (Tong & Nakayama, 1999) and for upright compared to rotated digits (Wang, Cavanagh, & Green, 1994). Most importantly, in the case of visual search for faces, this reduction in information load (indexed via search rate) was found with both familiar and unfamiliar faces, suggesting that the lower information load for highly familiar categories is not merely a result of pre-existing representations in LTM. Rather, it seems to result from a change in the way the category as a whole is processed. Interestingly, despite indirect evidence that VSTM may be greater for highly familiar categories, such as upright faces, the influence of this change in processing strategy on VSTM capacity is unknown. It is possible that experienced-based changes in the way information is processed translate to changes in VSTM capacity.

Interim summary and implications

It has been demonstrated that VSTM capacity is influenced by the perceptual organization of the information to be stored. Both the organization into discrete objects (Luck & Vogel, 1997) and the organization of features within objects (Xu, 2002a) impact VSTM. VSTM is also influenced by the complexity or information load of an object, with more complex objects requiring more ‘space’ in VSTM. Although a number of studies have failed to detect a reliable influence of LTM on VSTM capacity as would be expected from studies of memory expertise, there is indirect evidence that experience can influence the information load of an object category. One possibility is that experience and learning impact VSTM capacity by influencing the encoding and perceptual organization of visual information within a highly familiar category such as that observed with perceptual expertise. This leaves open the possibility that VSTM is influenced by an individual’s domain-specific processing biases that develop through extensive experience, as is the case for visual experts. The next section will review experience-based changes in performance within the context of face and object recognition.

Comparing face and object recognition

We have more specialized and extensive experience with faces than with any other object category, and this appears to have important consequences for the perceptual organization and processing of information in faces. For example, anecdotally, relative to the processing of non-face objects, face processing appears unique in its relative immunity to irrelevant featural changes and occlusion, such as a change in expression or the addition of a hat. We also have remarkable memory for faces; adults can recognize familiar faces with accuracy greater than 90%, despite not having seen some of these faces for over 35 years (Bahrick, Bahrick, & Wittlinger, 1975). Face perception is also characterized by its extreme sensitivity to subtle changes in the spatial-relations between features (Bruce, Doyle, Dench, & Burton, 1991; Haig, 1984; Hosie, Ellis, & Haig, 1988; Kemp, McManus, & Pigott, 1990). For example, adults can detect subtle changes in the configuration of facial features as small as one minute of visual angle, which approaches the limits of normal visual acuity (Haig, 1984). Such findings have inspired much research on face perception and how it differs from typical object perception.

Much of what we know about face recognition and how it differs from that of other object categories began with exhaustive studies on the *face inversion effect*. This effect was first reported by Yin (1969) and refers to the disproportionately large cost of inversion to face processing compared to that for other object categories, such as houses, planes, or stick figure people. Yin (1969) suggested that the cost to recognition due to inversion has two sources: (1) a general cost of inversion for mono-oriented[1] objects, and (2) a 'special' face-specific cost. Yin's subjects reported employing a global (or *holistic*) strategy for upright faces and a more analytic (or *feature-based*) strategy for the other categories, providing some clue as to what makes face processing 'special' compared to the processing of other object categories.

Holistic face perception

Research from the next three and half decades proved Yin's participants to be remarkably perceptive; not only is the analytic processing strategy reported for non-face objects consistent with many later theories of how non-face objects are represented (Biederman, 1987; Jolicoeur, 1990; Marr, 1982; Marr & Nishihara, 1978), but their report of a 'holistic' processing strategy

---

[1] Mono-oriented objects are objects that are predominantly seen in only one orientation (Yin, 1969)

for faces is also consistent with many of the hallmarks of face perception that are now known to distinguish it from the perception of other object categories (Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Young, Hellawell, & Hay, 1987). Consistent with the reports of Yin's subjects, Tanaka & Farah (1993) suggested that differences between face and object processing stem from a qualitatively different holistic representation for faces compared to a more piecemeal representation for objects. In holistic representations the information about the individual features (*featural information*) and the relations between features (*configural information)* are relatively inseparable (Farah, Wilson, Drain, & Tanaka, 1998; Tanaka & Sengco, 1997). Consistent with this account, observers demonstrated superior performance identifying a feature of a face (originally learned in the context of the whole face) when it was embedded within the context of a surrounding—but irrelevant— face compared to when it was presented in isolation (Tanaka & Farah, 1993). Importantly, this advantage was not present for non-face objects, such as houses, suggesting that features in non-face objects are processed more independently than those in faces. In addition, this advantage was also absent for inverted faces, suggesting that to elicit this holistic processing advantage faces must be configurally intact. However, the effect of face context remains even for face parts learned in isolation, although in this case the context *interferes* with, rather then facilitates, the identification of the parts (Leder & Carbon, 2005). Therefore, the nature of the impact of the face context depends on how the part is represented in memory. Nonetheless, these studies provide further evidence for the obligatory holistic processing strategy recruited for configurally intact faces.

The obligatory influence of task-irrelevant features on face part judgments has also been demonstrated in a number of other paradigms (Gauthier, Curran, Curby, & Collins, 2003; Young et al., 1987). For example, individuals were more accurate at identifying a part of a famous face when it was presented in isolation compared to when it appeared as part of a chimeric face made from halves of different famous people (composite effect; Young et al., 1987). Interestingly, when the halves were misaligned participants experienced a release from the interference of the to-be-ignored part (configural effect; Young et al., 1987). The two parts appeared to fuse together when aligned, creating a new identity and making it difficult to identify the two parts independently. Consistent with the absence of a whole-part advantage for inverted faces, alignment had no effect on judgments about inverted chimeric face halves. Similarly, when participants were asked to make a same/different judgment about parts of unfamiliar chimeric faces (e.g., the bottom halves) their performance was influenced by whether the same/different relation of the to-be-ignored top parts was consistent or inconsistent with the same/different relation between the task-relevant bottom halves (Gauthier et al., 2003). For example, if the bottom halves differed across the two faces in question, performance was better if the (irrelevant) top halves also differed. Once again, if the configuration of the faces was disrupted (e.g., by inverting the irrelevant half) participants demonstrated a release from this interference, suggesting that this integrated processing also depended on the faces being configurally intact (Gauthier et al., 2003).

Face processing effects such as the whole-part advantage, composite effect, and configural effect, led many to ask what makes configurally intact faces so special? What is lost when the face configuration is disrupted through manipulations such as inversion? Conclusions drawn in response to this question are consistent across most of the studies that have explored this issue: inversion selectively interferes with the processing of configural information in faces (Bartlett & Searcy, 1993; Bruce et al., 1991; Collishaw & Hole, 2002; Kemp et al., 1990; Leder & Bruce, 1998, 2000; Leder, Candrian, Huber, & Bruce, 2001; Murray, Yong, & Rhodes, 2000; Rhodes, Brake, & Atkinson, 1993; Searcy & Bartlett, 1996; Tanaka & Sengco, 1997; Thompson, 1980). For example, observers were far more sensitive to configural changes in upright faces than in inverted or non-face objects, but inversion did not influence the ability to detect featural changes (Tanaka & Sengco, 1997). The most striking example of this lack of sensitivity to configural changes in inverted faces is the Thatcher Illusion (Thompson, 1980). This illusion refers to observers' failure to notice that the eyes and mouth in a face are upright when the surrounding face is inverted; when the face is upright, however, (with eyes and mouth now up-side down) this change in the configuration of the features has a striking impact, rendering the face image grotesque. Therefore, inversion disproportionately impairs an observer's sensitivity to the configural information within faces, which appears to be a crucial element for the establishment of holistic face representations.

Holistic processing and expertise: Face processing as a model of visual expertise

Diamond & Carey (1986) suggested that faces are not inherently special, but rather that they acquire their uniqueness as a product of our extensive domain-specific experience with them. They suggested that what makes face processing unique is that it involves the encoding of configural information relative to the prototypical configuration of features within a face. The development of this prototypical representation is a by-product of our extensive experience and resulting expertise with faces in combination with the importance of configural information for distinguishing exemplars in highly homogeneous categories. The suggestion that faces may acquire their uniqueness through experience is consistent with other work showing that face recognition in young children is more susceptible to irrelevant featural changes, less sensitive to configural changes, and is marked by an interaction between inversion and age, with the magnitude of the inversion effect increasing through development (Carey, 1981; Carey, Diamond, & Woods, 1980; Diamond & Carey, 1977). Further support for this expertise-hypothesis of face processing is provided by the demonstration of a greater inversion effect for judgments about dogs among dog experts, compared to dog novices making the same judgment (Diamond & Carey, 1986). These results suggest that the ‘unique’ nature of face processing develops over many years of practice and that similar processing strategies can develop for non-face categories after extensive experience.

Recent evidence has generally been consistent with Diamond & Carey’s suggestion that the holistic processing mechanism that supports face recognition also supports expert visual recognition more generally (Diamond & Carey, 1986). For example, holistic processing effects such as sensitivity to configural changes can be obtained for non-face objects in observers trained to become experts with a novel category (Gauthier & Tarr, 2002). In addition, real-world car experts also have more difficulty processing only a part of a car when it is presented in the context of a whole car (Gauthier et al., 2003). These results support Diamond and Carey’s suggestion that holistic processing is associated generally with expert visual processing rather than with any single stimulus domain, such as faces.

Neurophysiological evidence is also consistent with notions of face perception as a model of expert processing. At a neurophysiological level, the robustly face-selective fusiform area (FFA) in the temporal lobe is believed to be a critical node in the system responsible for face perception (Kanwisher, McDermott, & Chun, 1997). After expertise training with a novel stimulus-category (“greebles”) fMRI has revealed that participants also recruited the FFA for the trained stimulus class (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). This FFA activity correlated with the degree of behaviorally induced holistic processing of the greeble stimuli (Gauthier & Tarr, 2002). In addition, FFA activity to non-face objects such as birds or cars was correlated with observers' real-world expertise (Gauthier, Skudlarski, Gore, & Anderson, 2000; Xu, 2005). Electrophysiological studies recording event-related brain potentials (ERPs) also reveal expertise effects in the brain as early as the first face-selective response, referred to as the N170, which occurs around 170 ms following stimulus presentation (Gauthier et al., 2003; Tanaka & Curran, 2001). These results suggest that the development of visual expertise with a particular category results in a shift towards the recruitment of a cognitive processing strategy and neural substrate similar to face processing.

Further evidence for a functional overlap between face and non-face object expertise has been provided by the demonstration of interference between face and car processing in car experts (Curby & Gauthier, 2001; Gauthier et al., 2003). The level of interference measured between the processing of these categories depended on the degree to which the car task recruited a holistic processing strategy. More specifically, this level of interference depended both on one's visual expertise with cars and also whether the cars were configurally intact. In an interleaved 2-back VSTM task that required car experts to process a face and a car in an overlapping manner, the degree of holistic processing of the faces was influenced by the format of the cars (intact or inverted tops). Faces in the context of normally configured cars were processed less holistically than those presented in the context of cars in a modified configuration (tops inverted). Importantly, cars in the modified configuration were processed less holistically, and therefore presumably competed less for holistic processing resources. Therefore, there appears to be some functional overlap for face and car processing that is related to an individual's level of expertise with cars.

Functional overlap between face and expert object processing was also demonstrated in lab-trained greeble experts (Rossion, Kung, & Tarr, 2004). In a flanker paradigm, subjects attended to a centrally presented greeble and reported on which side of the screen a distractor face appeared. After expertise training with greebles, the N170 in response to the distractor face was reduced when participants were concurrently processing a greeble compared to a novel (untrained) object. This study provided further evidence that the concurrent processing of non-face objects in a domain of expertise interferes with early face-selective responses, suggesting a functional overlap between face and non-face expert processing.

Summary and Implications

Configurally intact face processing and expert non-face object processing both show similar processing effects such as sensitivity to inversion and obligatory processing of the whole object even when only certain parts are task relevant. It has been suggested that these processing effects reflect qualitatively different representations created by expert and novice observers, in which the visual information is organized in a holistic rather than in a feature-based manner, respectively (Farah et al., 1998; Tanaka & Sengco, 1997). The recruitment of overlapping neural substrates, linked with behavioral measures of holistic processing, for faces and objects of expertise provides further support for the hypothesis that at least part of the ‘specialness’ of faces is a product of our extensive experience with this category. The role of category-selective temporal lobe brain areas in VSTM functions, briefly outlined earlier in this chapter, suggests an interesting possibility: if working memory functions are to some extent supported by posterior areas, such as the FFA, which demonstrate effects of expertise, then expertise might also influence VSTM functions.

The major goal of this dissertation is to answer a number of specific questions. First CHAPTER 2 asks whether there is a VSTM advantage for holistically encoded faces compared to more featurally encoded objects or inverted faces. After evidence for such an advantage was found, CHAPTER 3 asks if this advantage is face-specific, or if it extends to holistically processed non-face objects within a participant’s domain of expertise. This chapter also addresses whether the basis of this advantage is perceptual, or if instead it can be explained by contributions from verbal or long-term memory. The expert VSTM advantage was found to depend on sufficient encoding time, raising interesting questions about the relative time-course and capacity of encoding mechanisms among expert and novice observers. To explore some of these questions, CHAPTER 4 compares the time-course of processing for upright and inverted faces and cars among car experts and novices. Specifically, this chapter addresses the question as to whether the qualitative difference between expert and novice encoding mechanisms results in a quantitative or qualitative difference in the time-course of processing for object of expertise.

CHAPTER II

COMPARING VISUAL SHORT-TERM MEMORY FOR UPRIGHT AND INVERTED FACES AND OBJECTS

Introduction

An insight gleaned from the face and object recognition literature, and yet to be considered in the VSTM literature, is that different classes of stimuli can be processed in different ways. Can the processing strategy used to encode an object influence VSTM capacity for a given category? As noted in CHAPTER I, upright faces tend to be processed more holistically than non-face objects or inverted faces, which are both processed in a more feature-based manner (Tanaka & Farah, 1991, 1993; Yin, 1969). To review, this has been demonstrated through greater sensitivity to configural changes in upright faces compared to upright objects or inverted faces and also the obligatory processing of the whole (upright) face in part judgment tasks (Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Young et al., 1987). Although it has never been tested directly, some evidence raises the possibility that, despite their equivalent perceptual complexity, VSTM capacity for upright and inverted faces may differ. For example, search rate, used by Alvarez and Cavanagh (2004) as a measure of information load found to be highly correlated with VSTM capacity, is faster for upright than inverted faces (Tong & Nakayama, 1999). It may be that the processing strategy recruited for upright faces serves to reduce information load, which in turn leads to differences in encoding and/or VSTM capacity for upright faces versus that for inverted faces or other complex objects. Notably, like search rates for upright and inverted faces, search rates are also faster for familiar compared to unfamiliar objects (Mruczek & Sheinberg, in press; Schneider & Shiffrin, 1977; Wang et al., 1994). The relationship between information load (search rate) and VSTM capacity has been shown to be strongest when encoding time is limited, being reduced when encoding duration is extended (Eng et al., in press). Therefore, VSTM capacity may not be fixed by the perceptual features of a stimulus; experience and processing biases might influence it as well. In addition, the influence of experience on information load, and thus VSTM, may interact with encoding duration.

This chapter reports on three experiments that compare VSTM capacity for upright and inverted faces and different categories of non-face objects. Experiment 1 compares VSTM for upright and inverted faces under different temporal encoding limitations. To anticipate the results, an advantage for upright faces over inverted faces at long, but not short encoding times, was found. Experiment 2 then compares VSTM for faces and non-face categories under the same encoding limitations. Once again upright faces benefit more from additional encoding time, but this time compared to non-face upright objects rather than inverted faces. This finding suggests that the VSTM advantage for faces in Experiment 1 is not merely due to a general cost of inversion. However, VSTM performance for faces does not exceed that for the non-face objects at the longest encoding time (2500 ms). Experiment 3 compares VSTM capacity for faces and non-face objects under even less limiting encoding conditions, and finds an advantage for faces over non-face upright objects.

Experiment 1

Given the relatively high informational load of unfamiliar faces and the fact that they are intra-categorical items with no labels in long-term memory, VSTM capacity for this class of stimuli would be expected to be especially small (Alvarez & Cavanagh, 2004; Eng et al., in press; Olsson & Poom, 2005). One limitation to our understanding of VSTM for faces is that studies have not compared faces with objects of similar complexity—for instance, shaded photographs of other objects from within the same category. Therefore, it is difficult to separate the influence of perceptual complexity on VSTM capacity from more obligatory face-specific differences in encoding strategy. Inverted faces may be especially useful in this regard as they are particularly well-matched to upright faces in image complexity. Yet, because visual search rate is faster for upright than inverted faces (Tong & Nakayama, 1999), suggesting a lower information load, one might expect a VSTM advantage for upright faces at short encoding durations. When the presentation duration of the memory array is extended allowing for additional time to encode the items, prior findings (e.g., Eng et al., in press) suggest that the difference between upright and inverted faces should be reduced or abolished. However, holistic processing may confer an advantage to upright faces, leading them to be encoded more efficiently than inverted faces even at long encoding times. Experiment 1 examines this possibility by comparing VSTM capacity for upright and inverted faces while manipulating the time allowed for encoding the items in VSTM.

Methods

Participants. Participants were 24 (18 female) employees, undergraduate students or graduate students of Vanderbilt University with normal or corrected to normal vision, ranging in age from 19 to 33 (M = 24.3, SD = 3.86). Participants were compensated $10 per hour for their time.

Stimuli. Stimuli were 72 grayscale faces (1.9° x 2.3°) from the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany.

Procedure. The task was presented using a 19" monitor (1024 x 768 pixel resolution, 85 Hz refresh rate) driven by Matlab (Mathworks, Inc) together with the Psychophysics toolbox (Brainard, 1997) on a 450 or 733 MHz G4 Macintosh computer. Normal overhead lighting illuminated the testing room. Participants sat at an unconstrained distance of approximately 70 cm. For each participant, a random selection of half the faces appeared in the upright trials while the remaining faces appeared in the inverted trials. Participants performed a delayed match-to-sample probe recognition task (see Figure 1): the memory array, consisting of 1-5 faces either all upright or all inverted evenly spaced in a circular array (6.1° visual angle diameter), appeared for either 500, 1200, or 2500 ms, and after a 1200 ms delay participants were presented with a face-probe in one of the locations from the memory array. The probe remained on the screen until participants pressed a key to indicate whether the face was the same as or different from the one that appeared in that location in the memory array. Participants were instructed to respond as accurately and as quickly as they could. Their response and the time to respond were recorded. To minimize confusion, within each trial, the probe was never a face that had appeared at a different location in the memory array. However, participants were not informed of this. Participants also performed an articulatory suppression task to prevent verbal rehearsal: at the beginning of a trial, before the study array appeared, two digits were presented auditorily, to be rehearsed out loud throughout the trial. After making a response to the visual probe, a screen with two digits appeared and participants made a key press to indicate whether they were the same as those rehearsed.

Participants performed 900 trials across three sessions, each consisting of 10 alternating blocks (30 randomized trials/block) of either upright or inverted faces. In sum, across the experiment, there were 450 trials with upright faces and 450 with inverted faces. Within these stimulus categories, there were 15 conditions (5 set sizes x 3 presentation durations), each of which included 30 trials.

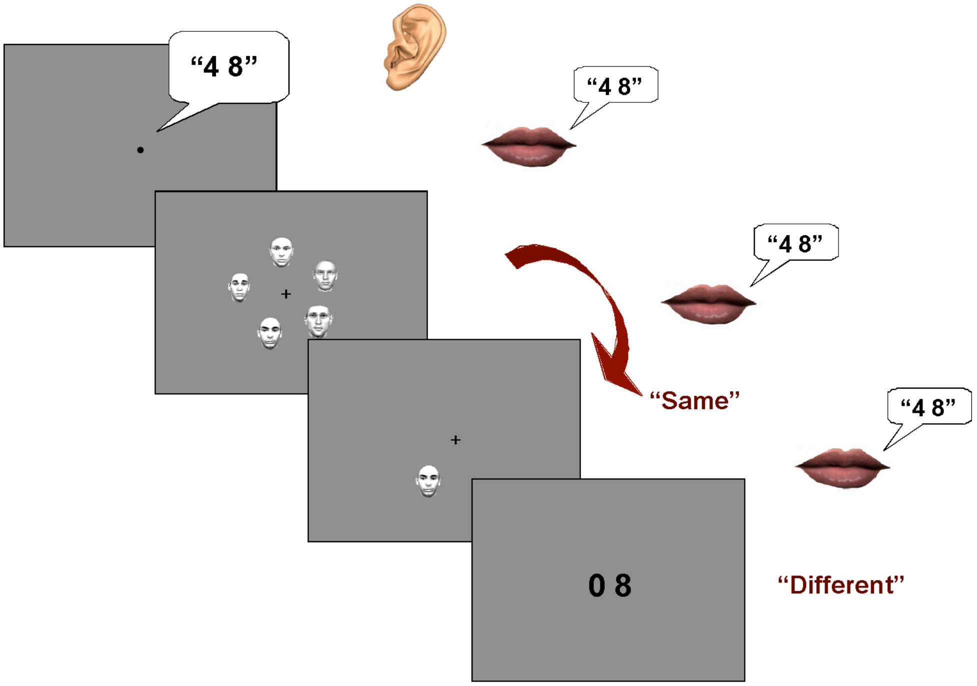


Figure 1. The sequence of events that occurred in each trial in Experiment 1. Participants first were presented with two digits followed by a visual stimulus that consisted of a screen that had 1, 2, 3, 4, or 5 upright or inverted faces in a circular array. This memory array appeared for either 500 ms, 1500 ms, or 2500 ms. After a 1200 ms delay a single face, serving as a probe, appeared in one of the locations filled in the memory array. Participants responded using the keyboard whether or not the probe face was the same as the one that appeared in this location in the memory array. After a response was made, a screen with 2 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.

Analysis. Data from one participant were removed from the analysis due to poor performance in the auditory task. Accuracy for this component exceeded 95% for all other participants. Incorrect articulatory suppression trials were removed from further analyses. For each participant, hit rate and correct rejection rate were calculated for each condition. These values were used to derive a measure of VSTM performance using Cowan's K, an estimate of the number of objects successfully encoded in VSTM (Cowan, 2001).[2] The maximum K (K-max) was identified for each presentation duration, regardless of set size, to provide an estimate of individuals VSTM capacity under these task conditions. All subsequent analyses were performed on these K or K-max values.

Results

As shown in Figure 2, increased encoding time led to increased VSTM capacity for both upright and inverted faces, although this effect was greater for upright faces. With 500 ms encoding time, orientation had little effect on VSTM capacity estimates, but longer encoding times (1500 ms or 2500 ms) led to greater VSTM capacity for upright than for inverted faces. Inverted faces benefited less than upright faces from additional encoding time. These results were not explained by a difference in response times between upright and inverted faces (see APPENDIX A).

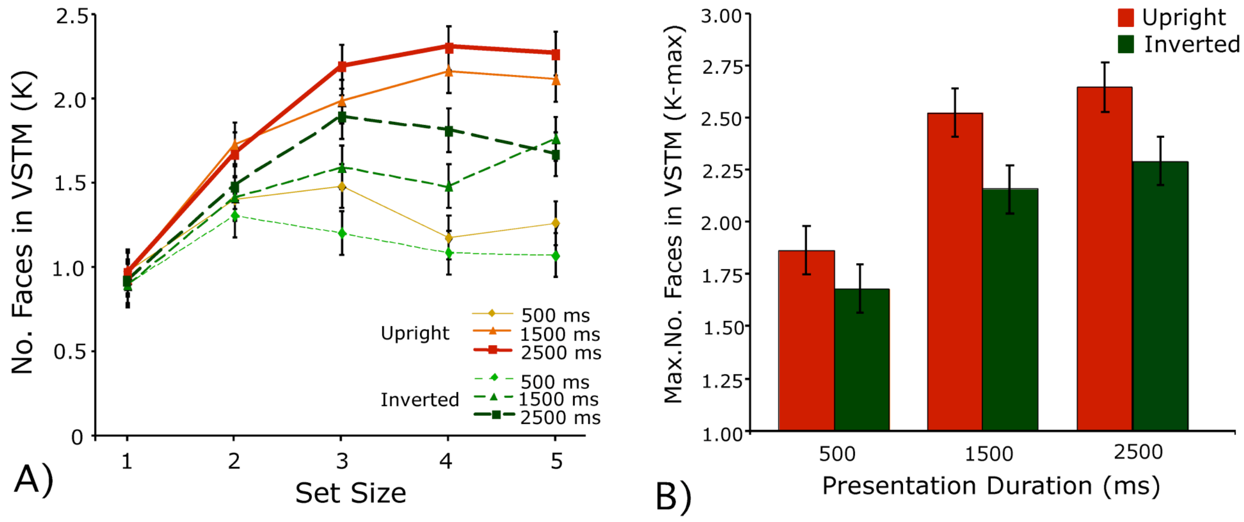


Figure 2. (A) Estimated number of upright or inverted faces stored in visual short-term memory (VSTM) using Cowan's (2001) formula at each of the different set sizes when the memory array was presented for 500 ms, 1500 ms, or 2500 ms, and (B) the maximum number of items in VSTM for each orientation and presentation duration. VSTM for upright and inverted faces were similar at the shortest presentation duration, and both increased with longer presentation durations. However, VSTM for upright faces were greater than that for inverted faces when ample encoding time was allowed, suggesting that affects of processing strategy on VSTM capacity emerge with sufficient encoding time. Error bars represent pooled standard error values.

---

[2] K=(hit rate + correct rejection rate - 1)*set size

A 2 (upright vs. inverted faces) x 5 (set sizes 1-5) x 3 (500 ms vs. 1500 ms vs. 2500 ms encoding time) ANOVA revealed a main effect of face orientation, with the overall K for inverted faces being lower than that for upright faces, $F(1,23)=29.38$, $p \leq .0001$. In addition, there was a main effect of presentation duration, with longer encoding times leading to greater VSTM capacity, $F(2,46)=83.43$, $p \leq .0001$. Most importantly, there was an interaction between orientation and presentation duration with the effect of orientation greater for longer presentation durations ($F(2,46)=4.35$, $p=0.019$; see Figure 2 A).[3]

An ANOVA on the maximum VSTM capacity (K-max) for upright and inverted faces across the different set sizes produced similar results, with a main effect of orientation. $F(1,23)=15.26$, $p=.0007$, and duration, $F(2,46)=30.76$, $p \leq .0001$. Although the interaction between orientation and duration was not significant ($F<1$, ns.), paired t-tests revealed a significant difference in K-max for upright and inverted faces in the 1500 ms, $t(23)=3.095$, $p=.0051$, and 2500 ms, $t(23)=2.59$, $p=0.016$, encoding conditions, but not the 500 ms presentation condition ($t(23)=1.66$, $p=.11$; see Figure 2 B).

Discussion

Despite equivalent perceptual complexity, VSTM for upright faces was greater than that for inverted faces. The lower VSTM capacity for inverted faces compared to upright faces at the longer encoding durations is consistent with the drop in face recognition performance reported in a wide variety of perceptual tasks when faces are inverted (Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Tong & Nakayama, 1999; Yin, 1969). The advantage for upright faces may result from holistic processing mechanisms not implemented for inverted faces. In addition, the proposed effect of these orientation-specific processing mechanisms increased with additional encoding time.

One possibility is that the VSTM capacity for inverted faces would be equivalent to that of upright faces if encoding time was extended beyond 2500 ms. That is, the absolute VSTM capacity for upright and inverted faces may be equivalent, but inverted faces are processed more

---

[3] Additionally, the main effect of set size, and its interactions with encoding duration and orientation were significant ($p<.05$). These effects are not central to our research question.

slowly. However, this is unlikely, as the maximum VSTM capacity for inverted faces failed to increase when the presentation duration was extend from 1500- to 2500 ms[4] (see Figure 2b).

An alternative interpretation of these results is that the orientation effect in VSTM capacity for faces may not be due to processing strategy, but may instead reflect other differences between upright and inverted stimuli more generally (e.g., additional costs incurred when processing objects in unfamiliar orientations; Lawson & Jolicoeur, 1998). However, this account would not predict that these effects would extend to a comparison between upright faces and other upright objects that are not processed in a holistic manner. This prediction is addressed in Experiment 2.

Experiment 2

Experiment 2 compares VSTM capacity for upright faces, cars, and watches. As discussed in CHAPTER I, previous studies suggest that these objects should be processed less holistically than faces. Cars, like faces, are typically seen in an upright orientation and thus provide a comparison to test if the results of Experiment 1 are merely due to the mono-oriented nature of our experience with faces (Yin, 1969). Watches were chosen because they are highly homogeneous and subjects would typically not have had much experience in discriminating among them. If the effects found in Experiment 1 were attributable to a more general cost due to inversion rather than to different processing mechanisms, then similar results should *not* be found when comparing upright faces to upright complex non-face objects. However, the opposite pattern of results is predicted because faces and non-face objects are thought to recruit different processing strategies, providing an advantage for upright faces over the more featurally processed non-face object categories. Experiment 2 tests this prediction by comparing VSTM capacity for upright faces to that for upright cars and upright watches while once again manipulating the time allowed for encoding the items in VSTM.

---

[4] $\underline{t}(23)=0.82, \underline{p}>.05$.

Methods

Participants. Participants were 21 (11 female) employees, undergraduate students, or graduate students of Vanderbilt University with normal or corrected to normal vision, ranging in age from 19 to 39 (M = 26.3, SD = 5.2). Participants were compensated $10 per hour for their time.

Stimuli. The stimuli were 216 grayscale images of upright human faces (1.9° x 2.3°), watch faces (1.9° x 2.3°), and cars (2.3° x 1.5°) (72 of each; see Figure 3). The faces were the same as those used in Experiment 1. The watch images all depicted front-on upright views. The viewpoint of the cars varied from three-quarter to side views across images.

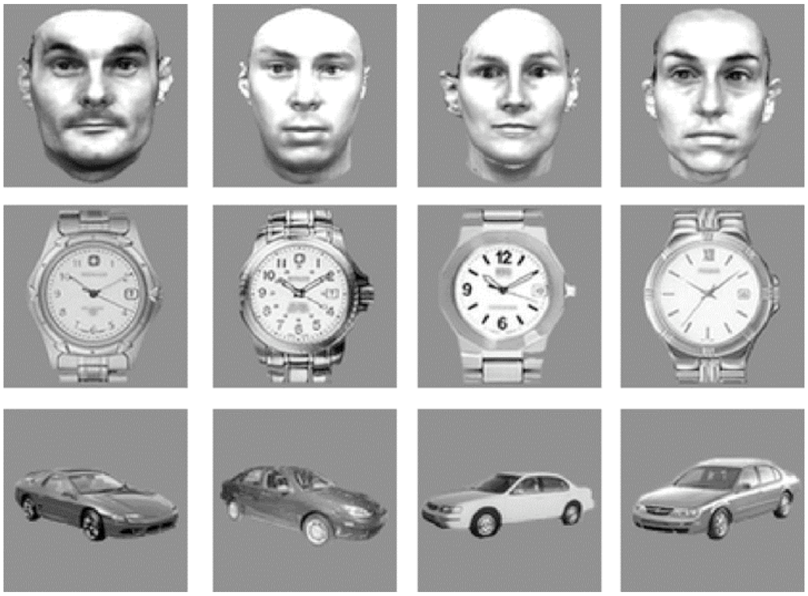


Figure 3. Examples of the stimuli presented in Experiment 2. The same faces (in either an upright or inverted orientation were also used in Experiment 1.

Procedure. The general procedure was the same as Experiment 1 except that participants performed 420 trials for each stimulus category (faces, cars, watches), in three separate sessions. The order of the sessions was counterbalanced across subjects. For each category there were 28 trials of each of the 15 conditions presented in a random order.

Results

As in Experiment 1, and as shown in Figure 4, providing observers with additional encoding time considerably increased their ability to correctly identify an object from the study array. Also, the relative VSTM capacity between categories again depended on encoding time. However, the relationship between VSTM capacity for faces and objects differed from that between upright and inverted faces; at short encoding times (i.e. 500 ms) capacity for faces was less than that for other categories, however when given 2500 ms encoding time there was no difference between face and object VSTM capacity (see Figure 4 and 5). These results were not explained by a difference in response times between faces and the other categories (see APPENDIX B).

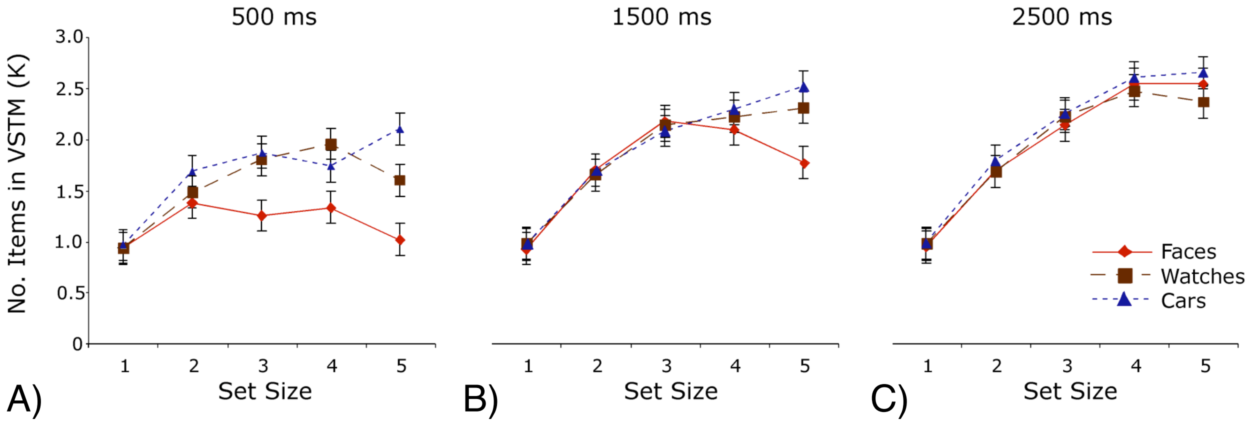


Figure 4. Estimated number of faces, watches or cars in visual short-term memory (VSTM) using Cowan's (2001) formula for each of the different set sizes when the memory array was presented for (A) 500 ms, (B) 1500 ms, or (C) 2500 ms. VSTM increases with additional encoding time for all three categories, and is similar across the categories when ample encoding time is allowed. However, VSTM for faces is lower in the shorter presentation duration conditions compared to that for the object categories. Error bars represent pooled standard error values.

An ANOVA performed on these data confirmed these results. There was a main effect presentation duration, F(2,40)=128.76, p≤.0001, with VSTM performance greater with longer presentation durations. There was also a main effect of stimulus category, F(2,40)=12.06, p≤.0001, with VSTM performance greater for watches or cars compared to faces. Most importantly, stimulus category interacted with presentation duration, with the effect of presentation duration being larger for faces than watches or cars, F(4,80)=6.52, p≤.0001.[5]

An ANOVA on the K-max for each category across the different set sizes (Figure 5) also supported the above results with a main effect of category, F(2,40)=5.13, p=.01, and presentation duration, F(2,40)=57.89, p≤.0001. The interaction between category and presentation duration approached significance, F(4,80)=2.14, p=.084. There appeared to be an overall advantage for cars, which may reflect participants' ability to use the variability in views to facilitate VSTM performance. When cars were removed from the analysis the interaction between presentation duration and stimulus category reached significance, F(2,40)=3.76, p=0.032. Paired t-tests revealed a significant difference in K-max between faces and cars, t(20)=4.05, p=.0006, and faces and watches, t(20)=3.27, p=.0038, in the 500 ms presentation condition, but not the 1500 ms or the 2500 ms conditions (all ps>.09; there was a trend for a general advantage for cars over the other stimulus classes).

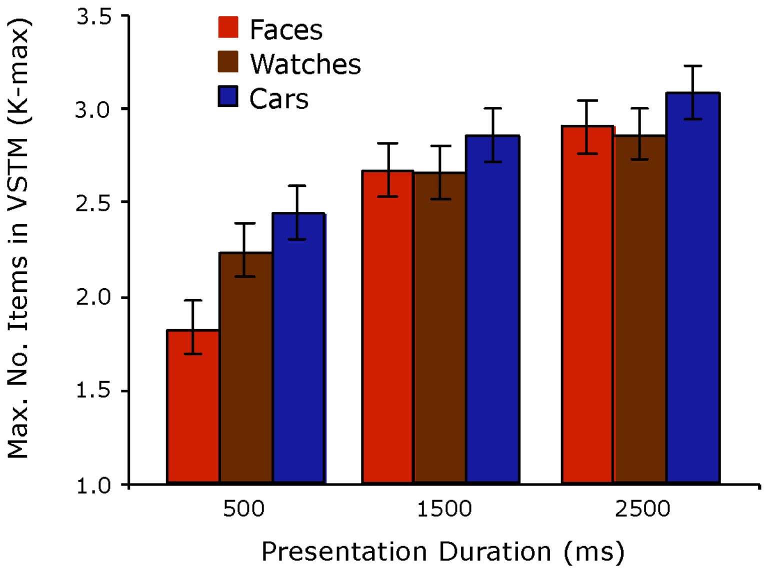


Figure 5. The maximum number of faces, watches, or cars (K-max) in visual short-term memory (VSTM) when the memory array was presented for 500 ms, 1500 ms, or 2500 ms. K-max increased with presentation duration for each of the categories, however the benefit of additional encoding time was greatest for faces. Error bars represent pooled standard error values.

---

[5] Additionally, the main effect of set size and its interactions with encoding duration and with stimulus category were significant, p<.05. These effects were not central to our research question.

Discussion

As predicted based on the proposed role of processing strategy on VSTM capacity, participants experienced a greater increase in VSTM capacity with additional encoding time for upright faces compared to objects. These results allow us to reject the possibility that the greater benefit for upright compared to inverted face VSTM capacity in Experiment 1 was due to differences between canonical and non-canonical stimulus views. Therefore, the results of Experiment 2 support suggestions that both encoding capacity limitations (encoding time) and processing strategy influence VSTM capacity. However, inconsistent with our predictions, VSTM for faces did not exceed that for objects. Possible explanations are discussed later and explored in Experiment 3.

The smaller VSTM capacity for faces compared to objects at the shortest presentation duration (500 ms) likely reflects a general difference in perceptual complexity or within category homogeneity between faces and objects. This had not been an issue in Experiment 1. The lower VSTM capacity for faces compared to objects at short encoding times suggests that faces may have a greater information load compared to watches or cars, and this is consistent with slower search rates for faces compared to those for multiple different object categories varying in complexity (Alvarez & Cavanagh, 2004; Eng et al., in press; Mruczek & Sheinberg, in press). Indeed, evidence suggests that our capacity for simultaneously processing faces in an identity recognition task may be as little as one (Bindemann, Burton, & Jenkins, in press). Importantly, the absence of a difference between upright and inverted face VSTM capacity at short encoding times in Experiment 1 suggests that the difference in encoding limitations between faces and objects in Experiment 2 may be unrelated to processing strategy. Instead, it may be due to stimulus factors such as perceptual complexity that are also reflected in measures of information load.

Our results, consistent with those of Eng et al. (in press), suggest that additional encoding time for items to be stored in VSTM can partly compensate for differences in information load between stimulus categories. Despite the smaller VSTM capacity for faces relative to cars or watches at short presentation durations, VSTM for faces approximates that of the other categories with sufficient encoding time. However, in Experiment 1, VSTM for inverted faces did not reach that of upright faces even when the memory array was presented for 2500 ms. Further, these results suggest that the influence of encoding time on VSTM capacity depends on processing strategy, with greater benefits of encoding time for more holistically processed faces compared to more featurally processed inverted faces and familiar objects.

In sum, based on the results of Experiment 1, we expected VSTM capacity for upright faces to exceed that of familiar objects at the longest presentation duration. However, the lower VSTM capacity for faces compared to objects at short presentation durations suggests that VSTM for faces may be disadvantaged at short encoding times due to the greater information load or higher homogeneity. Such differences between faces and objects may be too great to be fully offset by the recruitment of a different processing strategy. Alternatively, as suggested by Eng and colleagues (in press), 2500 ms may not be sufficient time to eliminate perceptual encoding limitations for faces. We address this possibility in Experiment 3.

Experiment 3

Experiments 1 and 2 demonstrated that VSTM for faces benefits more from additional encoding time than does VSTM for cars, watches, or inverted faces. It is possible that VSTM for upright faces would continue to benefit from even more encoding time (i.e. >2500 ms), perhaps then exceeding VSTM for objects. To investigate this possibility, Experiment 3 measured VSTM capacity for upright faces, watches and cars with 4000 ms encoding.

Methods

Participants. Participants were 36 (18 female) employees, undergraduate students or graduate students of Vanderbilt University with normal or corrected to normal vision, ranging in age from 18 to 33 (M =20.85, SD =3.01). Participants were compensated $10 per hour for their time.

Stimuli. The stimuli were the same as in Experiment 2 except that the viewpoint of the cars was controlled to avoid any advantage for VSTM capacity for cars due to the use of viewpoint information.

Design, Procedure, and Analysis. The general design, procedure, and analysis was the same as in Experiment 2 except that the study array was always presented for 4000 ms, and only set sizes 3, 4, and 5 were used. For each category there was a total of 84 trials (28 trials for each of the 3 set sizes). We note that the smaller number of trials per category than Experiment 2 (420) may lead to a general drop in performance, however this should be independent of any advantage for faces over cars or watches due to processing mode.

One participant with a VSTM capacity for cars greater than 2.7 SD above the mean was excluded as an outlier.

Results

As shown in Figure 6, when the memory array was presented for 4000 ms VSTM performance for faces exceeded that for either cars or watches. An ANOVA performed on these data confirmed this result. There was a main effect of category, F(2,70)=7.03, p=.0017, with VSTM capacity for faces exceeding that for both watches and cars. There was also a main effect of set size, F(2,70)=9.91, p=.0002, but no interaction between set size and category, F(4,140)=1.14, p=.3396, suggesting that the VSTM advantage for faces did not differ across the three set size conditions. The advantage for faces was confirmed by a priori 1-tailed paired t-tests (Faces vs. cars, t(107)=3.28, p=.0007; Faces vs. watches, t(107)=23.69, p=.0002). These results were not explained by a difference in response times for faces compared to cars or watches (see APPENDIX C).

An ANOVA on the K-max values across the different set sizes found a marginally significant effect of category, F(2,70)=2.66, p=.077. A priori 1-tailed paired t-tests comparing K-max for each stimulus category across the different set sizes once again confirmed the advantage for faces over watches, t(35)=2.46, p=.0094, and cars, t(35)=1.86, p=.035. K-max for watches and cars did not differ from each other, t(35)=0.01, p=.990 (2-tailed). VSTM performance was generally lower than in Experiment 2, suggesting that participants in Experiment 2 who performed five times the number of trials per stimulus category compared to those in Experiment 3, benefited from the greater exposure to the task and stimuli.

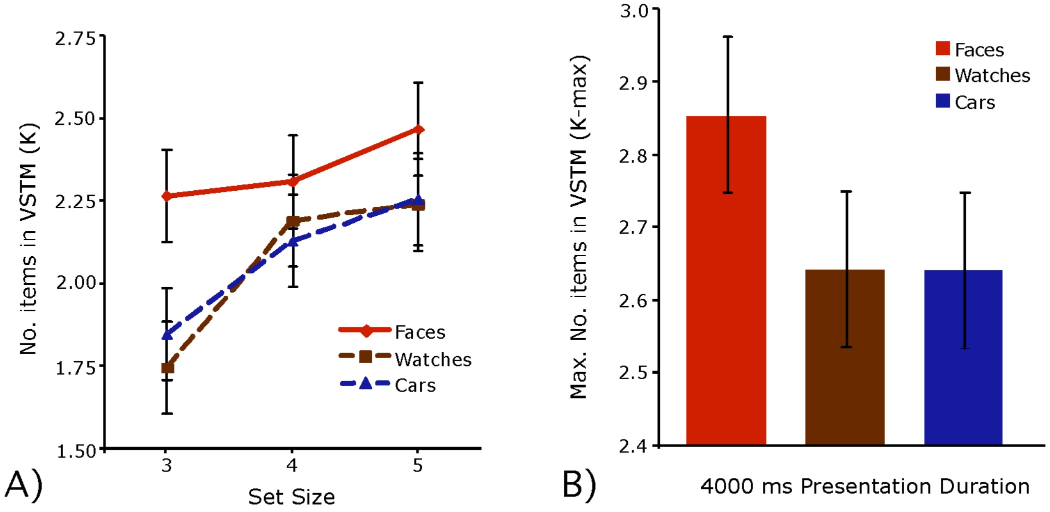


Figure 6. (A) Estimated number of faces, watches or cars stored in visual short-term memory (VSTM) using Cowan’s (2001) formula for each of the different set sizes when the memory array was presented for 4000 ms, and (B) the maximum number of objects in VSTM (K-max) for each category. K and K-max values were larger for faces than for the non-face categories with this longer presentation duration. Error bars represent pooled standard error values.

Discussion

Consistent with the proposed influence of face holistic processing mechanisms in increasing VSTM capacity, capacity for upright faces exceeded that for watches and cars with sufficient encoding time. Therefore, processing strategy was not only able to compensate for differences in complexity between stimulus categories, but it also resulted in an advantage for upright faces as long as sufficient time was allowed.

The set size condition in which VSTM capacity peaked (i.e. K-max) varied across participants. VSTM appeared to peak at set sizes less than five for some participants, especially in the face condition, with performance dropping for set sizes that exceeded participants’ VSTM capacity. This is also reflected in the results of Experiment 2 where face VSTM “dipped” at larger set sizes (see Figure 4 A and B). This result suggests that participants may have experienced more interference due to the presence of additional items beyond their VSTM capacity in trials with faces. This effect could possibly arise from the reported ability of faces to grab attention (Vuilleumier, 2000). This difference between face and object processing provides a potential explanation for the larger difference between face and object VSTM for smaller set sizes, compared to that at larger set sizes; averaging k-values across individuals whose VSTM capacity peaked at different set sizes may have led to a reduction in the average K values for the larger face set size conditions. Therefore, K-max may be a better measure for comparing VSTM performance for face and non-face categories.

General Discussion

Different classes of stimuli can recruit different processing strategies either through our experience with them or through innate biases. The holistic strategy recruited for face processing confers an advantage for face perception, making it less susceptible to irrelevant feature changes (e.g. hairstyle) and increases its sensitivity to subtle configural differences (Diamond & Carey, 1977; Tanaka & Sengco, 1997). The results from Experiments 1-3 suggest that such differences in processing strategy influence VSTM capacity, a consideration that has been unexplored until now. This benefit to upright face VSTM capacity cannot be explained by differences in image complexity, or by a general advantage for objects in a familiar orientation. The use of unfamiliar faces and of a concurrent articulatory suppression task also renders alternative explanations involving a contribution from verbal working memory unlikely. Long-term memory is also unlikely to explain these results: not only were the faces unfamiliar, but previous studies suggest that long-term memory traces created in the context of a VSTM task have limited impact on capacity (Olson & Jiang, in press). However, nonetheless this issue is explored further in the next chapter.

A recent study proposed that for intra-category objects that cannot be easily labeled, VSTM capacity is only one object (Olsson & Poom, 2005). The authors argued that estimates of VSTM capacity are inflated by observers assigning verbal labels to objects that cross category boundaries and by the use of the change-detection paradigm, which allows relational coding. However, our stimuli were intra-categorical and did not have obvious labels. That is, at least the watches and faces should not be any more easily labeled than any discriminable intra-category objects, such as those used by Olsson and Poom (2005). The car labels may have been more familiar, but it is unlikely that our novice subjects could label most of our stimuli accurately. The stimuli in the studies reported in this chapter were also more complex than those used by Poom & Olsson (2005), which should have reduced capacity (Alvarez & Cavanagh, 2004; Eng et al., in press). However, VSTM capacity surpassed two items for cars or watches, and even reached three items for faces. Indeed, one factor not considered by Olsson & Poom (2005) is that intra-category objects may require more encoding time than objects that cross category boundaries.

The results reported in this chapter and those of Eng et al. (in press) suggest that VSTM capacity for complex objects is underestimated when the memory array is presented for only 500 ms because of encoding limitations. For example, it is possible that capacity for the geometrical objects used by Olsson & Poom (2005) could reach the same level that we found for watches given enough encoding time.

The results reported in this chapter suggest that VSTM capacity is not a static feature of an object category, but instead can be influenced by both encoding time and processing strategy. This raises questions about the relationship between visual search rate and VSTM capacity (Alvarez & Cavanagh, 2004): Under what conditions does visual search rate predict VSTM capacity? Results of a recent study suggest that search rate may predominantly reflect encoding limitations due to perceptual complexity (Eng et al., in press). However, search rate can be influenced by differences between categories that do not differ in perceptual complexity, such as upright and inverted faces (Tong & Nakayama, 1999). Further studies are needed to better understand the relationship between visual search rate and VSTM capacity.

The recruitment of holistic processing mechanisms, generally linked with visual expertise (Gauthier & Tarr, 2002), may increase VSTM capacity by creating more efficient representations for storage. Existing work suggests that VSTM capacity for multi-featured items depends on the "binding" of features together (Delvenne & Bruyer, 2004; Wheeler & Treisman, 2002). Similarly, holistic processing may bind the information within a face more effectively than the more feature-based strategies used for non-face objects, leading to a more detailed and/or more efficient representation for faces. This would be consistent with the faster search rate for upright compared to inverted faces (e.g., Tong & Nakayama, 1999). Alternatively, holistic representations for upright faces may be more robust and less susceptible to decay or interference than are the more feature-based representations created for inverted faces or objects. Further studies manipulating the similarity and relevance of different dimensions of the probe and study items may provide insight into the nature of the representations stored in VSTM for objects and faces.

A recent study provides evidence against the possibility that holistic representations are simply more robust and less subject to decay or interference. If the VSTM advantage for faces arose due to holistic representation being more robust than more piecemeal representation, one would expect that the cost of inversion for VSTM performance would increase with increasing delay intervals between study and test. However, Freire, Lee, and Symons (2000) found no effect of memory delay (1 s – 10 s) on the size of the inversion effect. This result suggests that the advantage for face VSTM is not a result of more robust representations and is consistent with a more qualitative difference between holistic and more featural representations.

Thus far in the context of previous literature, the results suggest that not only do we perceive faces differently, but that this difference extends to our VSTM for this category. Indeed, more efficient perception of important categories in our environment would not be maximally adaptive if this advantage disappeared as soon as we lose sight of those objects. Interestingly, because holistic processing is also observed for objects of expertise (Busey & Vanderkolk, 2005; Gauthier & Tarr, 2002), the VSTM advantage reported for faces in this chapter may extend more generally to perceptual experts. If so, this would lead to an intriguing prediction: although practice on VSTM tasks has not been found to improve VSTM capacity substantially (Chen et al., in press), expertise training procedures that have been shown to increase holistic processing for objects might more likely impact VSTM. In addition, real-world experts, shown to demonstrate holistic processing for objects in their domain of expertise (e.g., cars among car experts) should also show increased VSTM capacity. This possibility is explored in CHAPTER III.

CHAPTER III

COMPARING VISUAL SHORT-TERM MEMORY FOR UPRIGHT AND INVERTED FACES AND CARS AMONG CAR EXPERTS AND CAR NOVICES

Introduction

Visual experts with non-face categories recruit a holistic processing strategy not unlike that recruited for upright faces (Diamond & Carey, 1986; Gauthier et al., 2003; Gauthier & Tarr, 2002). To review from CHAPTER I, classic holistic processing effects typically found with faces, such as sensitivity to inversion or configural changes, have been demonstrated among observers trained to become experts with a novel category ("greebles"; Gauthier & Tarr, 2002). In addition, real-world car experts also show holistic processing effects, namely the difficulty processing only a part of a car when it is presented in the context of a whole car (Gauthier et al., 2003). These results provide evidence that the holistic processing strategy recruited for faces is also associated with expert visual processing more generally. This suggests an interesting possibility: if the VSTM advantage for upright faces over inverted faces and non-face-objects (CHAPTER II) is a result of the more holistic processing style recruited for upright faces, visual experts with non-face categories should also demonstrate a VSTM advantage for objects within their domain of expertise. CHAPTER III explores this possibility.

CHAPTER III consists of four experiments that compare VSTM capacity for upright and inverted faces and cars among groups of car experts and car novices. Experiment 4 compares VSTM for upright and inverted faces and cars under similar temporal encoding limitations as those used in CHAPTER II. To anticipate the results, this study not only replicated the advantage for upright faces over inverted faces, but also found a similar inversion cost for cars among car experts but not novices. Car experts also had a greater VSTM capacity for cars compared to car novices. The other three experiments in this chapter were conducted in order to evaluate alternative explanations, unrelated to holistic processing, for the VSTM advantage for objects within one’s domain of expertise. Experiment 5 tested whether verbal working memory contributed to car experts’ superior VSTM performance: The articulatory suppression load was increased to a level previously shown to cause a large reduction in the level of performance on a verbal short-term memory task (Marsh & Hicks, 1998; Otani et al., 1997). The increase in articulatory suppression load did not lead to any qualitative changes to the general pattern of results, with car experts still demonstrating a VSTM advantage for cars and a greater inversion cost to car VSTM compared to novices. Experiment 6 explored whether LTM contributes to the VSTM advantage demonstrated for upright faces and upright cars among car experts. In this experiment participants performed a number of VSTM trials with a small (10 item) stimulus set, allowing them to establish LTM memory representations for these 10 items. The stimulus set was then changed to 10 new items from the same category. VSTM performance for faces and cars among car experts was unaffected by the change in stimulus set, but there was evidence that VSTM for cars was reduced among car novices. These results suggest that the VSTM advantage for objects from one's domain of visual expertise is unlikely to be a result of contributions from verbal short-term memory or LTM. Finally, Experiment 7 explored whether the VSTM advantage could result from a difference in eye-movement strategies employed to uptake the information from a study array. Imposing a restricted eye-movement pattern failed to eliminate the VSTM advantage for upright faces, suggesting that this advantage does not rely on strategic difference in the physical pattern of eye-movements employed to encode a study array. Together, the results reported in this chapter are consistent with a holistic processing account of the VSTM advantage demonstrated for objects of expertise.

Experiment 4

If the VSTM advantage for faces relies on the recruitment of specialized holistic processing mechanisms linked with visual expertise more generally (Gauthier & Tarr, 2002), experts with non-face object categories should show the same VSTM advantage as seen with faces. More specifically, car experts should experience a greater benefit to VSTM for cars with additional encoding time compared to novices. In addition, VSTM capacity for cars should be greater among car experts than novices as long as sufficient encoding time is allowed, but VSTM for upright faces should be similar across the groups regardless of encoding time. VSTM for cars among car experts should also be disproportionately reduced by inversion compared to novices, whereas the cost due to inversion for faces should be similar across the two groups. Alternatively, the advantage gained by recruiting a holistic processing style may be specific to faces. For example, it might be the result of an interaction between processing strategy and the perceptual features of faces, and thus not extend to other categories that recruit a holistic processing strategy. It is also possible that the VSTM advantage for faces is a product of their evolutionary importance, with special mechanisms in place to increase VSTM for this category but no other, consistent with the suggestion by some that there is an innate module dedicated to processing facial stimuli (Kanwisher, 2000). Experiment 4 examines these possibilities by comparing VSTM capacity for upright and inverted cars and faces among participants who are car novices or car experts while manipulating the time allowed for encoding the items in VSTM. If the VSTM advantage for faces reported in CHAPTER II is associated with visual expertise rather than being specific to faces, VSTM for cars among car experts, but not novices, should show the same pattern of results as that found for faces.

Methods

Participants. Participants were 36 individuals who where employees, undergraduate students, or graduate students of Vanderbilt University, or members of the surrounding Nashville community. Participants were compensated $10 per hour for their time. All participants had normal or corrected to normal vision and reported having a range of experience identifying cars, from very little to extensive. A self-report measure of participants' car and bird expertise was obtained from each individual in the form of a rating on a scale of one to ten. Participants were informed that "five" corresponded to average skill at identifying cars or birds whereas "ten" reflected perfect skill recognizing these categories.

Due to differences in criteria across individuals, an objective test of their expertise was also obtained. Participants' subordinate-level car expertise was quantified using a sequential matching task that has also been used in pervious studies (Gauthier, Curby, Skudlarski, & Epstein, 2005; Gauthier et al., 2003; Gauthier et al., 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 APPENDIX D). This task can be performed 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, participants who reported no expertise with birds also performed the same task with birds, in which they were required to make a same/different decision at the level of species about different images of passerine birds. Participants with a d' for car trials greater than two and at least a d' advantage of one unit for cars over birds[6] (i.e. car d' – bird d' > 1), were classified as experts (Gauthier et al., 2000).

Eighteen participants (11 males) met the criteria for car expertise (age, M = 22.28, SD = 4.71, car d' M = 2.55, bird d' M = 0.87), while the remaining 18 (10 male) were classified as car novices (age, M = 20.64, SD = 2.42, car d' M = 1.34, bird d' M = 0.84). One of the participants classified as an expert had a (car d' – bird d') less than one (0.83), but 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' prime measures for these categories. Car expertise scores from the matching task were generally consistent with subjects' self-report, with participants classified as novices reporting their car recognition skills at an average of 5.77/10[7]; those who met criteria for car expertise rated their skills, on average, as 8.00/10.

Stimuli. The stimuli were 72 grayscale faces (1.9° x 2.3°) from the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany and 72 grayscale images of cars (2.3° x 1.5°, profile view). The stimuli were the same as those used in Experiment 3.

Procedure. For each participant, a selection of half the faces and half the cars appeared in the upright trials while the remaining images appeared in the inverted trials. Participants performed a delayed match-to-sample probe recognition task like that described in CHAPTER II (see Figure 1). Participants performed a total of 1152 trials across four different one hour long sessions, each consisting of 8 alternating blocks (36 randomized trials/block) of upright and inverted images. Two sessions consisted of only face trials, while the other two sessions consisted of only car trials. Participants either completed the two car sessions or the two face sessions first. The order of the sessions was counterbalanced both within and across expertise groups. In sum, across the experiment, there were 288 trials for each of the four stimulus categories (upright faces, inverted faces, upright cars, inverted cars). Within these stimulus categories, there were 9 conditions (3 set sizes x 3 presentation durations), each of which was presented 32 times.

---

[6] Providing the participant did not also report expertise with identifying passerine bird species

[7] Self-ratings were not obtained from five participants who did not meet the criteria for expertise. However these participants were recruited as novices, not as experts.

Analysis. Data from one participant were removed from the analysis due to poor performance (73%) in the auditory task. Accuracy for this component exceeded 90% for all other participants. For remaining participants, incorrect articulatory suppression trials were removed from further analyses. For each participant, hit rate and correct rejection rate were calculated for each condition. These values were used to derive VSTM capacity using Cowan's K, K= (hit rate + correct rejection rate - 1) * set size, an estimate of the number of objects successfully encoded in VSTM (Cowan, 2001). An additional participant was removed due to poor performance on the VSTM task (average capacity for upright faces less than 1). These subjects were not included in the previous description of the participant groups to ensure that the descriptives reported reflect only those participants whose data was included in the statistics reported in the results section. The maximum K (K-max) was identified for each presentation duration regardless of set size. Since some participants tend to show a drop in performance in conditions where the set size exceeds their capacity, especially in the case with faces, VSTM capacity may be underestimated in the larger set size conditions (Experiment 3). To avoid this problem, all subsequent analyses were performed on the K-max values. In addition to the ANOVA analyses and the regression analyses exploring the relationship between level of car expertise and VSTM capacity, a series of planned t- tests were conducted to explore the specific predictions based on the proposed role of holistic processing in explaining the VSTM capacity advantage for faces found in CHAPTER II, that is (1) the presence of an inversion cost for cars among car experts but not novices, (2) greater VSTM for cars among car experts compared to novices, and (3) greater VSTM capacity for faces than cars among car novices, with sufficient encoding time.

Results

To summarize the results, VSTM for faces was similar across the two groups with inversion reducing face VSTM at all presentation durations (Figure 7A). In contrast, inversion of cars only impacted VSTM for cars among car experts and only when the presentation duration was long enough to allow sufficient encoding time (≥ 2500 ms) (Figure 7 B). In addition, while VSTM for faces was no different from that for cars among car experts regardless of the duration of the memory array presentation, car novices demonstrated an advantage for upright faces over upright cars at the longest presentation duration (4000 ms). Further confirming the importance of encoding time, car experts demonstrated greater VSTM for upright cars than novices only when the presentation duration was sufficiently long (≥ 2500 ms) (Figure 7 B). Finally, highlighting the continuous, rather than binary, nature of perceptual expertise, a participant's level of car expertise was correlated with the estimate of the maximum number of cars that that participant could successfully store in VSTM when the memory array was presented for 4000 ms (Figure 8 A). These results were not explained by a difference in response times between upright and inverted faces and cars (see APPENDIX E).

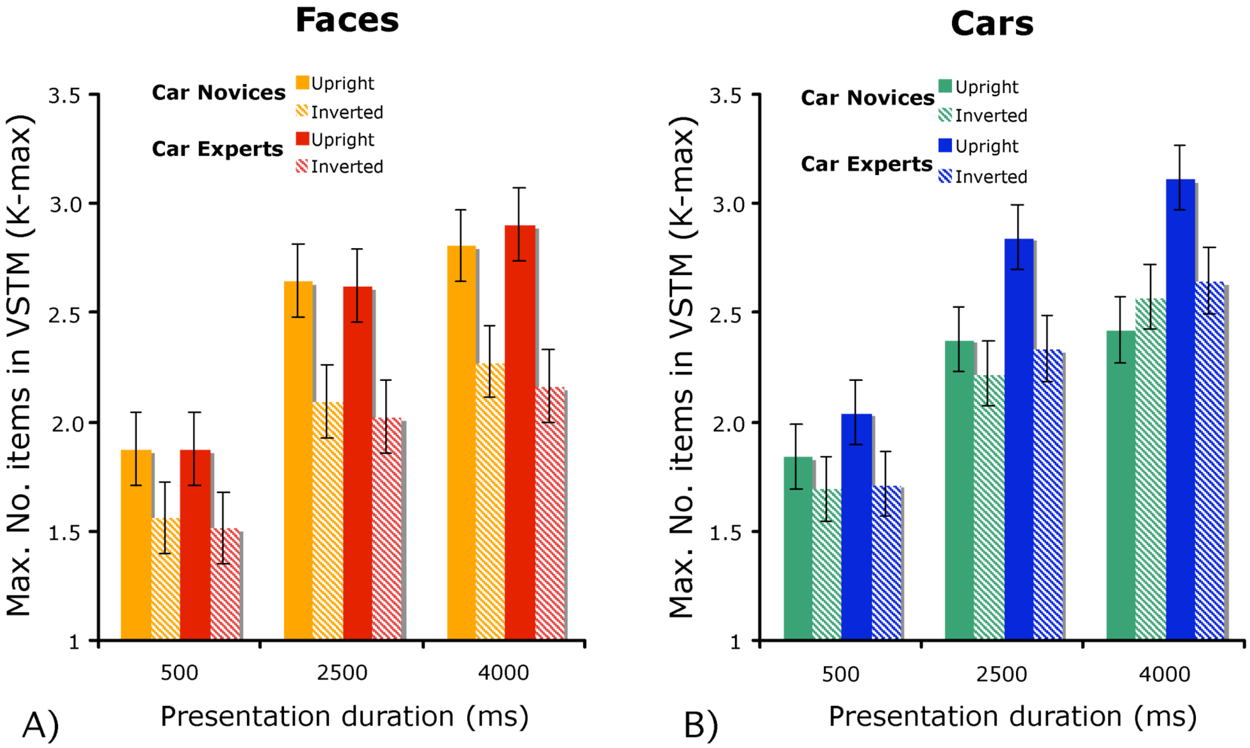


Figure 7. The maximum number of objects (K-max) in visual short-term memory (VSTM) for 500 ms, 2500 ms, and 4000 ms presentation durations for (A) upright and inverted faces and (B) upright and inverted cars among participants who were car experts and novices. Not only was there a VSTM advantage for upright cars among cars experts similar in magnitude as the advantage for upright faces, but also car experts, but not novices, showed an inversion effect for cars. Error bars represent pooled standard error values.

Omnibus test. Before proceeding directly to more specific planned comparisons, the data were submitted to an omnibus ANOVA. A 2 (category: face, car) x 2 (orientation: upright, inverted) x 3 (presentation duration; 500, 2500, 4000 ms) x 2 (car expertise group; novice, expert) ANOVA on the maximum VSTM capacity (K-max) revealed main effects of orientation, F(1,34)=82.81, p≤.0001, and presentation duration, F(2,68)=73.28, p≤.0001. The main effect of category also approached significance, F(1,34)=3.58, p=.067. These results suggest that VSTM was greater for upright items, longer presentation durations, and that a trend for a greater VSTM for faces compared to cars was also present. No interactions between presentation duration and any other variables were significant (all ps>0.253). There was an interaction between orientation and category, F(1,34)=5.48, p=.025, with the effect of orientation in general being greater for faces than cars. There was no main effect of expertise (F<1), but there was an interaction between orientation and expertise, F(1,34) = 8.31, p = .0068, and between category and expertise, F(1,34) = 4.97, p = .032, suggesting that the VSTM advantage of car experts over car novices was limited to cars (i.e. it did not extend to faces) and that car experts also experienced a greater cost due to orientation compared to novices. The three-way interaction between expertise, orientation, and category failed to reach significance, F(1,34)=1.45, p=.238. In order to better understand the patterns in these data, subsequent ANOVAs were performed separately for faces and cars as described below.

Comparing maximum VSTM for faces among car experts and novices. A 2 (orientation: upright, inverted) x 3 (presentation duration; 500, 2500, 4000 ms) x 2 (car expertise group; novice, expert) ANOVA on the maximum VSTM capacity (K-max) for upright and inverted faces found main effects of orientation, F(1,34)=69.53, p≤.0001, and duration, F(2,68)=40.24, p≤.0001, but importantly there was no main effect (F<1), or interaction involving car expertise (all ps>.425). The interaction between orientation and duration failed to reach significance, F(2,68)=1.77, p=.178. Therefore, VSTM for faces was significantly greater for longer presentation durations and for upright compared to inverted faces, although, the effect of orientation did not differ depending on presentation duration (Figure 7 A). In addition, an observer's expertise with cars did not impact VSTM for faces.

Comparing maximum VSTM for cars among car experts and novices. A 2 (orientation: upright, inverted) x 3 (presentation duration; 500, 2500, 4000 ms) x 2 (car expertise group; novice, expert) ANOVA on the maximum VSTM capacity (K-max) for upright and inverted cars found main effects of orientation, F(1,34)=9.12, p=.0048, and presentation duration, F(1,34)=61.93, p≤.0001, although the interaction between orientation and presentation duration did not reach significance (F<1). There was no main effect of car expertise, F(1,34)=1.95, p=0.172, but there was an interaction between car expertise and orientation, F(1,34)=5.61, p=.024, with greater VSTM for upright but not inverted cars, among car experts compared to car novices. The interactions between presentation duration, and expertise and/or orientation failed to reach failed to reach significance, (all ps>.230). Therefore, in general, VSTM was greater for longer presentation durations and for upright compared to inverted items. Notably, consistent with the predictions, VSTM for upright cars, but not inverted cars, was significantly greater among car experts than novices (Figure 7 B).

Planned comparisons. The specific predictions of the expertise account of the VSTM advantage for faces were generally confirmed by planned t-tests: maximum VSTM for cars was reduced by inversion for all presentation durations among experts (all ps<.022) but not novices (all ps>.206), while maximum VSTM for faces was reduced by inversion for all presentation durations among both expert and novice groups (experts, all ps<.010; novices, all ps<.0333). In addition, maximum VSTM for upright faces exceeded that for upright cars among car novices only when the presentation duration was sufficiently long (500 ms, t<1; 2500 ms, t(17)=1.48, p=.157; 4000 ms, t(17)=2.87, p=.011), however VSTM for cars and faces did not differ among car experts regardless of presentation duration (500 ms, t(17)=1.19, p=.251; 2500 ms, t(17)=1.16, p=.263; 4000 ms, t(17)=1.17, p=.257). Consistent with the need for sufficient encoding time in order for novices to show an advantage for faces over cars, a VSTM advantage for cars among car experts compared to car novices also only emerged when the presentation duration was sufficiently long (500 ms, t<1; 2500 ms, t(34)=1.68, p=.103; 4000 ms, t(34)=3.05, p=.0044).

Correlation between car expertise and maximum VSTM for cars and faces. There was a significant correlation between K-max for upright cars when the memory array was presented for 4000 ms and participants' sensitivity for cars in the car expertise measure (r=.388, p=.021). Notably, a participant's level of car expertise was not correlated with their VSTM for upright faces (r=.132, p=.451) or inverted cars (r=.149, p=.393), suggesting that the correlation between car expertise and VSTM for cars cannot be attributed to a more general difference in visual skills. The difference between a participant's sensitivity for cars and that for birds (i.e. car d' – bird d') may provide a better measure of a participant's expertise-specific skill. On the other hand, to the degree to which a participant's sensitivity for birds is increased by their experience with this category, this measure may underestimate learning-related increases in car sensitivity. Despite this possibility, there was a significant correlation between participants' delta d' score (i.e. car d' – bird d') and K-max with 4000 ms encoding time for cars (r=.364, p=.032; Figure 8 A), but once again not for upright faces (r=.032, p=.855; Figure 8 B) or inverted cars (r=.036, p=.838)[8].

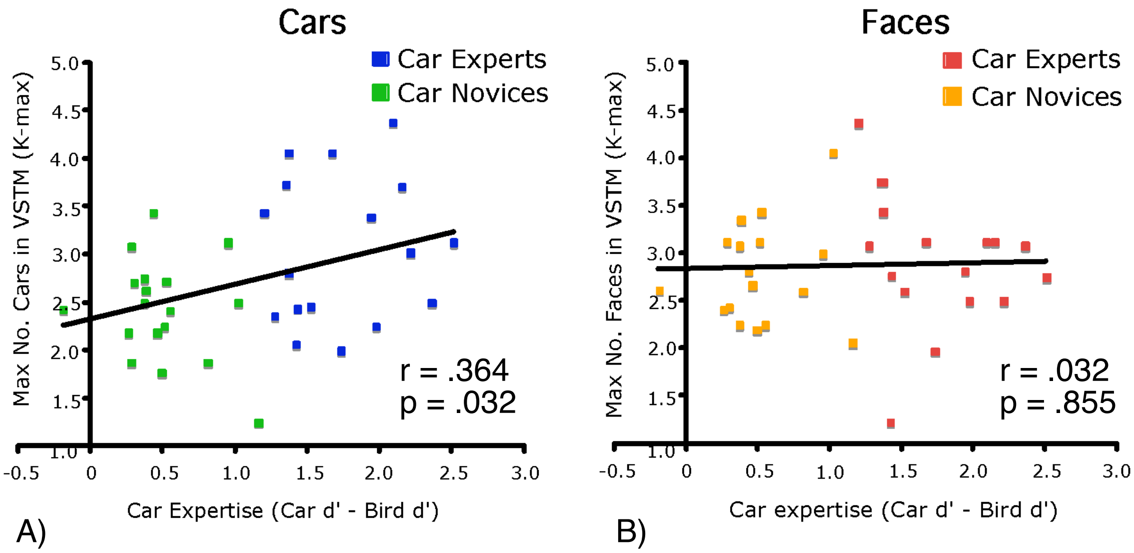


Figure 8. Scatter plot of individuals' car expertise scores (Car d' – Bird d') and their K-max for (A) upright cars and (B) upright faces when the memory array was presented for 4000 ms. There was a significant correlation between a participant's level of car expertise and their VSTM capacity for cars, but not for faces.

Discussion

As predicted based on the proposed influence of processing strategy on VSTM capacity, car experts demonstrated greater VSTM capacity for upright cars than car novices. Similar to the VSTM advantage for faces demonstrated in CHAPTER II, this advantage also depended on sufficient encoding time. In addition, inversion, believed to disrupt holistic processing, led to a reduction in VSTM capacity for cars among car experts but not novices. These results suggest that the VSTM advantage for faces is not due to a face-specific mechanism, as other objects within a domain of expertise can also demonstrate this advantage. These results are also

---

[8] The participant who reported above average skills with both cars and birds who had the reduced (car d' – bird d') expertise score was not included in the regression analysis.

consistent with the suggestion that the specialized holistic processing strategy, adopted for faces and by visual experts, may be the mechanism underlying this advantage.

There was a considerable amount of variability in the data, which prevented some effects in the omnibus ANOVA from reaching significance. This is not surprising given the numerous reports of robust individual differences in VSTM capacity among individuals (Todd & Marois, 2005; Vogel & Machizawa, 2004). Participants were grouped into experts and novices in order to reduce the impact of individual differences unrelated to expertise and to emphasize those specific to expertise, although perceptual expertise also falls on a continuum with some variability within the groups. Consistent with the presence of meaningful expertise-related individual differences within the groups, VSTM capacity for cars was correlated with level of visual expertise with cars. Sensitivity on this measure of car expertise has been previously shown to also correlate with behavioral measures of holistic processing (Curby & Gauthier, 2001; Gauthier et al., 2003) and also with a number of physiological markers typically associated with face processing, such as activation to cars in face-selective fusiform cortex (Gauthier et al., 2005; Gauthier et al., 2000) and the amplitude of a face-selective electrophysiological marker, the N170 potential, to cars (Gauthier et al., 2003). Therefore in the context of the findings from previous studies, the correlation between VSTM capacity for cars and participants' level of car expertise not only strengthens the link between the VSTM advantage demonstrated for faces with that for cars among car experts, but it also provides consistent, but indirect, evidence for a common holistic processing mechanism for this advantage.

The importance of sufficient encoding time to demonstrate an advantage of perceptual expertise is affirmed by the absence of many of the effects of expertise at short encoding times. For example, as in Experiment 3, an advantage for faces compared to upright cars among car novices was only present in the 4000 ms presentation duration condition. In addition, car experts also only demonstrated greater VSTM for cars compared to novices when the memory array was presented for 4000 ms. The similar temporal constraints and magnitude of the face and car expert VSTM advantages also provide support for the notion of a common underlying mechanism supporting these abilities.

One important difference between car experts and novices could provide another potential explanation for these results: they possess different degrees of knowledge and labels for the car stimuli. This was not directly measured in Experiment 4, as knowledge of the car model labels was not necessary to perform successfully on the measure of car expertise. It is likely although, that as a by-product of their experience with cars, participants in the expert group had more extensive knowledge of the labels of the models used in this experiment. There was anecdotal evidence consistent with this suggestion; upon completion of the study, car experts frequently and spontaneously mentioned some of the names of the cars that had appeared in the experiment. Therefore, it is possible that car expertise may facilitate performance on the VSTM task by allowing participants to verbally recode the study array items and thus benefit from a contribution from verbal short-term memory. The relatively low articulatory suppression load used in Experiment 4 may have been insufficient to prevent experts from utilizing verbal short-term memory stores (Marsh & Hicks, 1998). This strategy would have been less available to novices due to their more limited knowledge of the labels for these stimuli, and thus could potentially account for the difference in performance between the two groups. It is important to note that this account cannot explain the advantage for unfamiliar faces over familiar non-face objects, although it is possible that there may be a different underlying cause for the VSTM advantage for faces and that demonstrated for cars by car experts. If the VSTM advantage for cars among car experts relies on a contribution from verbal short-term memory, this advantage should be reduced, or even disappear entirely, if car experts' ability to recruit verbal memory stores is reduced. This possibility is tested in Experiment 5.

Experiment 5

This experiment specifically aimed to address whether the VSTM advantage demonstrated by car experts depends on their being able to recruit verbal memory stores to supplement VSTM. Since car experts are likely to be highly familiar with the labels associated with the car stimuli, the recruitment of a verbal strategy could potentially account for the VSTM advantage demonstrated by car experts. The articulatory suppression load used in Experiment 4 may have been insufficient to prevent a contribution from verbal short-term memory as previous work has demonstrated that participants can still perform a verbal memory task with reasonable accuracy (82%) with an articulatory suppression load equivalent to that used in Experiment 4 (i.e. two syllables) (Marsh & Hicks, 1998)[9]. However, verbal working memory performance drops considerably (54%) if the articulatory suppression load is increased to six syllables. Based on these findings, increasing the articulatory suppression load from 2-3 syllables to 5-6 syllables should reduce the ability of participants to adopt a verbal strategy to facilitate VSTM performance. This increase in articulatory suppression load was implemented in Experiment 5 to test this prediction. In addition, to further discourage participants from using a verbal strategy, the articulatory suppression load was semantically meaningful; participants were required to rehearse three car model names during car trials and three person names during face trials. Therefore, not only should this articulatory suppression load fill a substantial portion, if not all, of a participant's verbal memory, but it should further discourage the use of a verbal strategy due to the potential of the verbal load to semantically interfere with a verbal rehearsal strategy.

Methods

Participants. Participants were 31 employees, undergraduate students, or graduate students of Vanderbilt University, or members of the surrounding Nashville community. Participants were compensated $10 per hour for their time. All participants had normal or corrected to normal vision and reported having a range of experience identifying cars, from very little to extensive. Their car expertise was quantified using the same matching task as in Gauthier et al. (2000) and described in APPENDIX D. Fourteen participants (11 males) met the criteria for car expertise (age, M = 21.64, SD = 2.10, car d' M = 2.72, bird d' M = 0.93), while 17 (13 male) were classified as car novices (age, M = 22.41, SD = 3.02, car d' M = 1.27, bird d' M = 0.76). Once again, participants' self-report of their skill at recognizing cars was generally consistent with their performance on the subordinate car matching task, with participants meeting the criteria for expertise on the task rating themselves an average of 8.43/10; those who were classified as novices rated their skills, on average, as 6.03/10.

Stimuli. The visual stimuli were the same as those used in Experiment 4. The auditory stimuli were changed as described below.

---

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

Procedure. The procedure was the same as in Experiment 1, except that participants were required to rehearse three car models (“Spectra”, “Blazer”, “Accord”, “Corvette”, “Mustang”, ”Jetta”, “Civic”, ”Tribute”, “Probe”, “Vibe”) or person names (“Leanne”, “Sarah”, “Alan”, “Cathryn”, “Amy”, “Andrew”, “Peter”, “Mark”, “Jane”, “Robert”) instead of the two digits. This resulted in 5-6 syllables being rehearsed compared to 2-3 in Experiment 1. In addition, Experiment 2 also differed from Experiment 1 in that the probe for the auditory task was presented auditorily rather than visually.

Analysis. The analysis was the same as used in Experiment 1. Three participants were removed from the analysis due to poor performance in the articulatory suppression task (<90%). An additional two participants were excluded from the analysis due to poor performance in the VSTM with faces (average K for upright faces<1). In addition to the ANOVA analyses and the regression analysis exploring the relationship between level of car expertise and VSTM capacity, a series of planned 1-tailed tests were conducted to explore the specific predictions based on the results of Experiment 1, that is (1) the presence of an inversion cost for cars among car experts but not novices, (2) greater VSTM for cars among car experts compared to novices, and (3) greater VSTM for faces than cars among car novices.

Results

The results of Experiment 5, in general, provide a replication of those reported in Experiment 4. To summarize the results, VSTM for faces was similar across the two groups, with inversion reducing face VSTM at all presentation duration conditions (Figure 9 A). In contrast, inversion of cars only impacted VSTM among car experts but not novices (Figure 9 B). In addition, while VSTM for faces was no different from that for cars among car experts regardless of the presentation duration, car novices demonstrated an advantage for upright faces over upright cars when the presentation duration was sufficiently long (4000 ms). Further confirming the importance of encoding time, car experts also demonstrated greater VSTM for upright cars than novices only when the presentation duration was sufficiently long (≥ 2500 ms) (Figure 9 B). Finally, highlighting the continuous, rather than binary, nature of perceptual expertise, a participant’s level of car expertise was correlated with the estimate of the maximum number of cars that that participant could successfully store in VSTM in 4000 ms presentation duration condition (Figure 10 A). These results were not explained by a difference in response times between upright and inverted faces and cars (see APPENDIX F).

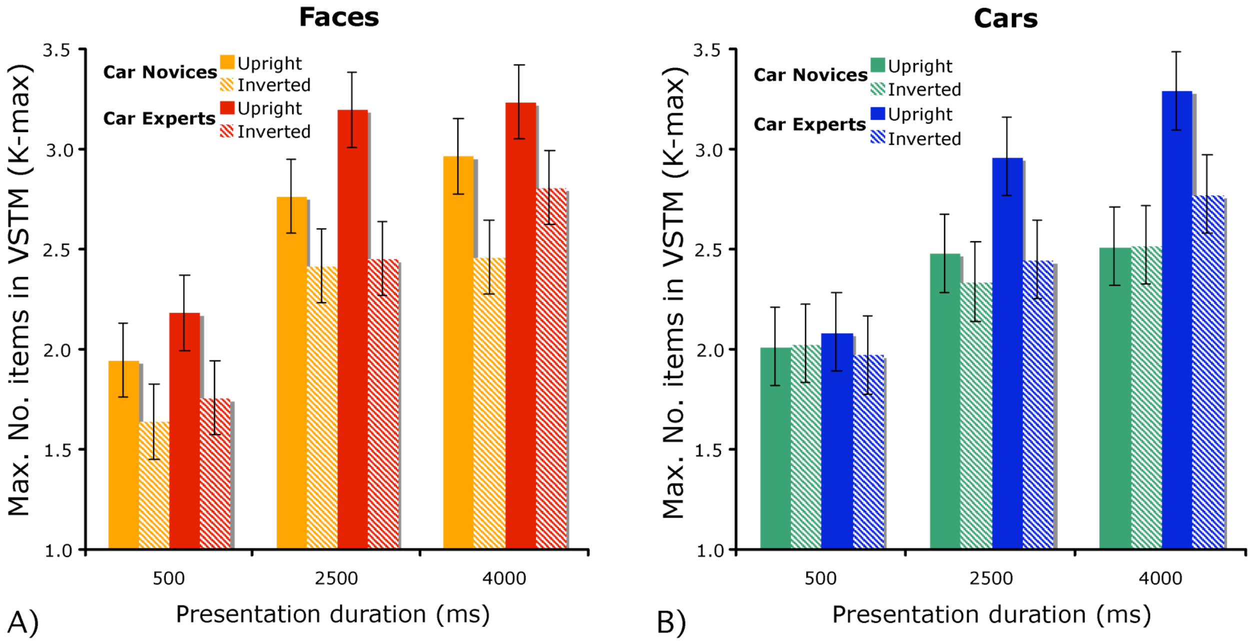


Figure 9. The maximum number of items (K-max) in visual short-term memory (VSTM) for 500 ms, 2500 ms, and 4000 ms presentation durations for (A) upright and inverted faces and (B) upright and inverted cars among participants who were car experts and car novices. Not only was there a VSTM advantage for upright cars among cars experts similar in magnitude as the advantage for upright faces, but also car experts, but not novices, showed an inversion effect for cars. Error bars represent pooled standard error values.

Omnibus test. Following the same procedure as with Experiment 4, before proceeding directly to more specific planned comparisons, the data were submitted to an omnibus ANOVA. A 2 (category: face, car) x 2 (orientation: upright, inverted) x 3 (presentation duration; 500, 2500, 4000ms) x 2 (car expertise group; novice, expert) ANOVA on the maximum VSTM capacity (K-max) across the different set sizes revealed main effects of orientation, F(1,29)=34.35, p≤.0001, and presentation duration, F(2,58)=65.30, p≤.0001. The main effect of expertise also approached significance, F(1,29)=3.66, p=.066, but there was no main effect of category (F<1). These results suggest that VSTM was greater for upright items, longer presentation durations, and that a trend for generally greater VSTM capacity among car experts compared to car novices. There was an interaction between orientation and expertise, F(1,29)=4.50, p=.043, and a trend for an interaction between expertise and presentation duration, F(1,58)=2.00, p=.144, but no other interaction with expertise approached significance (all ps >.267). The VSTM advantage for expertise was greater for upright than inverted items, with experts also tending to experience a greater benefit due to additional encoding time. There was also a trend for an interaction between presentation duration and category, F(1,58)=2.434, p=.097, but no other interactions with presentation duration approached significance (all ps>.269). There was an interaction between orientation and category, F(1,29)=7.91, p=.0087, with the effect of orientation in general being greater for faces than cars. In order to better understand the pattern of results in these data, subsequent ANOVAs were performed separately for faces and cars, as described below.

Comparing maximum VSTM for faces among car experts and novices. A 2 (orientation: upright, inverted) x 3 (presentation duration; 500, 2500, 4000 ms) x 2 (car expertise group; novice, expert) ANOVA on the maximum VSTM capacity (K-max) for upright and inverted faces found main effects of orientation, F(1,29)=32.61, p≤.0001, and presentation duration, F(2,58)=40.79, p≤.0001, but importantly there was no main effect, F(1, 29)=2.12, p=0.156, or interaction involving car expertise (all Fs<1). There was no significant interaction between orientation and presentation duration (F<1). Therefore, VSTM for faces was significantly greater for longer presentation durations and for upright compared to inverted faces, although, the effect of orientation did not differ depending on presentation duration (Figure 9 A). In addition, an observer's expertise with cars did not impact VSTM for faces.

Comparing maximum VSTM for cars among car experts and novices. A 2 (orientation: upright, inverted) x 3 (presentation duration; 500, 2500, 4000 ms) x 2 (car expertise group; novice, expert) ANOVA on the maximum VSTM capacity (K-max) for upright and inverted cars found main effects of orientation, F(1,29)=11.24, p=.0022, and presentation duration, F(1,29)=33.59, p≤.0001. The interaction between orientation and presentation duration did not reach significance, F(2,58)=1.10, p=.340. The main effect of car expertise did not reach significance, F(1,29)=2.71, p=0.110, but importantly there were interactions between car expertise and orientation, F(1,29)=7.40 , p=.011, and car expertise and presentation duration, F(1,29)=3.64, p=.032, with car experts not only demonstrating greater VSTM for upright but not inverted cars compared to car novices, but also a larger benefited from additional encoding time compared to novices. The interaction between presentation duration, expertise and orientation was not significant (F<1). Therefore, in general, VSTM was greater for longer presentation durations and for upright compared to inverted items. Notably, consistent with our predictions, VSTM for upright cars, but not inverted cars, was significantly greater among car experts than novices (Figure 9 B).

Planned comparisons. Consistent with the above results our specific predictions were confirmed by planned 1-tailed t-tests: maximum VSTM for cars among car experts was reduced by inversion when the memory array was presented for 2500 ms, t(13)=3.47, p=.0021, or 4000 ms, t(13)=2.52, p=.0128, but not when it was presented for only 500 ms (t<1). In contrast, among car novices inversion failed to impact VSTM capacity for cars regardless of the presentation duration (all ts<1). VSTM for faces was reduced by inversion for all presentation durations among both expert and novice groups (although only marginally for 4000 ms among experts, all ps<.0519; novices, all ps<.0228). In addition, maximum VSTM for upright faces exceeded that for upright cars among car novices as long as the memory array presentation was sufficiently long (500 ms, t<1; 2500 ms, t(16)=1.27, p=.1121; 4000 ms, t(16)=2.09, p=.0265), however VSTM for cars and faces did not differ among car experts regardless of presentation duration (500 ms, t<1; 2500 ms, t(13)=1.25, p=.117; 4000 ms, t<1). Consistent with the need for sufficient encoding time in order for novices to show an advantage for faces over cars, a VSTM advantage for cars among car experts compared to car novices only emerged with sufficiently long presentation durations (≥2500 ms) (500 ms, t<1; 2500 ms, t(29)=2.42, p=.011; 4000 ms, t(29)=3.90, p=.0003).

Correlation between car expertise and maximum VSTM for cars and faces. As in Experiment 1, there was a significant correlation between K-max for upright cars when the memory array was presented for 4000 ms and participants' sensitivity for cars in the car expertise measure (i.e. car d') (r=0.701, p≤.0001). Notably, a participant's level of car expertise was not correlated with their VSTM for faces when the memory array was presented for 4000 ms (r=.202, p=.276) suggesting the correlation between car expertise and VTSM for cars cannot be attributed to a more general difference in visual skills. The correlation between VSTM for inverted cars and a participant’s car expertise was significant (r=.389, p=.031), however this correlation was carried by two outliers (greater than 2 SD above the mean) and once these participants were removed the correlation dropped considerably (r=.131, p=.497). The correlation also remained significant when the car sensitivity score was replaced with the delta d’ score (i.e. car d’ – bird d’) (r=.575, p=.0007; Figure 10 A). In addition, consistent with the previous results the correlation between the delta d’ score and face VSTM (r=.001, p=.996; Figure 10 B) or inverted car VSTM (r=.327, p=.072; with the two outliers removed, r= -.027, p=.891) was not significant.

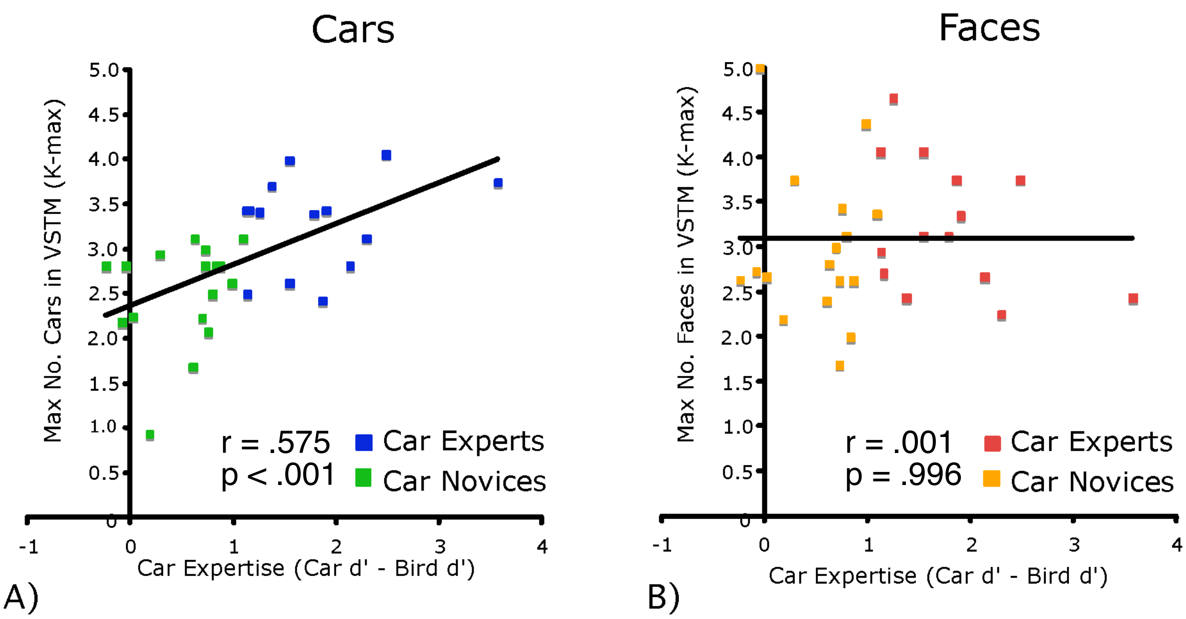


Figure 10. Scatter plot of individuals’ car expertise scores (Car d’ – Bird d’) and their K-max for (A) upright cars and (B) upright faces with a 4000 ms presentation duration. There was a significant correlation between a participant’s level of car expertise and their VSTM capacity for cars but not for faces.

Combined analysis of Experiments 4 and 5

To more directly assess the effects of the increase in articulatory suppression load on VSTM performance for faces and cars among both car experts and car novices ANOVA analyses were performed on the combined data. A 2 (orientation: upright, inverted) x 3 (presentation duration; 500, 2500, 4000 ms) x 2 (car expertise group; novice, expert) x 2 (Experiment; 4, 5) ANOVA on the maximum VSTM capacity (K-max) for upright and inverted faces found a main effects of experiment, F(1,63)=7.62, p≤.0076, with VSTM capacity generally larger in

Experiment 5 than in Experiment 4. There were also main effects of orientation, F(1,63)=95.12, p≤.0001, and presentation duration, F(2,126)=81.66, p≤.0001, but importantly there was no main effect, F(1, 63)=1.07, p=0.305, or interaction involving car expertise (all Fs<1.63, and ps>.206). There was also no significant interaction between orientation and presentation duration, F(2,126)=1.79, p=.172. Therefore, the results of the combine analysis on the K-max values for faces from Experiment 4 and 5 suggest that the increase in articulatory suppression load did not impact the qualitative pattern of results for upright and inverted faces, however there was a general increase in VSTM performance for faces in Experiment 5 relative to Experiment 4.

A 2 (orientation: upright, inverted) x 3 (presentation duration; 500, 2500, 4000 ms) x 2 (car expertise group; novice, expert) x 2 (Experiment; 4, 5) ANOVA on the maximum VSTM capacity (K-max) for upright and inverted cars found no main effect of, F(1,63)=1.07, p=.3052, or interaction with experiment (all Fs<.71, and ps>.496). There were main effects of orientation, F(1,63)=18.78, p≤.0001, and presentation duration, F(2,126)=91.67, p≤.0001, and car expertise, F(1, 63)=4.43, p=.039. There were also interactions between orientation and expertise, F(1, 63)=11.92, p=.0010, and presentation duration and expertise, F(2,126)=5.11, p=.007, with the expert VSTM advantage larger for upright items and for longer presentation durations. Therefore, the results of the combined analysis on the K-max values for cars from Experiment 4 and 5 suggests that the increase in articulatory suppression load did not impact the general pattern of results for upright and inverted cars for either car experts and novices.

Discussion

Despite the increase in the articulatory suppression load from 2-3 syllables to 5-6 syllables, an increase previously shown to significantly affect verbal short-term memory performance (Marsh & Hicks, 1998; Otani et al., 1997), the VSTM advantage for faces and other objects of expertise remained intact. In fact, the results of this experiment are similar to those obtained in Experiment 4; car experts, when given sufficient encoding time, demonstrated a greater VSTM capacity and inversion effect for cars than did car novices. Novice performance was also consistent across Experiments 4 and 5, with novices once again demonstrating greater VSTM for faces than cars when given sufficient encoding time. These results provide evidence that the greater VSTM capacity for faces among car experts does not rely on a contribution from verbal working memory.

The performance level in general was similar across Experiments 4 and 5 with the only difference being an increase in performance for upright and inverted faces, suggesting that there was no general cost to VSTM performance due to increasing the articulatory suppression load. The absence of a measurable effect of this manipulation on performance makes it difficult to ascertain if the manipulation actually impacted participants' potential to utilize verbal working memory capacity. However, a number of findings in the literature suggest that this load manipulation should have impacted verbal working memory capacity (Marsh & Hicks, 1998; Otani et al., 1997). In addition, the lack of an effect of this increase in verbal load on the VSTM task is predicted by Baddeley and Hitch's model of working memory (Baddeley, 1986; Baddeley & Hitch, 1974). Their model suggests that the only capacities shared by visual and verbal short-term memory systems are at an executive level, and it is conceivable that rehearsing two digits compared to three words would not require a discernable difference in executive resources. Recent findings are consistent with a lack of a measurable cost to executive resources under some rehearsal load conditions (Marsh & Hicks, 1998; Otani et al., 1997). Therefore, the absence of a measurable effect of the increase in articulatory suppression load on performance in the VSTM is not unexpected if participants are not utilizing verbal memory to aid their performance in the VSTM task.

Interestingly, the correlation between an observer's level of car expertise and their VSTM for cars with sufficient encoding time (4000 ms) was even stronger in this study compared to that reported in Experiment 4. It is possible that a small proportion of participants in Experiment 4 did adopt a verbal strategy and thus when this strategy was no longer available performance more strongly reflected the effects of one's visual expertise with cars.

It is important to note that 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 or not the label is explicitly accessed or used to aid recall. However, if one assumes 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, then the presence of this advantage for unfamiliar faces with no known labels provides evidence against this account. Notably, since only familiar categories were used, all stimuli could be labeled at some level. For example, even car novice could label each of the cars as 'car', just as each of the unfamiliar faces could be labeled as 'face'. However, even if one argues that the presence of the

VSTM advantage relies on the knowledge of individuating labels for the items, the presence of this effect for unfamiliar faces is inconsistent with this account.

While 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 LTM as an alternative resource. This could provide another potential explanation for their greater VSTM capacity compared to novices. Car experts may be better able to recruit or establish LTM memory representations that can contribute to performance in VSTM tasks. Experiment 6 explores this possibility.

Experiment 6

Studies of expert memory with chess experts suggest that LTM can play an important role in the apparent increase in experts' functional VSTM capacity for meaningful configurations of chess pieces (Chase & Simon, 1973; Gobet & Simon, 1998). Chess experts were found to have a superior memory for chess positions, sometimes even recalling the position of entire chess boards (around 25 pieces) from briefly (5 seconds) presented games (Chase & Simon, 1973). However, they were only able to recall around five pieces from randomly arranged boards, which was no better than novices' memory for chess pieces. Gobet and Simon (1998) suggest that chess experts use LTM to support their apparently larger functional VSTM capacity by storing pointers in VSTM to larger "chunks" of pieces stored in LTM. This account of chess experts' memory is similar to that describing expert memory more generally: Ericsson and colleagues suggested that the superior memory of experts depends on the deliberate recruitment of elaborate memory structures that allow for the chunking and indexing of information in LTM (Ericsson & Kintsch, 1995). This is proposed to allow for rapid and efficient retrieval of information from LTM in short-term memory tasks. With enough practice, functional short-term memory capacity can be increased by as much as 1000% (Ericsson et al., 1980). Therefore, the storage of larger memory chunks and the use of representations in long-term memory may allow experts to maximize the use of the limited object "slots" in VSTM by effectively increasing the amount of information that is associated with each slot (Vogel et al., 2001).

The large stimulus sets (72 items per category) used in Experiments 4 & 5 may also have encouraged participants to rely on information in LTM. Each item would have appeared infrequently (less than once in every 10 trials), and thus the relative familiarity of the probe could serve as a useful cue to aid performance. Therefore, the VSTM advantage for faces, and for cars among car experts, may rely on an interaction between LTM and expertise, with experts better able to utilize information about the familiarity of an item within the study. Such an interaction could result from experts' superior ability at distinguishing exemplars, increasing the reliability of familiarity cues. Therefore, it is possible that if these familiarity cues were no longer useful the difference between expert and novice VSTM would be eliminated.

It is important to note that an influence of LTM on expert performance, in at least some form, cannot be denied as expertise by its very nature is a product of long-term learning that likely affects many different levels of processing from perception to recognition. In addition, almost all models of working memory acknowledge that this system does not operate in isolation (Barnard, 1999; Ericsson & Delaney, 1999; Kieras, Meyer, Mueller, & Seymour, 1999; Schneider, 1999; Young & Lewis, 1999), with numerous models even suggesting that the contents of short-term memory are merely the contents of LTM currently activated (Cowan, 1999; Engle, Kane, & Tuholski, 1999; Lovett, Reder, & Lebiere, 1999; Norman, 1968; O'Reilly, Braver, & Cohen, 1999). However, an important question is whether the expert VSTM advantage demonstrated in the previous experiments in this dissertation depends on a more direct use of LTM such as described in the accounts of expert memory proposed by Ericsson and colleagues (Ericsson & Kintsch, 1995). Therefore, in addition to exploring the contribution of familiarity cues to the expert VSTM advantage, Experiment 6 also aims to address the question as the whether this advantage relies on the use of stimulus-specific representations in LTM.

If the VSTM advantage for objects of expertise relies on a difference in familiarity between foils and true probes, this advantage should disappear when a small stimulus set is used, as the build-up of proactive interference should eliminate most, if not all, of a participant's ability to rely on the relative strength of traces in LTM to distinguish foils from probes. That is, representations of all the items in the stimulus set should be similarly active in LTM, leading to confusion if one was to rely on familiarity cues from long-term memory to distinguish foils from true probes. In addition, if the VSTM advantage for faces, and cars among car experts, relies more generally on the creation of stimulus-specific LTM representations, then VSTM capacity should be reduced if the stimulus set is replaced by an entirely new set of items. These predictions are explored in Experiment 6, in which the stimulus set is reduced to just 10 items, so each stimulus would appear on average in a little more than once in every second trial, thus providing conditions for a considerable build-up of proactive interference. In addition, after participants complete two-thirds of the total number of trials the stimulus set is replaced by 10 new items to explore the extent to which VSTM performance relies on stimulus-specific representations for items in the stimulus set.

Methods

Participants. Participants were 36 employees, undergraduate students, or graduate students of Vanderbilt University, or members of the surrounding Nashville community. Participants were compensated $10 per hour for their time. All participants had normal or corrected to normal vision and reported having either extensive or very little experience identifying models of cars. All participants completed an updated version of the subordinate-level car-matching task to measure their expertise with cars. A version of the expertise measure with newer models cars (1998-2000 compared to 1995-1998 models used in the previous car expertise test) was used. This was done to better match the stimuli in the expertise test with those in the VSTM task. In every other aspect the expertise task was the same as that used for Experiment 1 and 2 in this chapter (see APPENDIX D). Pilot data showed that the two car expertise measures were highly correlated. Eighteen (13 male) participants (age M = 25.3, SD = 4.54, car d' M = 2.84, bird d' M = 1.02) met our criteria for car expertise, with the remaining 18 (8 male) classified as novices (age M = 27.2, SD = 8.36, car d' M =0.70, bird d' M = 0.83). Once again, participants' self-report of their skill at recognizing cars was generally consistent with their performance on the subordinate car matching task, with participants meeting the criteria for expertise on the task rating themselves an average of 8.42/10; those who were classified as novices rated their skills, on average, as 3.58/10.

Stimuli. The stimuli were 40 images of faces (1.9° x 2.3°) and 40 profile views of modern cars (2.3° x 1.5°). All images were grayscale, and cropped and scaled in the same way as those used in Experiments 2-5.

Design & Procedure. The design and procedure was similar to that used in Experiment 2, except the set size was held constant at 5 items, the items were always upright, and only two different presentation durations (500 ms, 4000 ms) were used. Participants performed 6 blocks of 36 trials for each category (faces, cars) resulting in a total of 432 trials. Trials for each category were performed in two separate (216 trial) sessions, with the order of the session counterbalanced across expert and novice groups. Twenty faces and 20 cars were randomly selected for each participant to appear in the VSTM task. Stimuli presented in the first 4 blocks of the VSTM task for each category were selected from a smaller set of only ten images. After performing these trials the stimulus set was changed to a new set of ten images from the same category. As a result of the smaller stimulus set each item would appear on average in every one in two trials, and thus could even occur in sequential trials. Twenty of the faces and 20 cars were used as foils in a same/different recognition task completed at the end of session 2. Participants performed this recognition task for only faces or cars depending on which session they completed last. This was done to ensure that the recognition task was a surprise. In this component, participants viewed a series of 40 faces or cars and made a key press to indicate whether or not the item appeared in the VSTM task. Each image remained on the screen until the participant made a response.

Analysis. The analysis procedure was similar to that used in Experiments 1 & 2. Data from two participants were removed from the analysis due to poor performance in the auditory task. Accuracy for this component exceeded 91% for all other participants. The data from the first two-thirds of the experiment were grouped separately from the data from the last third for which a new stimulus set was used. For each participant, hit rate and correct rejection rate were calculated for each condition in each of these bins. These values were used to derive VSTM capacity using Cowan's K, an estimate of the number of objects successfully encoded in VSTM (Cowan, 2001). Data from one participant (novice) were excluded due to poor performance in the visual working memory task (K< 1 for faces). This resulted in data from 18 novices and 18 experts being included in the final analysis. Data from one participant in the recognition memory task were not used as there was a technical failure that occurred mid-way through this component.

Results

The use of a smaller stimulus set and the change of stimulus set failed to eliminate the VSTM advantage demonstrated for objects within one's domain of expertise (Figure 11). To summarize the results, only performance among car novices in the long presentation duration condition was impacted by the change in stimulus set, with VSTM for cars in this condition dropping significantly. As in Experiment 4 & 5, VSTM capacity for faces exceeded that for cars among novices for both pre- and post-change stimulus sets, providing the presentation duration was long enough to allow sufficient encoding (4000 ms). Face and car VSTM did not differ among car experts in both pre- and post- change stimulus set conditions, regardless of presentation duration. In addition, the VSTM advantage for cars among car experts, compared to car novices, was apparent in the post-change stimulus set condition once the presentation duration was sufficiently long, although this difference failed to reach significance in the pre-change stimulus set condition. Consistent with the stronger expertise effects in the post-change stimulus set condition, the correlation between a participant's level of car expertise and their VSTM for cars when the memory array was presented for 4000 ms was only significant for the post-change stimulus set condition (Figure 11). These results were not explained by a difference in response times between upright and inverted faces and cars (see APPENDIX G). Performance in a surprise recognition test revealed better recognition memory in general for faces (Figure 12). In addition, consistent with the drop in performance for cars among novices after the stimulus set was changed, car novices had poorer memory for post-change cars, although equivalent memory for pre-change cars compared to car experts.

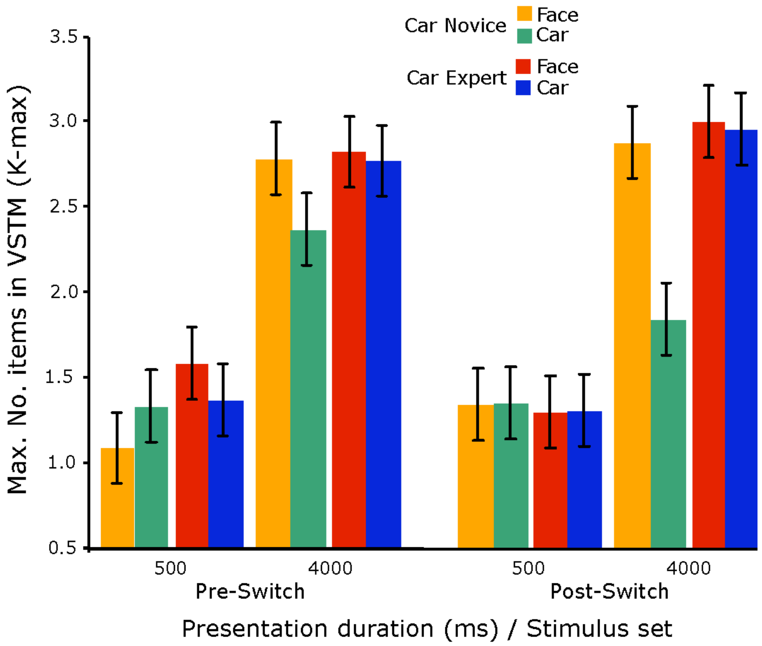


Figure 11. The maximum number of upright faces or cars (K-max) in visual short-term memory (VSTM) before and after a stimulus set change. Participants had performed 140 trials with the same small (10 item) stimulus set before the stimulus set was switched and they performed 70 additional trials. Participants were either car experts or novices and the memory array was presented for 500 ms or 4000 ms. Only novice VSTM performance with cars in the 4000 ms presentation duration condition was influenced by the change in stimulus set. Error bars represent pooled standard error values.

Comparing VSTM for faces and cars among car experts and novices. To test for an increase in performance due to experience with the images, trials from the first 2/3s of the experiment were divided into two bins. A 2 (category; faces, cars) x 2 (presentation duration; 500 ms, 4000 ms) x 2 (block; first 1/3, second 1/3) x 2 (car expertise groups; expert, novice) ANOVA found no main effect or interaction with order (all ps>.229). Data from the first two sections were combined for the subsequent analyses.

A 2 (category; faces, cars) x 2 (presentation duration; 500 ms, 4000 ms) x 2 (stimulus set; pre-change, post-change) x 2 (car expertise groups; expert, novice) ANOVA revealed a main effect of presentation duration, F(1,34)=177.87, p≤.0001, and the effect of category, F(1,34)=3.47,p=.071, and expertise, F(1,34)=3.92, p≤.056, approached significance. There was no main effect of changing the stimulus set (F<1). The interactions between expertise and category, F(1,34)=1.25, p=.272, expertise and presentation duration, F(1,34)=2.35, p=.135 were not significant. However, consistent with the predicted results there was an interaction between category and presentation duration, F(1,34)=5.20, p=.029, and category, presentation duration and expertise, F(1,34)=6.89, p=.013, with the effect of presentation duration on VSTM capacity generally greater for faces than cars, although the effect of presentation duration on car VSTM was greater among cars experts than car novices. Interestingly, there was also an interaction between presentation duration, expertise, and stimulus set, F(1,34)=4.36, p=.044, with VSTM performance at the longest presentation duration decreasing after the stimulus set was changed among novices but not experts. The interaction between category, stimulus set and expertise approached significance, F(1,34)=3.34, p=.076, with novices tending to experience a reduction in car VSTM performance while VSTM performance among car experts appeared to be unaffected by the change in stimulus set. The interaction between category, duration, expertise group, and order was not significant F<1, nor was any other interaction with stimulus set (all ps>.313).

Comparing VSTM for faces among car experts and novices. A 2 (presentation duration; 500 ms, 4000 ms) x 2 (stimulus set; pre-change, post-change) x 2 (car expertise groups; expert, novice) ANOVA revealed a main effect of presentation duration, F(1,34)=135.50, p≤.0001, but no effect of stimulus set (F<1) or car expertise, F(1,34)=1.00, p=.324. In addition, no interactions between presentation duration and/or expertise and/or stimulus set were significant (all ps>.1960).

Comparing VSTM for cars among car experts and novices. A 2 (presentation duration; 500 ms, 4000 ms) x 2 (stimulus set; pre-change, post-change) x 2 (car expertise groups; expert, novice) ANOVA revealed main effects of presentation duration, F(1,34)=73.39, p≤.0001, and car expertise, F(1,34)=4.30, p=.046, but no order (F<1). In addition, there was an interaction between presentation duration and expertise, F(1,34)=8.19, p=.0.0072, with car experts only demonstrating a VSTM advantage for cars with a sufficiently long presentation duration. The interaction between presentation duration, expertise and stimulus set, F(1,34)=3.37, p=.075, approached significance, with car novices VSTM performance dropping after the change in stimulus set while car experts performance showed a slight increase.

Planned comparisons. Planned two-tailed t-tests exploring the effect of the change of stimulus set on VSTM capacity revealed that VSTM among car experts for both cars, 500 ms, t(17)=1.32, p=.205, 4000 ms, t<1, and faces, 500 ms, t(17)=1.28 p=.218, 4000 ms, t<1, regardless of presentation duration, was unaffected by the change in stimulus set. Among novices, VSTM performance was only reduced for cars at the long presentation duration, 4000 ms, t(17)=2.16 p=.045, 500 ms, t<1, with face VSTM performance unaffected by the change in stimulus set regardless of presentation duration, 4000 ms, t<1, 500 ms, t(17)=1.21 p=.241.

In addition, planned one-tailed t-tests assessing the presence of a VSTM advantage for experts revealed that VSTM for upright faces exceeded that for upright cars among car novices providing the presentation duration was sufficiently long (4000 ms) regardless of stimulus set (pre-change, 500 ms, t(17)=1.24 p=.883; 4000 ms, t(17)=2.08, p=.027; post-change, 500 ms, t<1; 4000 ms, t(17)=3.18, p=.0027), however VSTM for cars and faces did not differ among car experts for either stimulus set regardless of presentation duration (pre-change, 500 ms, t(17)=1.32 p=.103; 4000 ms, t<1; post-change, 500 ms, t<1; 4000 ms, t<1). Consistent with the need for sufficient encoding time in order for novices to show an advantage for faces over cars, a VSTM advantage for cars among car experts compared to car novices only emerged in the 4000 ms presentation duration condition for the post-change stimulus set (500 ms, t<1; 4000 ms, t(34)=3.50, p=.0007). Surprisingly, this advantage for cars among car experts failed to reach significance regardless of presentation duration for the pre-change stimulus set (500 ms, t<1; 4000 ms, t(34)=1.29, p=.103), although there was a trend for such an effect in the long presentation duration condition.

Correlation between car expertise and VSTM for cars and faces. As in Experiments 1 & 2 there was a significant correlation between K-max for cars when the memory array was presented for 4000 ms and participants sensitivity for cars in the car expertise measure, but only for the post-change stimulus set (pre-change, r=.226, p=.185, post-change, r=.509, p=.0015). Notably, participants' level of car expertise was not correlated with their VSTM for faces for either the pre- (pre-change, r=.014, p=.935, post-change, r=.053, p=.760) stimulus sets. The general pattern of results was the same for correlations between the delta d' score (i.e. car d' – bird d') and K-max with a 4000 ms presentation duration for cars (pre-change, r=.202, p=.236, post-change, r=.452, p=.006) (Figure 12 A and C) and faces (r=.028, p=.870) or post- change (r=.074, p=.667) (Figure 12 B and D).

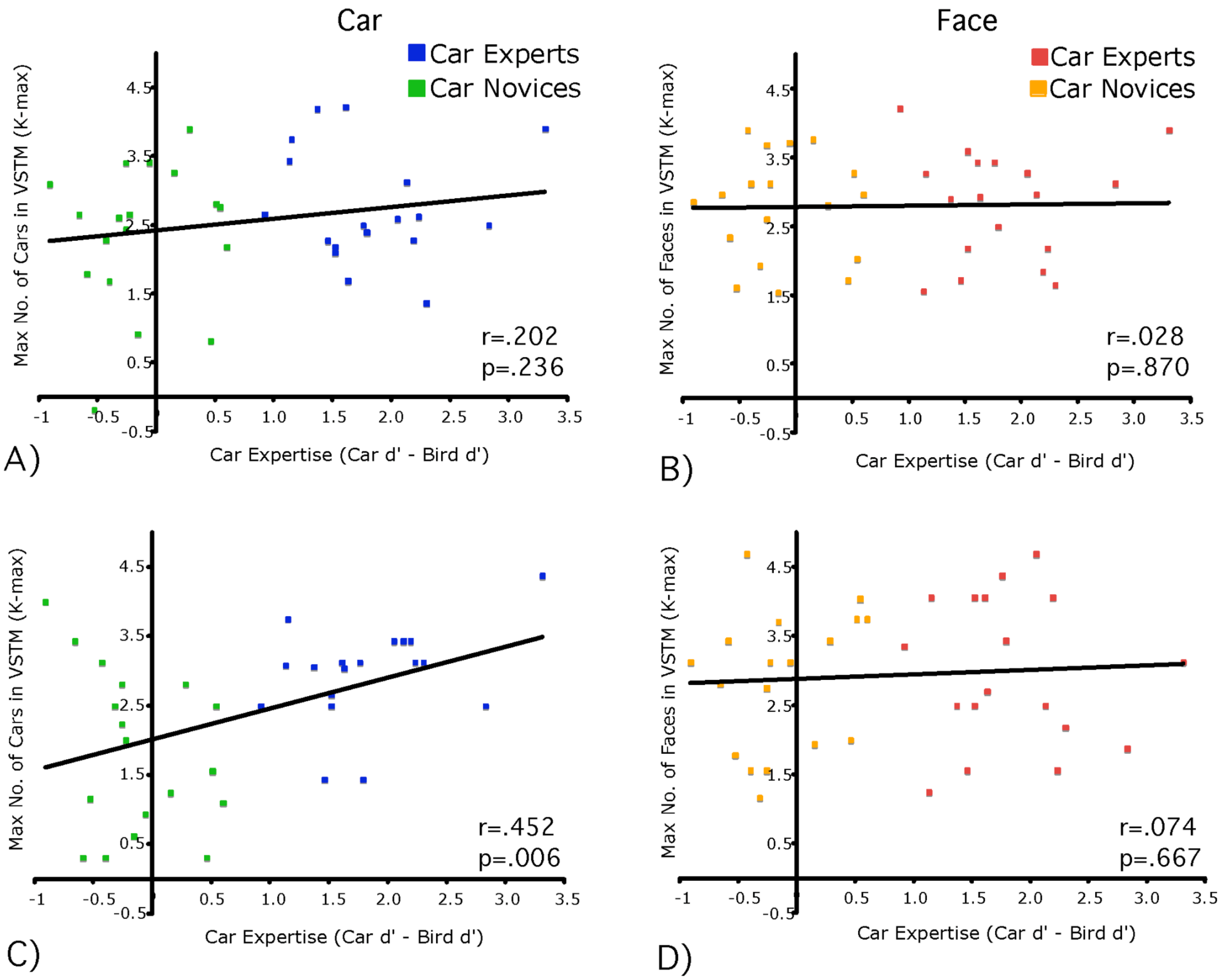


Figure 12. Scatter plot of individuals' car expertise scores (Car d' – Bird d') and their K-max for cars (A) pre- and (C) post- stimulus set switch and for faces (B) pre- and (D) post stimulus set switch with a 4000 ms presentation duration. There was a significant correlation between a participant's level of car expertise and their VSTM capacity for cars in the pre- but not post-stimulus switch condition. VSTM with faces was not correlated with level of car expertise, regardless of condition.

Recognition memory performance. A 2 (Category; faces, cars) x 2 (stimulus set; old, new) x 2 (car expertise group; expert, novice) ANOVA found main effects of category, F(1,31)=4.59, p=.040, but not stimulus set (F<1) or expertise (F<1). There was also an interaction between category and expertise, F(1,31)=5.05, p=.032, with Scheffé tests revealing that car novices have better memory for faces than cars, while car experts' memory did not differ across the categories. There was an interaction between category, expertise group, and stimulus set, F(1,31)=6.56, p=.016. Scheffé tests revealed that car novices, relative to car experts, had worse memory for the post-change cars (p=.001) but equivalent memory for the pre-change cars (p=.190), while they had better memory for the post-change faces (p=.0001) but equivalent memory for the pre-change faces (p=.159), see Figure 13.

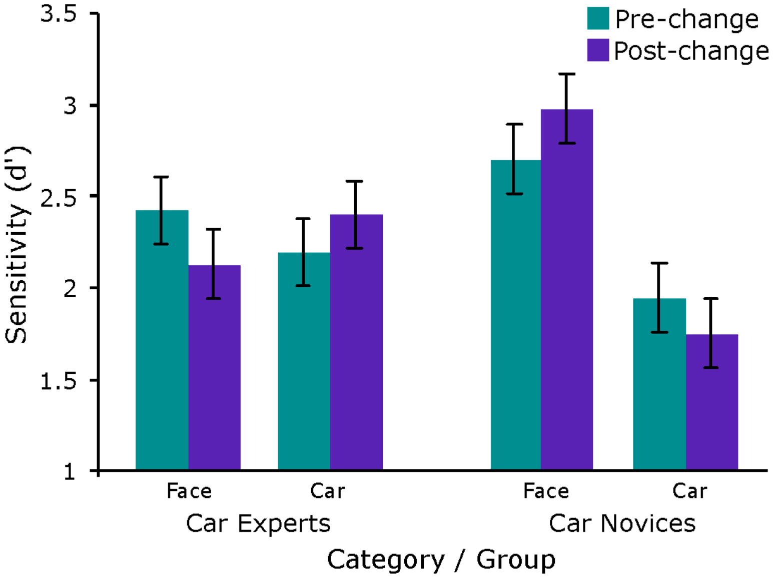


Figure 13. Recognition performance for cars and faces in the pre- and post- change stimulus sets among car experts and novices in the surprise old/new recognition task. There was a general advantage for faces. Car novices had worse memory for post-change cars, but better memory for post-change faces compared to experts. However, car experts and novices had equivalent memory for faces and cars in the pre-change stimulus set. Error bars represent pooled standard error values.

Discussion

In contrast with suggestions that LTM contributes heavily to expert VSTM performance, the expertise advantage was not eliminated by the build-up of proactive interference expected from the use of a small stimulus set. This suggests that the expertise advantage does not rely on the more efficient use of familiarity information present when study items repeat infrequently. Further support for the limited role of LTM in contributing to the expert VSTM advantage is the absence of any detectable cost to expert performance when the stimulus set is replaced by an entirely different set of items after a period of learning. VSTM for cars among car novices at the longest presentation duration was the only condition to be significantly impacted by the change in stimulus set. Therefore, these results suggest that the expert VSTM advantage does not depend on a contribution from LTM.

Alternatively, it is possible that car experts do rely on LTM to support their VSTM advantage for objects within their domain of expertise, but they were not affected by the change in stimulus set in Experiment 6 because they already had establish LTM representations for the cars in both the pre- and post-change stimulus sets. In contrast, novices—who were presumably less familiar with the cars—were influenced by the additional exposure to the cars in the pre-change set as it provided a greater opportunity to establish and rely on representations in LTM. Therefore, the drop in VSTM among novices reflects the extent to which the novices were relying on representation created during the first section of the experiment. However, this account would have predicted that VSTM for the unfamiliar faces would have incurred a cost due to the change in stimulus set, as participants could not have had pre-existing representation of these faces in LTM. This explanation would have also predicted that novice's VSTM for cars, and also VSTM for the unfamiliar faces more generally, would have increased from the first to the second half of the pre-change trials, which it did not. Therefore, the use of pre-existing representations in LTM by experts cannot account for the pattern of results found in Experiment 6.

Another possibility is that experts can create LTM representations more rapidly than novices due to a more established representational space for exemplars within their domain of expertise (Palmeri, Wong, & Gauthier, 2004). More specifically, experts may be able to more optimally map objects onto learned categories or identities due to their extensive experience. If experts can rapidly establish LTM representations for new items, any cost due to the change in stimulus set may be minimal and washed out in the numerous trials (~30 per set size) required to calculate a reliable estimate of VSTM capacity. In contrast, if novices develop LTM representations more slowly, they would demonstrate a larger cost due to the change in stimulus set. This possibility is difficult to address due to the limited frequency at which VSTM performance can be sampled reliably, making the monitoring of changes in VSTM performance on a finer temporal scale almost impossible with this paradigm. However contrary to novices' actual performance, this account would have predicted that their performance would have increased during the pre-change stimulus set condition. In addition, performance in the old/new recognition task also is inconsistent with this account of the results; memory for cars failed to differ across the pre-change and post-change stimulus sets for either the expert or novice groups. This suggests that if there is a difference in the speed with which the two groups can establish LTM representations, it cannot account for the drop in performance among car novices for the post-change stimulus set.

Despite the drop in VSTM performance for cars among car novices with the change in stimulus set, it is questionable whether this reflects a use of LTM representations to increase VSTM performance in the pre-change condition by novices. As mentioned above, participants failed to show the increase in VSTM performance after exposure to the items that would be expected if LTM representations were used to facilitate VSTM performance: Performance during the first 72 trials in the study, during which each of the 10 items appeared approximately 40 times, did not differ from that during the second set of 72 trials with the same stimulus set. Instead, this cost to performance with the change in stimulus set might reflect interference from the previous set of stimuli. More specifically, consistent with the characteristic processing style of novices discussed in CHAPTER I, the cost incurred to novice VSTM performance may reflect the recruitment of a rigid feature-based strategy in which the cars are identified by a single salient feature. The salient feature chosen is likely to be task- and stimulus-specific, with the salience and relative usefulness of a feature as a memory cue critically depending on the features of the other items in the stimulus set. Such a strategy could be quite effective for a small specific stimulus set. However, a stimulus specific strategy would presumably show very poor transfer to a new stimulus set, thus providing a potential explanation for the cost to novice performance with the change in stimulus set. Costs due to similarly inflexible memory representations have been documented in the literature (Finke, 1989; Reed, 1974). In contrast, a more holistic strategy believed to be recruited for faces by both experts and 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 was changed may reflect a cost due to interference rather than the elimination of facilitation from their experience with the items.

Consistent with the benefits of a small stimulus set for novices, when the memory array was presented for 4000 ms VSTM for cars among experts and novices in the pre-change stimulus set did not differ significantly, nor was there a correlation between car expertise and VSTM capacity. These results are in contrast with those from Experiment 4 and 5 in which a considerably larger stimulus set was used. Therefore, the presence of an expert VSTM advantage may depend on a sufficiently large stimulus set and/or minimal exposure to the items within a set, thereby preventing novices from using stimulus-specific compensatory strategies to increase their VSTM performance.

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). In some cases, the most recently stored item can be distinguished based on its temporal context. It is unlikely that this temporal information would be sensitive enough to be reliable under conditions such as those in Experiment 6; items frequently appeared in consecutive trials (a little more than a few seconds apart at times) and subjects performed 144 of such trials 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, while the digits typically used in Ericsson's studies can be easily encoded as a running time, a zip code, or a birthday, such reliable alternative encoding strategies that could differentiate unfamiliar faces at the individual level are unlikely to be as easy to implement. Presumably this would be especially true under conditions such as those in Experiment 6 where each stimulus repeats between 40 – 80 times, and thus a large number of different reliable alternative encoding strategies would have to be generated in order to avoid proactive interference. Therefore, it is unlikely that participants could have used the strategies suggested by Ericsson and Kintsch (1995) to successfully overcome the effects of the build-up of proactive interference in Experiment 6.

Experiments 5 & 6 have addressed the potential role of a contribution from other complex cognitive systems such as verbal or long-term memory, but have overlooked another alternative account for the differences in VSTM performance for expert and novice categories and for the greater benefit of additional encoding time for items within ones domain of expertise. A difference in the physical strategy (e.g., the pattern of eye-movements) employed to gather the information within a study array could potentially account for the performance of experts and novices and also the influence of encoding time on VSTM performance. Experiment 7 explores this possibility.

Experiment 7

It is possible that the VSTM advantage for categories of expertise may rely on strategic differences unrelated to holistic processing, such as a difference in the eye movements strategies employed by expert and novice observers. For example, a non-expert observer may make a rapid series of fixations to each of the stimuli in the study array and continue to repeat this sequences while the study array remains present. This might be a strategy adopted when viewing multiple objects that are difficult to distinguish, as it is likely to facilitate the detection of differences between the stimuli. In contrast, the items may be easier to distinguish for an expert observer and thus they may view the items in the array more deliberately, maintaining fixation on a particular item until enough information is extracted for a complete representation. If true, this difference in eye-movement strategies would provide an alternative account for the data in Experiments 4 and 5: non-expert items would be advantaged at short presentation durations because more items could be viewed in the limited time. The representation of these rapidly viewed items might be less complete or robust compared to the representations created with longer viewing fixations of fewer items by expert observers. Longer presentation durations may allow expert observers to reap the benefits of longer fixations times, leading to an advantage for expert-category items when sufficient encoding time is allowed. The trade-off between fixation duration and the number of items viewed may favor short fixation durations for shorter presentations but longer fixation durations when encoding time is less limited.

If this is the case, the advantage shown for faces should disappear when participants who are car novices are required to adopt the same pattern of eye movements for faces and cars. This prediction is tested in Experiment 7 by comparing VSTM for study arrays of faces and cars presented either simultaneously or sequentially. In addition to exploring the trade-off between fixation length and number of fixations for a given encoding time, two different sequential presentation conditions were also compared; items in the study arrays were either encoded during one or two separate fixations with the total presentation time held constant.

Methods

Participants. Participants were 29 (16 female) employees, undergraduate students, or graduate students of Vanderbilt University, or members of the surrounding Nashville community. All participants had normal or corrected to normal vision (age M = 23.48, SD = 3.84). Participants were compensated $10 per hour for their time. To ensure all subjects were novices, participants' expertise with cars was measured as in Experiments 1. Two participants were excluded as they met the criteria for car expertise.

Stimuli. Stimuli were 10 grayscale faces (1.9° x 2.3°) from the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany and ten profile views of modern cars (2.3° x 1.5°).

Design & Procedure. The design and procedure were similar to that used in Experiment 6. Once again the set size was held constant at 5 items and only upright faces and cars were displayed. Unlike Experiment 6, three different presentation durations (1000 ms, 2000 ms, 4000 ms) were used. Participants performed 7 blocks of 36 trials for each category (faces, cars) resulting in a total of 504 trials. Trials for each category were performed in two separate (252 trial) sessions, with the order of the sessions counterbalanced across subjects. The study items were presented either simultaneously or sequentially. If the items appeared sequentially, the items would appear in the same locations as in the simultaneous presentation conditions starting from the lower right location. Participants were informed of this order during the instruction period, and were told to move their eyes to each item as it appeared and to maintain their gaze there until the next item appeared (between 200 ms – 800 ms). In the trials where the items appeared sequentially, this sequence was either presented once *('Sequential 1'*) or twice *('Sequential 2'*), with the total presentation time for the study array being equivalent to the presentation durations in the simultaneous conditions (i.e. 1000 ms, 2000 ms, or 4000 ms). For example if the total presentation duration was 4000 ms, in the 'Sequential 1' condition the presentation interval for each item was 800 ms, whereas in the 'Sequential 2' condition each items appeared for two 400 ms intervals. Due to the fundamental limitations of the visual system in terms of the time required to make a saccade, the 'Sequential 2' condition with a total duration of 1000 ms (100 ms/item) was not included in the design.

Analysis. The analysis procedure was the same as that used in Experiment 1-6, where the hit rate and correct rejection rate were calculated for each condition to derive VSTM capacity using Cowan's K. Data from three participants were removed from the analysis due to generally poor performance for faces in the VSTM task (average K<1). Accuracy for the auditory component exceeded 95% for all remaining participants.

Results

To summarize the results, the VSTM advantage for faces over cars among car novices was not affected by restricting participants eye-movements, as participants still demonstrated a VSTM advantage for faces, providing there was sufficient encoding time regardless of whether they viewed the items simultaneously, and thus were free to adopt any eye-movement strategy, or if the items were viewed sequentially in a fixed order and for a fixed amount of time (Figure 14). However, there was a general reduction in VSTM performance when participants viewed the items over two short fixations compared to one longer fixation equivalent in length to the sum of the two presentations. These results were not explained by a difference in response times between upright and inverted faces and cars (see APPENDIX H).

To provide a stringent test of whether presentation format impacted the VSTM advantage for faces, the data from each presentation format condition were first analyzed separately in separate 2 (category; faces, cars) x 3 (presentation duration; 1000 ms, 2000 ms, 4000 ms) ANOVAs for the simultaneous and Sequential 1 conditions, and in a 2 (category; faces, cars) x 2 (presentation duration; 2000 ms, 4000 ms) for the Sequential 2 condition. As expected from the results of Experiment 1 & 2, there was a main effect of category, F(1,28)=8.12, p=.0081, and presentation duration, F(2,56)=23.54, p≤.0001, and an interaction between category and presentation duration, F(2,56)=4.17, p=.021, for trials in which the study array was presented simultaneously. Planned paired t-tests found no advantage for faces with 1000 ms (t<l), but an advantage with 2000 ms, t(28)=3.95, p=.025, and 4000 ms encoding time, t(28)=2.42, p=.022. Similarly, when the study array was presented sequentially just once, there was also a main effect of category, F(1,28)=4.40, p=.045, and presentation duration, F(2,56)=24.35, p≤.0001, and an interaction between category and presentation duration, F(2,56)=4.13, p=.021. Planned paired t-tests found no advantage for faces with 1000 ms, t(28)=1.39, p=.018, but an advantage with 2000 ms, t(28)=2.4, p=.023, and 4000 ms encoding time, t(28)=2.31, p=.0283. When presentation of the study array was divided into two sequential loops, there was still a main effect of presentation duration, F(1,28)=28.89, p≤.0001, however the main effect of category, F(1,28)=3.14, p=.087, and interaction between category and presentation duration,

F(1,28)=3.15, p=0.087, approached significance. Planned paired t-tests found no advantage for faces with 2000 ms (t<1), but an advantage in the 4000 ms presentation duration condition, t(28)=2.36, p=.0253. The weakening of these effects for trials in which the study array presentation was divided over two loops may be due to the general drop in performance for both cars and faces in this condition.

To explore possible interactions with presentation format, a 2(category; faces, cars) x 2 (presentation duration; 2000 ms, 4000 ms) x 2 (presentation format: simultaneous, sequential 1) was conducted. The sequential 2 condition was not included in this analysis as every condition was not represented in this presentation format. Main effects of category, F(1,28)=10.97, p=.0026, presentation duration, F(2,56)=57.94, p≤.0001, and presentation format F(1,28)=7.12, p=.0125, were observed, with better VSTM performance for faces, longer total presentation durations, and for the simultaneous presentation format. There was no interaction between presentation format and category, F(1,28)=1.98, p=.170, or presentation duration (F<1). Importantly, there was an interaction between category and presentation duration, F(2,56)=6.86, p=.0022, although this did not interact with presentation format (F<1) suggesting that VSTM for faces was greater than that for cars regardless of presentation format providing there was sufficient encoding time.

To more directly examine a potential trade-off between the number of fixations and the length of fixation an ANOVA comparing the sequential 1 and sequential 2 presentation formats (the 1000 ms presentation duration condition was not included in this analysis as there was not a corresponding presentation duration in the sequential 2 condition) found main effects of category, F(1,28)=14.11, p=.0008, presentation duration, F(1,28)=42.11, p≤.000, with better VSTM performance for faces and for longer total presentation durations. There was also a main effect of presentation format, F(1,28)=11.51, p=.0021, with a single longer fixation resulting in better performance than two shorter durations this appeared to be especially true with shorter total presentation duration although the interaction between total presentation duration and presentation format did not reach significance, F(1,28)=2.45, p=.129.

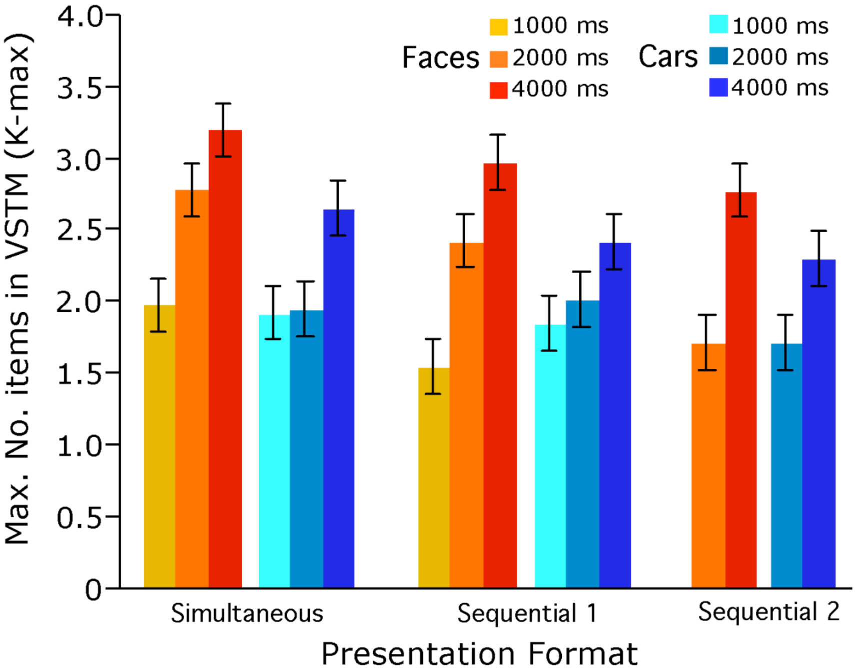


Figure 14. The maximum number of faces or cars (K-max) in visual short-term memory (VSTM) for the 1000 ms, 2000 ms, or 4000 ms presentation duration conditions. The study arrays were either presented simultaneously, in a sequential order once (sequential 1), or twice in the same sequential order (sequential 2). The VSTM advantage for faces was present regardless of whether the items were presented simultaneous or sequentially, however a single longer presentation generally facilitated better VSTM performance than two shorter fixations. Error bars represent pooled standard error values.

Discussion

The finding of a VSTM capacity advantage for faces compared to cars among car novices, regardless of whether the study array was presented simultaneously or sequentially, suggests that the VSTM advantage for faces and other expert categories does not rely on a difference in the pattern of eye movements employed during the encoding of the study array. However, shortening the total presentation duration by requiring participants to make an additional set of saccades to each item did result in a general cost to performance. This result suggests that if given the same total viewing time, an eye movement strategy that favors longer, but fewer, fixations would result in better VSTM performance providing the fixations were sufficiently short to still allow each object in the study array to be fixated.

The approximately equivalent performance in the ‘sequential 1’ and simultaneous presentation conditions suggests that participants gain very little, if any, benefit from the potential to encode multiple items in parallel. This is consistent with Rousselet and colleagues account of object encoding: Parallel processing occurs for a very brief amount of time after a stimulus appears to provide a crude analysis of the visual scene, but more detailed object processing occurs in a relatively serial fashion (Rousselet, Thorpe, & Fabre-Thorpe, 2004). The weak trend of a general cost in the 'sequential 1' compared to the simultaneous condition, may be due to a cost incurred by the inability to divide the encoding time to favor more difficult or less distinctive items that may require additional encoding time. Therefore, these results support object recognition models suggesting that to extract the fine-level information required for identity judgments, objects must be encoded in serial.

General Discussion

The four experiments in this chapter establish the presence of a VSTM advantage for cars among car experts, compared to car novices. This advantage is remarkably similar to that demonstrated for faces in CHAPTER II, and which is also replicated in this chapter; this advantage requires sufficient encoding time, is orientation-specific, and is similar in magnitude to the VSTM advantage for faces. In the context of the literature linking face perception with expert visual processing more generally, these findings are consistent with a general perceptual expertise mechanism underlying the VSTM advantage for faces. This advantage was not eliminated by the introduction of a verbal memory load previously demonstrated to have a significant impact on verbal short-term memory performance, suggesting that it does not rely on a contribution from verbal working memory (Experiment 5). Nor was it eliminated by the use of a small stimulus set, which increased the potential for proactive interference, or a surprise change in stimulus set to eliminate advantages due to previous exposure to the images (Experiment 6). Finally, the VSTM advantage for objects within ones domain of expertise, in this case faces, was also still present when participants were required to adopt a restricted pattern of eye movements to encode the items in memory (Experiment 7). Therefore, this effect is robust and evidence suggests that it does not depend on the recruitment of additional capacity from other memory systems such as verbal or long-term memory.

The correlation between VSTM for cars and sensitivity on an established measure of visual expertise with cars is also consistent with a perceptual locus of the superior VSTM of car experts (Gauthier et al., 2003). In addition, this correlation provides indirect evidence that the locus of this effect may lie more specifically with the specialized holistic processing strategy recruited for faces and other objects within one's domain of expertise: this behavioral measure of perceptual expertise with cars is correlated with measures of holistic processing of cars (Gauthier et al., 2003). This measure is also correlated with the level of activity for cars in functionally defined face-selective cortex in the fusiform gyrus in the temporal lobe (Gauthier et al., 2000). Activation in these areas has also been linked to holistic processing measures (Gauthier & Tarr, 2002). In addition, the orientation specific nature of the expert VSTM advantage is also consistent with a contribution from holistic processing mechanisms: the inversion effect for faces is a result of the disruption of access to configural information critical for holistic processing (Bartlett & Searcy, 1993; Bruce et al., 1991; Collishaw & Hole, 2002; Kemp et al., 1990; Leder & Bruce, 1998, 2000; Leder et al., 2001; Murray et al., 2000; Rhodes et al., 1993; Searcy & Bartlett, 1996; Tanaka & Sengco, 1997; Thompson, 1980). Therefore, many aspects of the VSTM advantage for objects from one's domain of expertise are consistent with the locus of this effect being with the holistic encoding strategy recruited by visual experts.

The influence of specialized expert encoding mechanisms may also impact VSTM capacity by mediating which information gets retained (Olson et al., 2005). Alvarez & Cavanagh (2004) suggest that there is a core set of features that form the minimal representation of object identity, and which are always encoded regardless of task demands. Presumably, this core set of features would depend on the level at which an object is first identified, for example, a vehicle can be identified as a car, a sedan or a Honda Accord 2002. The level at which one first identifies an object shifts with experience from a basic level (e.g. car) to a subordinate level (e.g. Honda Accord) (Tanaka & Taylor, 1991). Conceivably, the set of core features encoded may change depending on one's experience. For example, a person who possesses expertise recognizing different model cars may have a core set of features that includes more diagnostic information than that of someone who has very little experience identifying cars. Therefore, expert representations for objects could support more reliable recognition than the less diagnostic novice representations, providing a potential account for the difference in VSTM capacity between novices and experts. Future studies should probe the nature of the representations created by experts and novices by systematically manipulating different dimensions of the target and probe items.

In addition, holistic encoding may also result in the tighter binding of the information in an object representation, potentially impacting VSTM in a number of ways. Holistic representations may serve to better "chunk" object features together at a perceptual level and thus reduce the number of information units that must be stored for a given object. The differential impact of a face-like configuration on feature and conjunctive target searches is consistent with this possibility; when observers searched through triplets of up or down arcs organized either to form schematic faces or meaningless patterns, the face-like configuration facilitated the conjunction search (Suzuki & Cavanagh, 1995). This advantage for conjunctive searches was only present for upright, but not inverted face configurations. Notably, the face configuration slowed search in the feature search condition. Therefore, holistic processing appears to group the constituent features into a unitized representation facilitating their processing as a unit, but rendering the individual features less accessible to rapid parallel search mechanisms (Treisman, 1982; Treisman & Gelade, 1980).

Holistic encoding may also impact VSTM by reducing the impact of limitations in attentional mechanisms, required to bind information together within an object, on VSTM capacity. Feature-based theories of attention suggest that VSTM is limited not only by the capacity of independent feature stores, but also by the capacity of attentional mechanisms required to maintain the binding between these features. For example, a holistic representation of the shape of a face may fill less shape feature 'slots' than a more piecemeal representation of the same shape information, such as that which might be used by a novice observer. Holistic representations may also require little or no attentional resources to maintain binding information compared to representations in which the features are represented more independently. Consistent with this hypothesis, holistic processing has been reported to occur equally well with and without attention (Boutet, Gentes-Hawn, & Chaudhuri, 2002), however there may be a limit to the number of faces that can be holistically processed simultaneously (Palermo & Rhodes, 2002). Therefore, VSTM for faces and other objects of expertise might be less susceptible to attentional capacity limits rendering performance less prone to errors due to mismatching of features across items.

It is important to note that the proposed advantage due to holistic processing is unlike that suggested to support the greater functional VSTM capacity of chess experts for meaningful chess configurations. An important difference between the "chunking" of information and holistic processing is that the former operates over multiples objects (i.e. chess pieces) while the later operates within an item (e.g. a car). Similarly, Ericsson's digit span experts develop "chunking" strategies that bind *multiple* digits together. Importantly, there is no evidence that participants are chunking different faces together in memory, as this would be expected to result in a large increase in VSTM capacity in the magnitude of that demonstrated by chess experts or digit span experts. In addition, the items within chunks require 'semantic glue' to hold them together, for example one of Ericsson's digit span experts utilizes semantic knowledge about running times to chunk digits together (Ericsson et al., 1980). Consistent with the need for semantic knowledge in order to chunk items together in LTM, chess experts VSTM advantage breaks down for meaningless chess configurations (Chase & Simon, 1973). In contrast, the holistic binding that occurs within objects of expertise operates on unfamiliar exemplars, presumably with no or at most very few semantic associations, providing they fall within the domain of one's expertise. Therefore, holistic processing, unlike the chunking used by Ericsson's memory experts, operates more at a perceptual level rather than a semantic level and has a far more restricted range, operating within and not between objects.

The difference in magnitude of the VSTM advantage demonstrated for faces, and cars among car experts, compared to that exhibited by Ericsson and colleagues'(Ericsson et al., 1980) memory experts is one of the most convincing piece of evidence suggesting that these two effects of expertise are qualitatively different. More specifically, while visual expertise increased VSTM for faces and cars among car experts by about 20-25%, Ericsson's memory experts increased their functional VSTM capacity by 1000% (Ericsson et al., 1980). Notably, the VSTM capacity of visual experts remains within the range of typical VSTM limits for simple objects (Vogel et al., 2001). If car experts were recruiting LTM resources, Ericsson's account would predict that their capacity should be far greater, easily exceeding normal VSTM capacity limits. Therefore, I suggest that the VSTM advantage demonstrated for items within one's domain of visual expertise reflect the impact of expert *perceptual* processes on VSTM. In contrast, the VSTM advantage reported by Ericsson reflects the use of deliberate and explicit strategies relying on *semantic* information to chunk and LTM to store, what would otherwise be, independent objects together (Ericsson & Kintsch, 1995).

An interesting question is why the advantage for faces in VSTM capacity only appears at presentation durations longer than 500 ms, given that differences between the neural response to upright faces and inverted faces or objects occur as early as 170 ms after presentation (see Rossion & Gauthier, 2002 for a review). One possibility is that holistic processing already has an effect at 500 ms, but that its contribution cannot overcome perceptual limitations under these conditions (Eng et al., in press). Alternatively, because consolidation into VSTM is capacity limited, it may not have sufficient time to store more objects of expertise even if more are perceived (Jolicoeur & Dell' Acqua, 1998). Thus, the advantage for upright faces, and upright cars among car experts, may only appear when there is sufficient time to complete consolidation. Estimates of consolidation time have been inferred to be as large as 500 ms per item (Chun & Potter, 1995; Jolicoeur & Dell' Acqua, 1998; Ward, Duncan, & Shapiro, 1996), although it could be as little as 50 ms for very simple objects (Vogel, Woodman, & Luck, in press). It is unknown how perceptual processing efficiency influences consolidation mechanisms—this would be a worthy topic for future research. However, a first step to better understand the results reported in this chapter would be to explore the temporal dynamics of holistic processing as it is possible that holistic processing may follow a different time-course than that of more feature-based encoding strategies. CHAPTER IV will explore this issue.

CHAPTER IV

EXPLORING THE TIME-COURSE OF ENCODING UPRIGHT AND INVERTED FACES AND UPRIGHT CARS AMONG CAR EXPERTS AND NOVICES

Introduction

The capacity of information processing is limited at several stages along the visual processing pathway. For instance, we are limited in the amount of information we can select for processing, the speed with which we can process and respond to this information, and the amount of information we can maintain in conscious awareness without external stimulation. A processing limitation at any one of these stages could lead to an apparent capacity difference at later stages. Capacity limitations at a perceptual level may be especially important, as limitations at this level could potentially have a cascading effect on functions in many other systems, such as VSTM. For example, limitations at a perceptual level could limit both the amount of information that is processed in parallel and also the rate at which the information is processed.

The studies reported in CHAPTERS II and III focused on VSTM capacity—the limitation of maintaining visual information in awareness in the absence of external input. The results from these chapters suggest that VSTM storage is influenced by the manner in which an item is perceptually encoded earlier in the visual information processing pathway. More specifically, the studies in CHAPTERS II and III demonstrated that a VSTM advantage exists for objects within one's domain of expertise. These studies provide consistent, but indirect, evidence that the specialized holistic encoding style recruited for these objects may benefit VSTM capacity.

Crucially, it appears that the VSTM advantage for objects of expertise emerges only with a sufficient encoding duration. An open question is whether the different processing strategies recruited by expert and novice observers result in a difference in the time-course of encoding. More specifically, do holistic and more feature-based encoding strategies require different amounts of time? One may expect expert encoding to be faster, given the superior performance of experts compared to novices under a wide range of conditions. However, the requirement of sufficient encoding time to demonstrate an expert VSTM advantage might be indicative of a more capacity demanding perceptual processing strategy for objects of expertise, and thus, counter-intuitively, encoding by experts may be slower than encoding by novices. A third alternative could be that the rate of encoding does not differ between experts and novices. This would suggest that the additional encoding time required by experts is merely a consequence of their ability to store more objects in VSTM. That is, if experts and novices have a similar rate of encoding, experts would need more encoding time to reach their VSTM capacity. If a differential time-course exists for expert and novice encoding, it could provide insight into the underlying cause for the dependence of the expert VSTM advantage on sufficiently long encoding time and also the nature of expert processing more generally.

This chapter first outlines the current state of knowledge about the time-course of processing for face and non-face objects and the factors that influence it. After the review of this literature, the results of two experiments comparing the time-course of expert and novice object encoding and recognition are reported. The results of these studies will be discussed in terms of the possible implications for our understanding of the temporal dynamics of expert perceptual processing.

Time-course of object and face processing

For the purpose of this dissertation, the term "*time-course of processing*" refers to the temporal dynamics of object processing from the moment the stimulus is presented to an individual to the point where recognition performance asymptotes. By its nature, this time-course incorporates both the time necessary for perceptual encoding and the time necessary for non-perceptual mechanisms underlying recognition performance (e.g. semantic processing). If one were interested in post-perceptual mechanisms, an appropriate design might be to keep stimulus presentation durations constant while varying the time limit within which participants need to respond. However, the purpose of this chapter is to explore the early encoding phase of this processing time-course. Thus, the current design will manipulate stimulus presentation duration rates. Since we do not have a behavioral performance measure that would give us direct access to this encoding stage, inferences will instead be made based on the impact of different encoding manipulations on subsequent object recognition performance. For the purposes of this dissertation**,** the "onset time" of recognition is the encoding time required before recognition performance exceeds chance-level. More specifically, this point reflects the minimum encoding time to detect a behavioral marker of successful encoding of an object. Another important aspect of the time-course of processing is the rate at which performance approaches asymptote. This rate can be inferred to reveal differences in the rate at which task-relevant information is accumulated. A difference in rate may reflect either more efficient use of perceptual information or a faster rate of encoding.

Factors that can influence the time-course of object processing

Visual categorization tasks that require different levels of specificity have been reported to follow different time-courses (Grill-Spector & Kanwisher, 2005; Jolicoeur, Gluck, & Kosslyn, 1984; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Tanaka, Luu, Weisbrod, & Kiefer, 1999). However, although one might expect that more specific categorization judgments require more time than do less specific categorization judgments, the speed with which an item can be identified cannot be predicted based only the specificity of the judgment to be made; objects are identified first at a particular level of abstraction which is typically neither the most general nor the most specific possible (Rosch et al., 1976). For example, Rosch et al. (1976) reported that observers are faster to label an image as a 'dog' (basic-level), than to say more specifically that it is a 'beagle' (subordinate-level), or more generally that it is an 'animal' (superordinate-level). This categorization level is referred to as the *basic-level*, and reflects the level at which most knowledge is organized (Rosch et al., 1976).

According to Rosch et al. (1976), the basic-level is not arbitrary, but instead is strongly influenced by the structural properties of the stimulus. The basic-level is defined as the level of categorization at which a single mental image can be formed, category members share similar shape, and similar motor actions are used to interact with category members (Rosch et al., 1976). It has been proposed that these basic-level representations have a privileged status being the point of first contact for perceptual categorization judgments (Jolicoeur et al., 1984). Empirically, this point is referred to as the "entry-point" of recognition, as it appears to be where the perceptual stimulus first makes contact with representations in memory (Jolicoeur et al., 1984).

The slower categorization of objects at more specific (subordinate) and at more general (superordinate) levels than the basic-level appears to reflect limitations in different processing mechanisms or stages. The response time of participants to provide more superordinate labels than the "entry point" (e.g. 'animal') was related to the degree of association between the "entry point" concept and the superordinate label. This was demonstrated by showing that the time to provide a superordinate label after reading a basic level word was positively correlated with the time to provide the same superordinate-category label for the corresponding picture (Jolicoeur et al., 1984). Therefore, the longer processing time for superordinate categorizations appears to reflect limitations at a semantic, rather than perceptual, level. Consistent with the suggestion that the slower response times for superordinate categorizations, relative to basic-level categorizations, was not due to additional perceptual processing, manipulating the presentation duration had no effect on the time difference between verifying a basic-level and a superordinate-level categorization of a picture (Jolicoeur et al., 1984). In contrast, providing more subordinate labels (e.g. beagle) than that corresponding to the entry point of categorization was affected by encoding duration limitations (Jolicoeur et al., 1984). Thus, the slower response times for subordinate-level categorizations appear to reflect limitations at a perceptual level, as such categorizations require more detailed perceptual analysis. Therefore, the time-course of object encoding is influenced by the increasing perceptual demands that are associated with increasingly specific categorization judgments. In contrast, the time-course of later post-encoding stages of object processing, especially for superordinate judgments, appear to be influenced by the accessibility of semantic information.

Neurophysiological evidence is consistent both with the suggestion of a different underlying cause for the increased response time for categorizations at subordinate and superordinate levels and with the greater perceptual demands of subordinate level categorizations, relative to more general categorizations. A recent study that recorded electrophysiological responses while participants made visual categorization judgments at different levels of specificity reported that superordinate-level categorizations, compared to basic-level categorizations, were associated with increased amplitude of relatively late frontal potentials associated with the retrieval of semantic information (Tanaka et al., 1999). In addition, consistent with suggestions that subordinate-level categorizations require additional perceptual analysis, this study also reported that subordinate-level categorizations were associated with increased amplitude of early posterior potentials associated with visual processing, namely the N1 potential, relative to basic-level categorizations. These results provide further evidence of the greater demands placed on perceptual mechanisms during subordinate level categorizations, relative to those during both superordinate- and basic-level categorizations.

Another factor that might affect the time-course of object processing is experience. However, Fabre-Thorpe et al. (2001) explored the influence of stimulus-specific experience on the time-course of encoding and argued against this possibility. In this study, participants were extensively trained with a set of images in a categorization task in which the stimuli appeared in a rapid serial visual presentation (RSVP) stream. Participants were required to detect the presence of an animal in a briefly flashed (20 ms) scene. In this experiment, Fabre-Thorpe et al. (2001) found that performance increased for only the sub-set of trained images that were difficult to classify at the beginning of training. However, despite this limited improvement in behavioral performance, across all stimuli the stimulus-specific training did not influence indices of the time-course of object processing at a neurophysiological level, as the ERPs in response to trained and novel stimuli were indistinguishable (Fabre-Thorpe et al., 2001). Electrophysiological responses to target and non-target trials could be distinguished after as little as 150 ms, replicating the results a previous study in which participants were not trained (Thorpe, Fize, & Marlot, 1996). Notably, this time frame of 150 ms approaches the hypothetical lower bound (130 ms) for the time-course of object recognition, even if one assumes a purely feed-forward scheme (Allison, Puce, Spencer, & McCarthy, 1999). The authors suggested that experience could not compress the mechanisms and/or the time-course of visual processing.

It is important to note that this study, as well as others that suggest a 150 ms time-course for object recognition, is limited by its use of a superordinate-level categorization task (finding an “animal” or “human” face in the picture) (e.g., Fabre-Thorpe et al., 2001; Thorpe et al., 1996). As described above, previous work suggests that superordinate categorizations requires less perceptual processing than subordinate categorizations where the items have to be individuated, such as in the types of tasks commonly performed by visual experts. It is possible that a more perceptually challenging recognition task (e.g., a subordinate- rather than a superordinate- or basic-level categorization task that places additional demands on perceptual encoding processes) is required to reveal experience-based changes to the time-course of object recognition.

In sum, the time-course of object processing has been shown to be influenced by the level at which objects must be identified in categorization tasks. Participants can more readily access information at a basic-level, as such categorizations are not only faster than more general superordinate-level categorizations, but they are also faster than more specific subordinate-level categorizations. The slower time-course for subordinate-level categorizations reflects costs associated with the greater amount of perceptual information that needs to be extracted for more specific recognition tasks. Stimulus-specific training does not appear to influence the time-course of object processing for categorizations at a superordinate-level, but previous studies have shown that categorizations at this level rely more on semantic information, and less on perceptual information, than those at the subordinate level. The full potential for experience to influence the time-course of object processing is unclear because, although studies using simple categorization tasks show little impact of experience, more perceptually demanding tasks and more extensive training may be required to reveal the benefits of experience. Consistent with this suggestion, the next section discusses evidence that experience can impact the time-course of more specific subordinate-level categorizations.

Visual expertise and the time-course of object processing

Although Rosch et al. (1976) emphasized the importance of *structural* information in determining the basic-level categorization of an object, they also suggested that experience may influence this basic-level. They reported that one of their participants, an expert plane mechanic, appeared to recognize planes at a more subordinate-level. Tanaka and Taylor (1991) explored this possibility further in groups of real-world experts, namely bird experts and dog experts. Consistent with Rosch et al.'s (1976) observation with the plane mechanic, in a category verification task, car and bird experts were as fast to categorize objects from their domain of expertise at the subordinate-level as they were to categorize these objects at the basic-level. The authors suggested that the entry point of recognition may shift with expertise to a level that is more subordinate than the typical basic-level demonstrated among novice observers. Thus, expertise appears to impact the time-course of subordinate-level categorizations.

Equally fast subordinate- and basic-level categorization performance is considered a hallmark of visual expertise (Gauthier & Tarr, 1997; Tanaka, 2001; Tanaka & Taylor, 1991). In fact, it has been used in expertise training paradigms as a benchmark of successful training: Individuals participating in lab-based expertise training paradigms with novel objects, such as "greebles", are considered experts when they are as fast at identifying greebles at a basic-level (family) as they are at identifying them at a more subordinate-level (individual) (Gauthier & Tarr, 1997, 2002; Gauthier et al., 1999). Notably, lab-trained experts who have achieved this benchmark of visual expertise show many other general characteristics associated with real-world expertise (Gauthier et al., 2003; Gauthier et al., 2000), such as the greater difficulty recognizing aligned, relative to mis-aligned, halves of greebles or the differential recruitment of functionally defined face-selective areas of the brain (Gauthier & Tarr, 2002; Gauthier et al., 1999).

As a test of the expertise theory of face recognition (outlined in detail in Chapter I), Tanaka (2001) explored the speed with which typical observers can categorize faces at superordinate- ("living thing"), basic- ("human"), or subordinate-levels ("Bill Clinton"). Consistent with an expertise account of face processing, the entry point of face recognition was shifted to a more subordinate level. In fact, observers were faster to verify that a face was "Bill Clinton" than to verify that it was "human". In a paradigm similar to that used by Jolicoeur et al. (1984), Tanaka (2001) verified that the accuracy of subordinate-level categorizations of dogs and birds (among novices) were substantially impacted by a reduction in encoding time (75 ms compared to 950 ms). However, this same manipulation failed to impact subordinate-level categorizations of faces. This result provides evidence that sufficient perceptual information can be extracted during the first 75 ms of encoding to access identity-level representations of faces, but not for categories of objects for which an observer has less experience. These results were interpreted as reflecting the development of specialized "perceptual routines" by visual experts that permit them to rapidly analyze objects from the domain of expertise (Tanaka, 2001).

The earlier availability of identity-level information for faces compared to other categories suggests that face processing has a temporal advantage over the processing of non-face stimuli. Therefore, face processing may follow a different time-course compared to the processing of other categories, at least in the early stages of processing. The difference in the *amplitudes* of the peaks of the N170 potentials evoked by faces and those evoked by non-face objects suggests that there may be important differences in relatively early perceptual processing stages between face and non-face object processing (Bentin, Allison, Puce, Perez, & et al., 1996). However, the absence of a difference in the *timing* of the peaks of the N170 potentials appears to suggest that the *time-course* of face and object processing do not differ. However, when faces are inverted, the peak of the N170 is delayed (~10 ms) compared to that for upright faces (Rossion et al., 2000). Importantly, this delay is not present for the N170 in response to upright relative to inverted objects, ruling out accounts of the delay involving a general cost due to processing images in non-canonical orientations. However, inconsistent with an expertise account of this temporal advantage for upright faces, this study found no evidence of a similar delay for upright non-face processing relative to upright face processing.

However, a recent ERP study using a more sensitive analysis technique provided evidence of a temporal advantage for face processing, relative to the processing of upright non-face objects (Caldara et al., 2003). This analysis involved a spatio-temporal segmentation procedure that aimed to define intervals where surface ERP map topographies remain stable for a period of time, and the points in time where they change from one stable configuration into another. These segments are believed to reflect different functional stages in the brain (Brandeis & Lehmann, 1986; Michel, Henggeler, & Lehmann, 1992). Notably, this analysis found evidence that the stable segment that contained category-specific responses for face and non-face object stimuli occurred about 25 ms earlier for faces relative to non-face stimuli, a result not previously detected in analyses based only on the peak of the N170 potential (Caldara et al., 2003). Therefore, there is evidence that processing of non-face objects is also delayed relative to that for faces, although previous methods were not sensitive enough to reveal this delay. However, as with all comparisons between different categories of stimuli, a role of featural differences in producing this delay in the electrophysiological response cannot be ruled out.

In sum, there is evidence that visual expertise may change the time-course of object processing resulting in faster subordinate-level categorizations of objects of expertise, relative to typical objects. Given that subordinate-level categorizations require additional perceptual processing compared to less specific categorizations (Tanaka et al, 1999; Jolicoeur et al., 1984; Tanaka, 2001), the effect of expertise on subordinate-level categorizations may more specifically reflect an influence of expertise on the time-course of *perceptual encoding*. Neurophysiological evidence from studies recording ERPs is also consistent with expert processing being associated with a temporal advantage (Caldara et al., 2003; Rossion et al., 2000). However, it is unclear from studies demonstrating a temporal advantage for faces and object of expertise what processes underlies this temporal advantage. One possibility is that there is a general difference in the rate with which recognition approaches asymptote, whereas a second possibility is that there is a delay in when recognition of non-face objects begins. Unfortunately, current interpretations of the influence of experience on the time-course of encoding are limited by temporal sampling that is too coarse (75 ms, 950 ms; Tanaka, 2001), or more general problems associated with comparisons across stimulus categories (Caldara et al., 2003). A more thorough mapping of the time-course of object processing, from the onset of a stimulus to when performance asymptotes, for both expert and novice performance, is required to better understand the impact of expertise on the temporal dynamics of object processing.

The experiments in CHAPTER IV address the questions raised above regarding the underlying cause for the temporal advantage for face processing reported in previous studies; they explore whether this temporal advantage occurs at an encoding level, and if so, whether it is the result of a difference in the onset of recognition or rate of performance increase. Finally, they explore whether patterns are similar for face and non-face objects of expertise. The studies in this chapter use a "structural" backward masking paradigm, where masks share features with the object and are presented so that they cover the location of the original object. Varying stimulus-mask-onset-asynchronies are used to manipulate the time available to encode a stimulus. The essence of these studies is to vary the time that participants are given to encode the study item and then to measure how performance changes as a function of the pre-mask encoding duration.

To anticipate, performance in the experiments in CHAPTER IV followed an exponential function, with participants at chance performance at the shortest presentation duration (12 ms) and reaching asymptote by the longest presentation duration (1000 ms). A shifted exponential curve with three parameters was fitted to the data to better quantify the resulting performance function (Wickelgren, 1977; Wickelgren & Corbett, 1977). The three parameters estimated from the data were: the *intercept*, or time at which the curve first rises above chance; the *asymptote*, or ultimate performance level; and *rate*, or rate at which the asymptote is reached (see Figure 15). The choice of a shifted exponential curve to fit the data was theory-neutral, although its parameters provided a description of aspects of performance that were of specific interest to the aims of this study.

The first experiment in this chapter, Experiment 8, compared how performance varies with encoding time for upright and inverted faces. To anticipate the results, recognition of upright faces required less encoding time to exceed chance-level than did inverted faces. In addition, the time-course functions for upright and inverted faces could be distinguished quite early in processing. The exponential curve fitted to these data suggested that these differences in the time-course of processing for upright and inverted faces were a result of a difference in the onset time of recognition rather than a difference in the rate of performance increase. To ascertain whether these differences between upright and inverted faces were the result of a more general stimulus load difference between the encoding of objects in canonical and non-canonical views, Experiment 9 compared the performance/encoding time function for upright cars among car experts and car novices. The behavioral results of Experiment 9 were strikingly similar to those in Experiment 8. In addition, similar to the results of Experiment 8, the exponential curves fitted to the data from Experiment 9 were also more consistent with a different onset time of recognition for cars among car experts and novices, rather than a difference in the rate of increase in performance. This chapter concludes by discussing possible explanations and mechanisms for how learning may influence the time-course of object recognition.

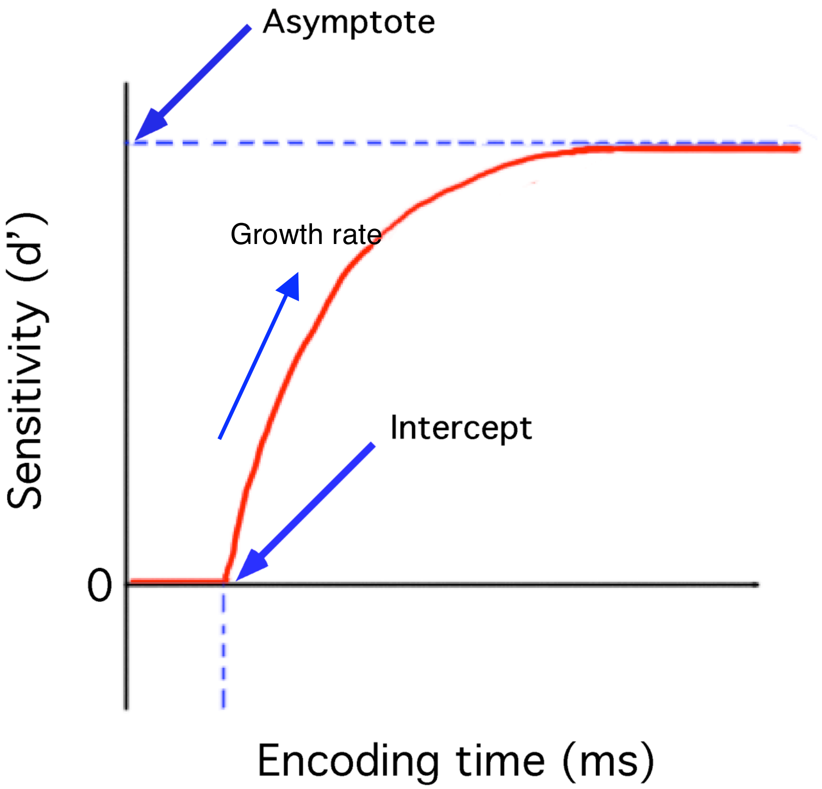


Figure 15. Diagram of a shifted exponential curve. This type of curve will be fitted to the data in this chapter to estimate the asymptotic performance level, intercept or performance onset, and the rate of growth of performance.

Experiment 8

Experiment 8 maps the time-course of encoding for upright and inverted unfamiliar faces in order to explore whether the dependence of experts' greater VSTM performance on sufficient encoding time (Experiments 1-7) might be rooted in differences at perceptual encoding stages. More specifically, Experiment 8 aims to test if the temporal advantage for upright over inverted faces detected in electrophysiological recordings, specifically the N170 component, is due to a difference in the time-course of *encoding* between upright and inverted faces (Rossion et al., 2000). If so, this same advantage should be detected in a behavioral backward masking paradigm in which temporal encoding limitations are manipulated across eight different stimulus-mask onset asynchronies. If an advantage is detected for upright relative to inverted faces, the fine sampling of performance under different encoding limitations should also provide insight into the source of this advantage, namely whether it is an earlier onset of recognition or a faster rate of performance increase with additional encoding time. Alternatively, the temporal advantage suggested by previous studies may reflect an advantage at later stages of processing, as the N170 occurs later than some estimates of the time-course of object processing (Fabre-Thorpe et al., 2001; Thorpe et al., 1996). If so, it is conceivable that experts may even demonstrate a greater cost to the rate of performance increase than novices do when encoding time is limited (although the asymptotes will likely be higher). Holistic encoding may be more burdensome on perceptual processing mechanisms compared to more feature-based encoding strategies. This pattern of results could provide a potential account for the greater influence of encoding time on experts' VSTM performance, relative to novices'.

Methods

Participants. Participants were 34 undergraduate students at Vanderbilt University enrolled in Psychology classes (age, M=18.94, SD=3.25, 25 females). All participants reported having normal or corrected to normal vision. Participants received course credit for their participation in this study.

Stimuli. Stimuli were 320 grey-scale, front-view images of faces in both upright and inverted orientations, resulting in a total of 640 images. Images were obtained from the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany, the Harvard Face database, Stirling face data base (http://pics.psych.stir.ac.uk/), Nottingham scans face database (http://pics.psych.stir.ac.uk/), and the Belgian face database. There were 240 different facial identities, with the additional images including a modified copy of 80 of the faces. The adjusted faces were cropped differently around the outer edge and the overall luminance was adjusted (Figure 16). Images subtended 3.5 x 4.7 degrees.

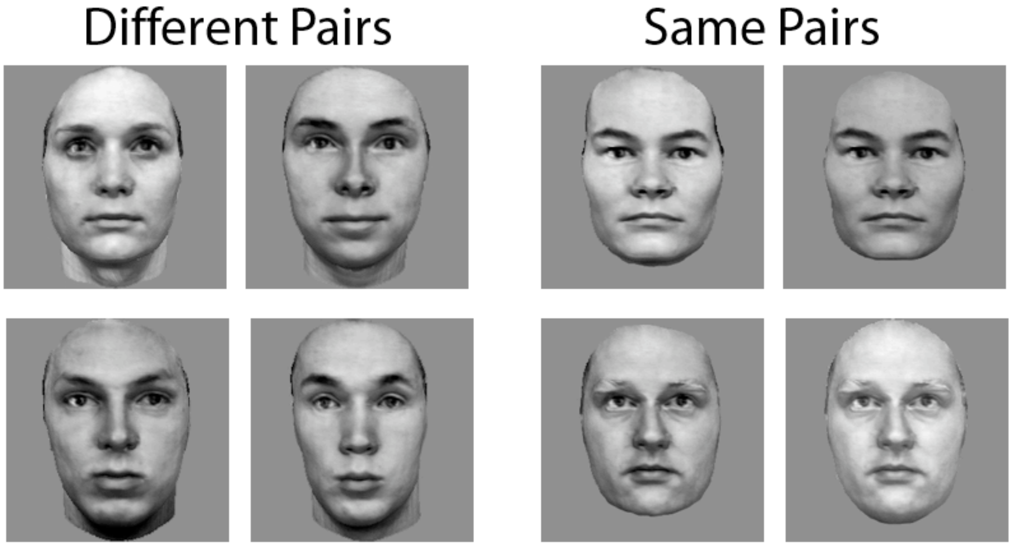


Figure 16. Examples of pairs of stimuli that appeared in Experiment 8 for which the correct response was “different” (right) and “same” (left). The stimuli used in the “same” trials were cropped differently and the overall luminance was changed.

Procedure. The task was presented using a 19” monitor (1024 x 768 pixel resolution, 85 Hz refresh rate) driven by Matlab (Mathworks, Inc) together with Psychophysics toolbox (Brainard, 1997) on a 450 or 733 MHz G4 Macintosh computer. Normal overhead lighting illuminated the testing room. Participants’ heads were stabilized at a fixed distance (75-cm) from the screen using a standard chin-rest before they performed the sequential matching task in which the first images were backward masked. A between subjects design was used, in which participants were randomly allocated to an upright or inverted face condition. A between subjects design was used due to stimulus limitations, as the same faces could not appear in both the upright and inverted conditions. Seventeen participants performed in each of the upright (age, M= 19.29, SD=1.83, 11 female) and inverted (age, M= 18.59, SD=0.94, 14 female) conditions. Participants initiated each trial by a space-bar press, which proceeded as follows: an image of a face appeared for either 12-, 47-, 82-, 118-, 153-, 235-, 494-, or 1000 ms. This was followed by a mask that appeared for 494 ms. Finally, a second image of a face appeared and remained on the screen until the participant made a key-press indicating if the faces had the same or different identities (Figure 17). Participants were told to respond as accurately and as fast as they could. All images were presented centrally with a small random jitter of 0, +/- 5, or +/- 10 pixels vertically or horizontally. Participants performed 160 trials, which consisted of 20 trials for each of the different presentations times with an equal number of trials in which the correct response was "same" and "different".

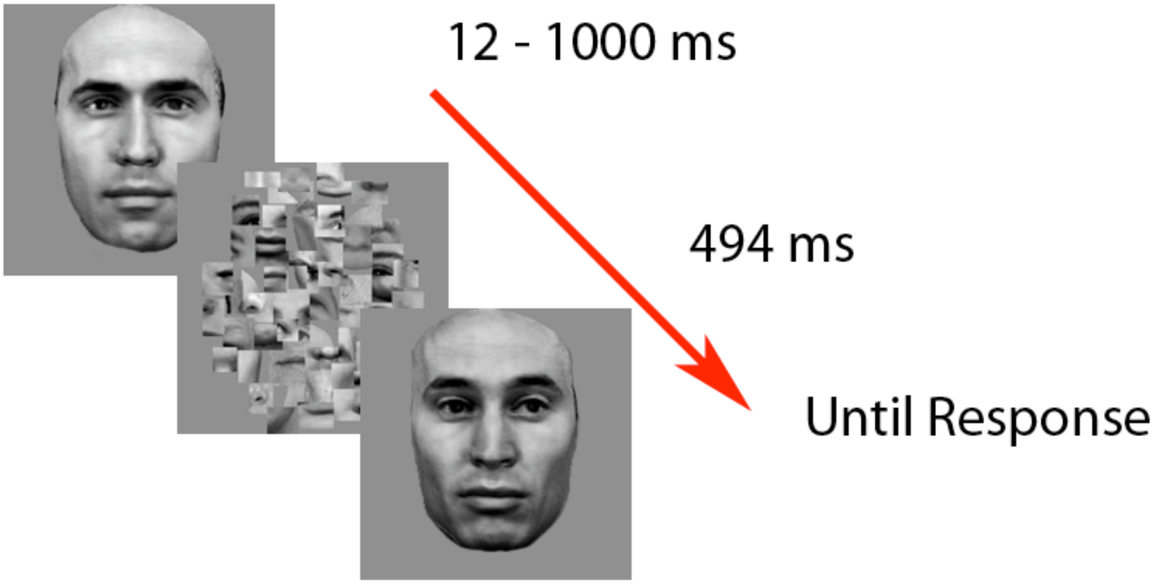


Figure 17. The sequence of events that occurred in each trial in Experiment 8. Participants were first presented with an upright or inverted face image that was masked after a 12, 47, 82, 118, 153, 235, 494, or 1000 ms. The masked remained on the screen for 494 ms after which a second face image appeared and remained until participants made a key press indicating whether the two faces had the same or different identities.

The faces in "same" trials were always different in some superficial ways, such as the way the outer edge of the face was cropped, the lighting in the image, or the general luminance. These differences between images in same trials were included deliberately in order to discourage participants from using a strategy in which they based their decision on a specific diagnostic feature of the image (e.g. the outer edge of the face) rather than the identity of the face. Participants were instructed to ignore any superficial difference between the face images (e.g., a general difference in luminance, or the way the outer edge of the face was cropped from its background) and to base their decision solely on the identity of the face. In addition, faces in the different trials were matched together in such a way that similar looking faces appeared together to further discourage subjects from adopting a strategy based on salient low-level features.

Sensitivity and response time analysis. Trials with a response time less than 200 ms or greater than 4000 ms were removed from the data set (2.6%). For each participant, sensitivity (d') and response time were calculated for each presentation duration condition for the upright and inverted orientations. An omnibus ANOVA was performed on both the response time and sensitivity data. In addition, planned t-tests were performed to identify the encoding time conditions in which sensitivity differed from chance (i.e. chance d' = 0) for upright and inverted faces (providing an estimate of the onset of performance), and to also compare sensitivity for upright and inverted faces to identify for which encoding conditions upright face performance could be distinguished from that for inverted faces.

Curve fitting. In order to compare the dynamics of upright and inverted face encoding, a shifted exponential function were fitted to the averaged d' values at each time point (Wickelgren & Corbett, 1977):

$$\hat{d}' = A\left\{1 - \exp\left[-R(t-I)\right]\right\} \textit{ for } t > I, \text{ otherwise } \hat{d} = 0. \tag{1}$$

Here, A is the asymptote, R is the rate of approach to asymptote, I is the intercept, and t is the stimulus presentation duration. The inverse of the rate, $1/R$, is expressed in seconds.

The eight possible models derived from this exponential function, which differed in the number of free parameters, were fitted to the 16 (2 stimulus types x 8 presentation durations) data points, d_i. More specifically, the models differ in whether the intercepts, rates of approach to the asymptote, or the asymptotes were the same or different for the curves fitted to the observed data. This was notated with a 1 or 2 respectively (e.g. 2I, 1R, 2A). The least constrained model (6 parameters; 2 I, 2 R, 2 A), allowed different intercepts, rates of approach to the asymptote, and asymptotes for the performance functions for upright and inverted faces. Three models in which one parameter was constrained were also tested (5 parameters) where the intercept (1 *I*, 2 *R*, 2 *A*), the rate (2 *I*, 1 *R*, 2 *A*), or asymptote (2 *I*, 2 *R*, 1 *A*), could not vary between upright and inverted faces. Three models in which 2 parameters were constrained were also tested (4 parameters) where the intercept and rate (1 *I*, 1 *R*, 2 *A*), intercept and asymptote, (1 *I*, 2 *R*, 1 *A*), or rate and asymptote (2 *I*, 1 *R*, 1 *A*), could not vary between upright and inverted faces. The most constrained model possible (3 parameters; 1 *I*, 1 *R*, 1 *A*), assuming equivalent intercept, rate, and asymptotes for upright and inverted faces, was also tested. Fits were calculated for the group averaged sensitivity values to increase power as the individual data were quite variable due to the relatively small number of trials (20) contributing to each data point.

Data fitting was implemented in Matlab (Mathworks) using a simplex hill-climbing function to iteratively adjust parameters to maximize the value of the variance accounted for (r^2). These fits were replicated using Microsoft Excel Solver (Mircrosoft Corp., Redmond, WA). The range of values from which the starting point of the curve fitting functions could be selected were limited as follows: intercept, 0.01 - 0.5, growth rate 0.01 – 10.0, asymptote, 0.1 - 5.0). To increase the plausibility of the resulting best fit models, the fitting procedure was restricted so that the asymptote parameter for novices could not exceed that for experts. Two thousand iterations were performed per model to fit the data. Goodness of fit was assessed by calculating r^2 values (equation 2).

$$r^2 = 1 - \frac{\sum_{i=1}^{N}\left(d_i - \hat{d}_i\right)^2}{\sum_{i=1}^{N}\left(d_i - \bar{d}\right)^2} \qquad (2)$$

Nested models (whose parameters are proper subsets or super-sets) were statistically compared by application of a *F*-test comparing a full (greater number free parameters) to a reduced model (fewer number of free parameters) as follows in equation 3:

$$F\left(df_1, df_2\right) = \frac{\left(r_{full}^2 - r_{reduced}^2\right)/df_1}{\left(1 - r_{full}^2\right)/df_2} \qquad (3)$$

where $df_1 = k_{\text{full}} - k_{\text{reduced}}$, and $df_1 = N - k_{\text{full}}$. The k's are the number of free parameters in each model, and N is the number of predicted data points. This F-test incorporates an adjustment for the number of free parameters. The resulting F-values were converted to a t-value to evaluate the nested models. This conversion was done so that a one-tailed test could be applied. A one-tailed test is most appropriate because, logically, the fit of the nested model (i.e. the model with less free parameters) must be lower than, or at best equal to, that for the less constrained model.

Goodness of fit was also evaluated by comparing—between models— the r^2_{adjusted}, which is the proportion of variance accounted for after it is adjusted for the number of free parameters, k (Reed, 1976).

$$r^2 = 1 - \frac{\sum_{i=1}^{N}\left(d_i - \hat{d}_i\right)^2 / (N-k)}{\sum_{i=1}^{N}\left(d_i - \bar{d}\right)^2 / (N-1)} \tag{4}$$

where N is the number of data points (d_i), $\hat{d}_i$ is the value predicted by equation 1, k is the number of free parameters, and $\bar{d}$ is the overall mean.

Results

To summarize the results, given sufficient encoding time, sensitivity (d') for upright faces was greater than that for inverted faces. Recognition sensitivity (d') for upright and inverted faces was significantly different with as little as 83 ms presentation (Table 1). In addition, sensitivity for identifying upright faces differed from chance by the 48 ms presentation duration condition, while 118 ms was required before the identification of faces in an inverted orientation differed from chance (Figure 18 A, Table 2). These effects were not explained by differences in response times for trials with upright and inverted faces (Figure 18 B). The exponential curve that best fit these data was one in which the rate of approach to the asymptote with increasing presentation duration was the same for upright and inverted faces (Figure 19, Table 3). In this model, the presentation duration required for performance to exceed chance-level (i.e. the intercept value), as well as the asymptote value, differed for upright and inverted faces. This model predicted that the onset for upright face performance was 33 ms earlier than that for inverted faces.

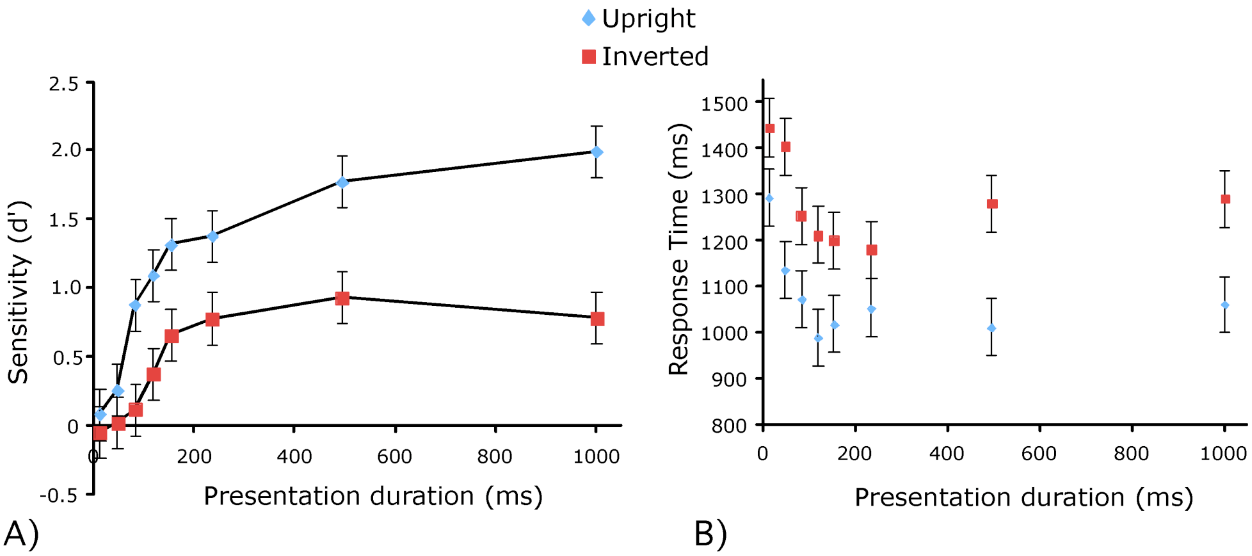


Figure 18. (A) Sensitivity (d') and (B) response time as a function of presentation duration in the backward masked sequential (identity) matching task with upright and inverted faces in Experiment 8. Error bars represent standard error of the mean.

Sensitivity analysis. A 2 (orientation; upright, inverted) x 8 (presentation duration; 12 ms, 48 ms, 83 ms, 118 ms, 153 ms, 236 ms, 495 ms, 1000 ms) ANOVA performed on the sensitivity measures (d') found main effects of presentation time, F(7,224)=30.85, $p \leq .0001$, and orientation, F(1,32)=30.09, $p \leq .0001$. Sensitivity was greater for longer presentation durations and for upright compared to inverted faces. There was an interaction between orientation and presentation duration, F(7,224)=3.25, p=.0027, with the benefit of longer presentation durations being greater for upright than inverted faces. Increased presentation duration led to a rapid increase in performance, although the rate of increase started to slow with approximately153 ms presentation duration as performance neared the asymptote.

Planned t-tests (Table 1) revealed that sensitivity for the upright and inverted trials could only be distinguished at a minimum of 83 ms presentation duration. Although sensitivity differed from chance with as little as 48 ms presentation duration for upright faces, while 118 ms presentation duration was required for sensitivity in the matching task to increase above chance for inverted faces (Table 2).

Table 1. Statistical comparisons (*t*-test) between sensitivity (d') for upright and inverted faces at each presentation duration.

Condition	Upright	Inverted	df	t	p
12 ms	.086	-.047	32	.66	.5150
48 ms	.264	.022	32	1.35	.1851
83 ms*	.879	.119	32	2.69	.0113
118 ms	1.092	.381	32	3.65	.0009
153 ms	1.318	.660	32	3.02	.0049
236 ms	1.384	.781	32	3.53	.0013
495 ms	1.775	.931	32	3.70	.0008
1000 ms	1.995	.787	32	6.20	≤.0001

* shortest presentation duration at which a difference in performance can be distinguished for upright and inverted faces using one-tailed planned t-tests.

Table 2. Statistical comparison of sensitivity (d') for upright (left panel) and inverted (right panel) faces at each encoding duration to determine the shortest presentation duration required for performance to differ from chance.

Condition	*df*	*t*	p	Condition	*df*	*t*	p
Upright faces				*Inverted faces*			
12 ms	16	.55	.5880	12 ms	16	-.36	.7205
48 ms *	16	2.38	.0300	48 ms	16	.16	.8750
83 ms	16	4.38	.0005	83 ms	16	.60	.5571
118 ms	16	7.85	.0001	118 ms *	16	2.80	.0130
153 ms	16	8.69	≤.0001	153 ms	16	4.23	.0006
236 ms	16	10.40	≤.0001	236 ms	16	7.32	≤.0001
495 ms	16	10.78	≤.0001	495 ms	16	5.90	≤.0001
1000 ms	16	15.13	≤.0001	1000 ms	16	5.49	≤.0001

* shortest presentation duration at which performance can be distinguished from chance (0 d') using one-tailed planned t-tests.

Response time analysis. A 2 (orientation; upright, inverted) x 8 (presentation duration; 12 ms, 48 ms, 83 ms, 118 ms, 153 ms, 236 ms, 495 ms, 1000 ms) ANOVA performed on the response time data found main effects of presentation duration, $F(7,224)=9.03$, $p\leq.0001$, and orientation, $F(1,32)=4.75$, $p=.0369$. Response times were faster for longer presentation duration conditions and for judgments about upright faces than inverted faces. There was no interaction between orientation and presentation duration ($F<1$). Increased encoding led to a relatively uniform reduction in response time until the 153 ms presentation duration condition after which mean response times started to plateau for both upright and inverted faces.

Curve fitting. Table 3 and Figure 19 shows the best fits to the average sensitivity measure across the subjects within each group for the eight models tested. Constraining the rate parameter (2 I, 1 R, 2 A) so that it was the same for both upright and inverted faces did not result in a significant drop in the variance accounted for across the two groups, relative to the full model (2 I, 2 R, 2 A), t(10)=1.48, p=.084. However, constraining the intercept parameter (1 I, 2 R, 2 A), t(10)=2.48, p=.0164, or the asymptote (1 I, 2 R, 2 A), t(10)=6.78, p<.0001, so they were the same for both upright and inverted faces did result in a significantly worse fit relative to the full model (2 I, 2 R, 2 A). Therefore, the data were best described by a model in which the asymptote and intercept, but not the rate of approach to asymptote, differed for upright and inverted faces. This model suggests that not only is there a considerable difference in the asymptote level of performance between the two groups (.92 d' difference), but also that the onset of recognition performance for upright faces occurs approximately 33 ms before that for inverted faces. However, recognition performance for upright and inverted faces with increasing presentation duration approached asymptote at the same rate.

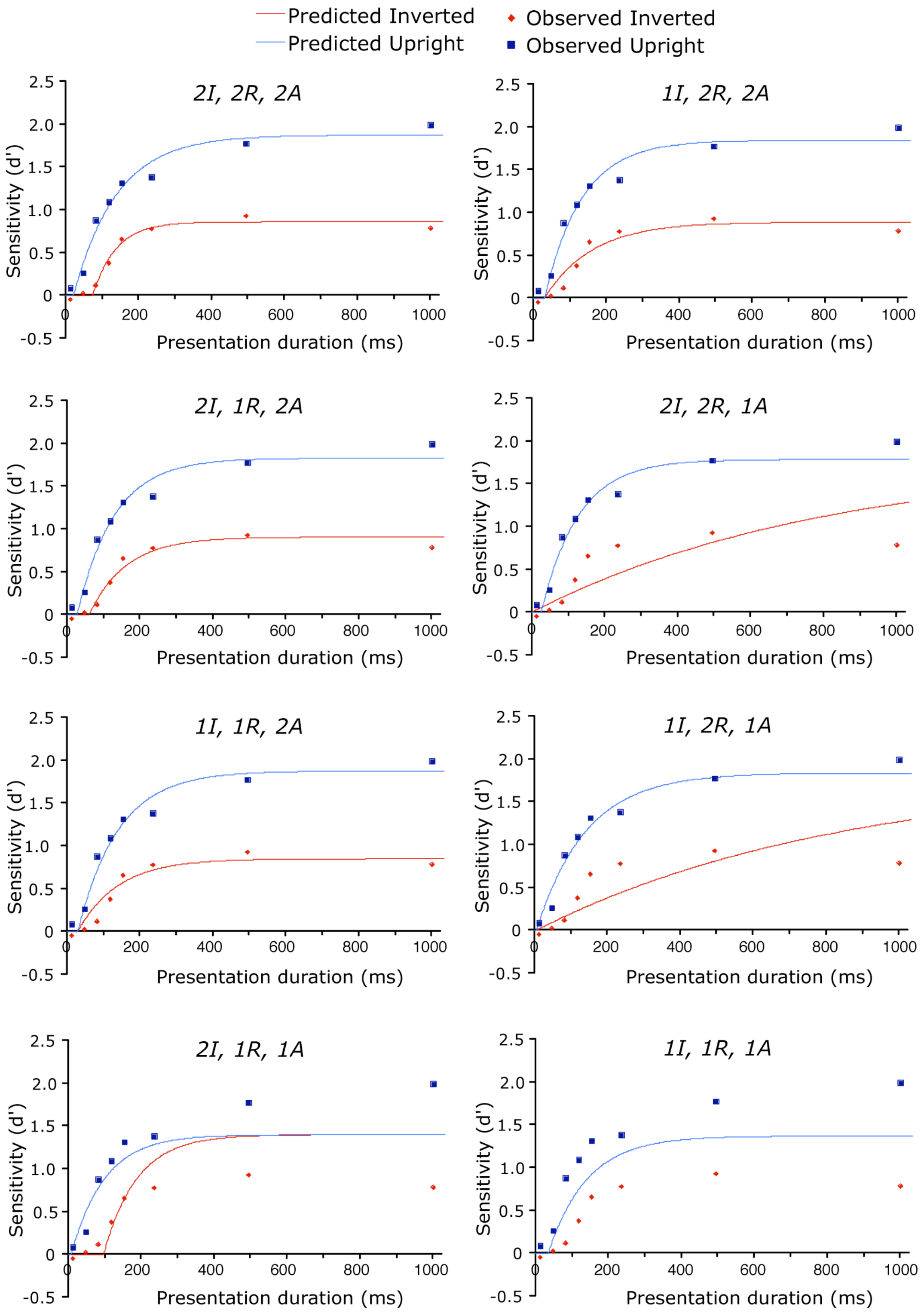


Figure 19. The curves that best fit the observed average d' values for upright and inverted faces for each of the eight possible models derived from the shifted exponential function. The models differed in terms of whether the intercept (*I*), rate of approach to asymptote performance level (*R*), and the asymptote (*A*) parameters were fixed or free to vary (notated as a 1 or 2, respectively) for the curves fitted to the performance for upright and inverted faces across the different presentation duration conditions.

Table 3. Best Fits of the Exponential Function in Experiment 8

Condition	Intercept	Rate	Asymptote	$r^2_{\text{(adjusted)}}$
2 I, 2 R, 2 A				
Upright	.021019	8.4211	1.8752	
Inverted	.07364	15.6659	.86014	
				.98127 (.971905)
1 I, 2 R, 2 A				
Upright	.029894	9.7598	1.8469	
Inverted	.029894	7.6219	.89059	
				.96979 (.95880)
2 I, 1 R, 2 A				
Upright	.027290	9.8626	1.8306	
Inverted	.060108	9.8626	.90636	
				.97715(.96884)
2 I, 2 R, 1 A				
Upright	.025273	9.8246	1.7877	
Inverted	2.6365e-012	1.2332	1.7877	
				.89528(.8572)
1 I, 1 R, 2 A				
Upright	.029074	9.0355	1.8778	
Inverted	.029074	9.0355	.8499	
				.96796(.95995)
1 I, 2 R, 1 A				
Upright	.0054576	7.3429	1.8398	
Inverted	.0054576	1.1811	1.8398	
				.89050(.86313)
2 I, 1 R, 1 A				
Upright	.007568	10.929	1.401	
Inverted	.096323	10.929	1.401	
				.77408(.7176)
1 I, 1 R, 1 A				
Upright	.034286	9.3891	1.369	
Inverted	.034286	9.3891	1.369	
				.64415(.58940)

Note: The models differed in terms of whether the intercept (*I*), rate of approach to asymptote performance level (*R*), and the asymptote (*A*) parameters were fixed or free to vary (notated as a 1 or 2, respectively) for the curves fitted to the performance for upright and inverted faces across the different presentation duration conditions.

Discussion

The results of Experiment 8 suggest that recognition performance for upright and inverted faces not only differs in magnitude, but that they also follow different time-courses to reach these different levels of performance. As expected based on the widely reported face inversion effect, sensitivity for matching upright faces was projected to reach a higher asymptotic level than that for inverted faces. More notably, the onset time for recognition performance for inverted faces was delayed relative to that for upright faces, as inverted faces required additional encoding time to exceed chance-level compared to the recognition of upright faces. However, the absence of a difference in the rate of increase in performance suggests that the temporal advantage for upright faces found in previous studies may be a result of a more qualitative difference in processing, rather than a general "speeding up" of performance. Notably, the *longer* response times in the inverted face condition, relative to the upright face condition, is inconsistent with the possibility that the temporal advantage for upright over inverted faces relies on additional post-perceptual processing for upright faces, but not inverted faces, and thus is consistent with the difference arising from early stages of processing.

There are a number of possible explanations for the initial delay in the performance onset in the inverted face condition relative to the upright face condition. One possibility is that this delay reflects the need to perform an additional process for inverted, but not upright, faces before they can be processed for recognition. For example, observers may apply some transformation to the inverted faces, such as mentally rotating them. This may be a cost experienced more generally when encoding inverted objects. Alternatively, the difference in the onset times for upright and inverted faces may be driven by an especially rapid onset of processing of upright faces rather than a slowed onset for inverted face processing, relatively to the processing of other upright or inverted object categories. For example, faces may have privileged access to capacity-limited attention necessary for guiding processing. Consistent with this possibility, some suggest that faces can capture attention (Theeuwes & Van der Stigchel, in press; Vuilleumier, 2000). In contrast, others suggest that holistic face processing can occur in the absence of attention (Boutet et al., 2002). Therefore, the processing of inverted faces may be delayed relative to the processing of upright faces due to the need to trigger processes that may occur automatically for upright faces, or due to attentional limitations not experienced for upright processing. This advantage, whether it be the absence of a need for attention or ability to preferentially capture attention compared to inverted faces, may be specific to upright faces. Alternatively, this temporal advantage may be acquired as a by-product of our extensive experience with faces and thus may generalize to non-face objects of expertise. Experiment 9 explores these possibilities.

Experiment 9

If the early advantage for upright, relative to inverted, face recognition judgments found in Experiment 8 reflects an acquired advantage due to our extensive experience with faces, car experts should show a similar "head start" relative to car novices, for similar recognition judgments about cars. However, if the difference in performance onset between upright and inverted faces instead reflects a general disadvantage for processing mis-oriented items, the processing of cars among both experts and novices should follow the same time-course pattern as that for faces. Therefore, two possible accounts of the temporal advantage for upright faces compared to inverted faces are proposed; the temporal advantage for upright faces may be a by-product of our extensive experience with upright faces or a more general advantage for processing objects in their canonical orientation. These different explanations for the temporal advantage for upright relative to inverted faces can be delineated by comparing the temporal dynamics of encoding among car experts and car novices. Experiment 9 uses the same general paradigm as Experiment 8 to map the time-course of car recognition among car experts and novices.

Methods

Participants. Participants were 38 (29 male) employees, undergraduate students, or graduate students of Vanderbilt University, or members of the surrounding Nashville community. Participants were compensated $10 per hour for their time. All participants had normal or corrected to normal vision and reported having either extensive or very little experience identifying models of cars. Their level of car expertise was quantified using the same task as used by Gauthier, et al. (2000) and described in detail in APPENDIX D. Nineteen participants (15 males) met the criteria for car expertise (age, M = 24.37, SD =5.02, car d' M = 2.78, bird d' M = 0.94), while 19 (14 males) were classified as car novices (age, M = 23.42, SD = 3.55, car d' M = 0.99, bird d' M = .90). Participants' self-reports of their car recognition skill were generally consistent with their performance on the subordinate car matching task, with participants meeting the criteria for expertise on the task rating themselves an average of 8.45 on a scale of 10; those who were classified as novices, on average, rated their skill as 4.97 on a scale of 10.

Stimuli. Stimuli were 320 grey-scale profile images of cars. There were 240 different car models, with the additional images including a modified copy of 80 of the car models. Similar to the duplicate face images used in Experiment 8, the overall luminance and/or shade of grey of the car panels and/or the tinting of the windows were adjusted (Figure 20). All the wheel-covers on the cars were replaced with one of six different kinds in such a way that cars appearing within the same trial always had the same wheel-covers. Images subtended 2.9 x 6.7 degrees of visual angle at the fixed viewing distance of 70-cm.

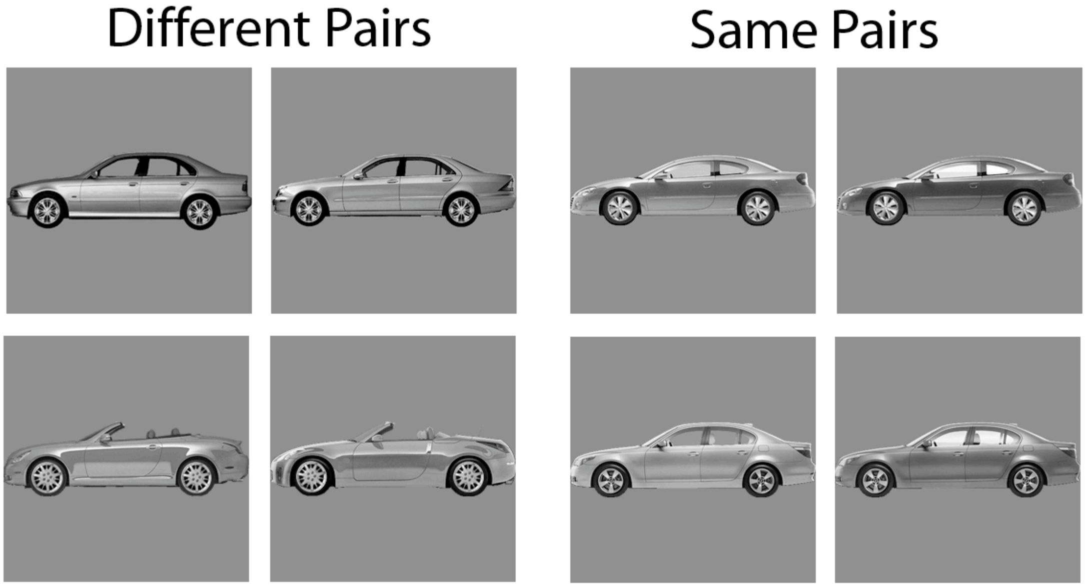


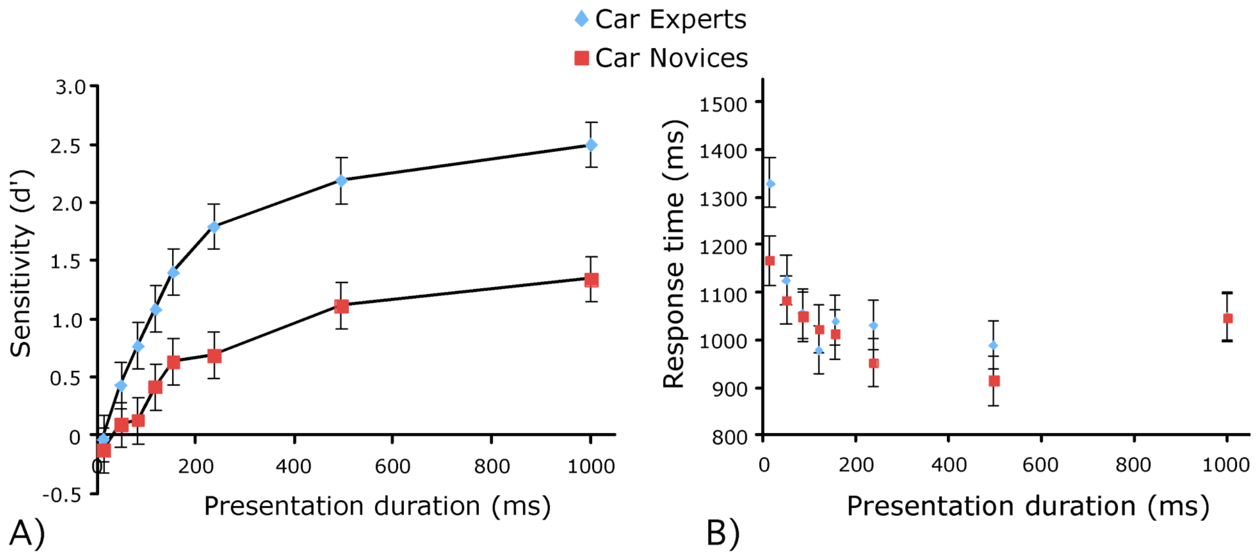
Figure 20. Examples of pairs of stimuli that appeared in Experiment 9 for which the correct response was “different” (right) and “same” (left). The tinting of the windows and the overall luminance was changed for stimuli appearing in “same” trials.

Procedure. The procedure was similar to that used in Experiment 8, except that images of cars, instead of faces, were used as stimuli, and all participants viewed the cars in an upright orientation. Once again, the images in the “different” trials were paired together in such a way as to maximize their similarity. In addition, the cars in “same” trials were always different in some superficial ways that were similar to the non-diagnostic differences found between the different model cars in “different” trials. For example, cars in “same” trials could differ in terms of the degree of tinting of the windows, the lighting in the images, and/or the color (shade of grey once the images were converted to grayscale) of the car. Once again, participants were instructed to ignore any superficial differences between the cars and to base their decision solely on the model of the car.

Analysis. Analysis was the same as that used in Experiment 8. Data from one participant in the novice group whose average performance did not exceed chance (50%) were excluded from further analysis. In addition, response times were found to differ between the expert and novice groups, with experts responding more slowly than novices. This introduced a potential confound. The five novices with the fastest average response times were excluded from the analysis and as a result there was no longer a significant difference in the response times between the two groups. This also resulted in an equal number of participants in the novice and expert groups.

Results

To summarize, the pattern of behavioral performance for cars among car experts relative to that among car novices in Experiment 9 was strikingly similar to that for upright faces relative to inverted faces in Experiment 8. Sensitivity for the expert and novice groups could be distinguished in the 48 ms presentation duration condition, which is slightly earlier than that found for upright and inverted faces (Table 4). If the presentation duration is sufficient, sensitivity for the matching judgments among car experts was greater than that among car novices. Car experts’ sensitivity differed from chance by the 48 ms presentation duration condition, however novices required 118 ms before their performance differed from chance (Figure 21 A, Table 5). This mirrors the results found for upright and inverted face trials in Experiment 8. These effects were not explained by a differential trade-off between response time and recognition performance among car experts and car novices (Figure 21 B). There appeared to be a trade-off between the rate and intercept parameters in accounting for the observed performance among car experts and car novices: the observed performance was best explained by either a model in which the rate of approach to asymptote was the same for experts and novices, but the intercept and asymptote parameters differed for experts and novices (Figure 23, Table 7), or a model in which the intercept was the same for the two groups, but there was a difference in the rate and asymptote parameters (Figure 21, Table 6). However, when the fit was restricted to the earlier part of the function, the model that allowed the intercept parameter to vary could better explain the performance of experts and novices when the plausibility of the parameter estimates and the relative fit (adjusted for the number of free parameters) of the different models was considered.



·*Figure 21.* (A) Sensitivity (d') and (B) response time as a function of presentation duration in the backward masked sequential matching task with cars among groups of car experts and car novices in Experiment 9. Error bars represent standard error of the mean.

Sensitivity analysis. A 2 (group; expert, novice) x 8 (presentation duration; 12 ms, 48 ms, 83 ms, 118 ms, 153 ms, 236 ms, 495 ms, 1000 ms) ANOVA performed on the sensitivity measure (d') found main effects of presentation duration, F(7,252)=49.13, p≤.0001, and expertise group, F(1,36)=58.43, p≤.0001. Sensitivity was greater among car experts than car novices and was also greater for longer presentation durations. There was also an interaction between expertise group and presentation duration, F(7,252)=3.71, p=.0008, with the benefit of longer presentation durations being greater for car experts than car novices. Increased presentation duration led to a rapid increase in performance, although the rate of increase started to slow with approximately153 ms presentation as performance approached the asymptote.

Planned t-tests revealed that sensitivity for the upright and inverted trials could only be distinguished at 48 ms and greater presentation durations (Table 4). However, sensitivity differed from chance with as little as 48 ms presentation duration for upright faces, while 118 ms was required for car novices' sensitivity to increase above chance (Table 5).

Table 4. Statistical comparisons (*t*-test) between sensitivity (d') for cars among car experts and car novices at each presentation duration.

Condition	Expert	Novice	df	t	p
12 ms	-.025	-.128	36	.0023	.9982
48 ms*	.434	.097	36	2.09	.0440
83 ms	.774	.134	36	3.12	.0035
118 ms	1.093	.421	36	3.25	.0025
153 ms	1.405	.635	36	3.47	.0014
236 ms	1.799	.690	36	4.45	≤.0001
495 ms	2.193	1.116	36	5.57	≤.0001
1000 ms	2.500	1.348	36	5.06	≤.0001

* shortest presentation duration at which performance can be distinguished for car experts and novices using one-tailed planned t-tests.

Table 5. Statistical comparison (*t*-test) of sensitivity (d') for cars among car novices (left) and car experts (right) at each presentation duration to determine the shortest encoding time required for performance to differ from chance.

Condition	*df*	*t*	p	Condition	*df*	*t*	p
Car Novices				*Car Experts*			
12 ms	18	-1.042	.3112	12 ms	18	-.225	.8244
48 ms	18	.7431	.4670	48 ms*	18	3.059	.0068
83 ms	18	.9144	.3726	83 ms	18	5.213	≤.0001
118 ms*	18	3.405	.0032	118 ms	18	6.540	≤.0001
153 ms	18	3.324	.0038	153 ms	18	9.266	≤.0001
236 ms	18	4.861	≤.0001	236 ms	18	9.643	≤.0001
495 ms	18	9.163	≤.0001	495 ms	18	15.14	≤.0001
1000 ms	18	8.441	≤.0001	1000 ms	18	19.04	≤.0001

* shortest presentation duration at which performance can be distinguished from chance (0 d') using one-tailed planned t-tests.

Response time analysis. A 2 (group; expert, novice) x 8 (presentation duration; 12 ms, 48 ms, 83 ms, 118 ms, 153 ms, 236 ms, 495 ms, 1000 ms) ANOVA performed on the response time data found a main effect of presentation duration, F(7,252)=12.58, p≤.0001, but no main effect of expertise, F<1. There was also no interaction between expertise and presentation duration, F(7,252)=1.48, p=.174. Therefore, although response times were faster with longer presentation durations, they did not differ across the expert and novices groups. Increased encoding led to a reduction in response time until the 153 ms presentation duration condition after which the mean response times appeared to plateau for both experts and novices. (Figure 21 B).

Curve fitting. Table 6 and Figure 22 show the best fits to the averaged sensitivity measures across the subjects within each group, for all eight models tested. Constraining the asymptote (2 I, 2 R, 1 A), t(10)=4.99, p=.00025, resulted in a significantly worse fit relative to the full model (2 I, 2 R, 2 A). Constraining the intercept parameter (1 I, 2 R, 2 A), t(10)=1.02, p=.167, or the rate parameter (2 I, 1 R, 2 A), t(10)=1.46, p=.087, so that they were the same for both upright and inverted faces did not significantly effect the fit relative to the full model (2 I, 2 R, 2 A). However, constraining both the rate and intercept parameters within the same model (1 I, 1 R, 2 A), t(11)=3.51, p=.0025, resulted in a significant impact on the fit relative to the full model (2 I, 2 R, 2 A). This suggests that the intercept and rate parameters traded-off in their ability to account for the difference in performance between car experts and novices.

From observing the fits to the intercept constrained (I, 2 R, 2 A) or rate constrained (2 I, 1 R, 2 A) models it appears that a trade-off may be occurring between different stages in the time-course or object processing, namely the earlier versus the later stages of the time-course (see Figure 22). The intercept constrained model appeared to better account for the longer presentation durations while the rate constrained model appeared to better account for the shorter presentation durations. A similar reduction in fit to the early data points was observed for the intercept constrained (1 I, 2 R, 2 A) model in Experiment 8 (see Fig 19). Given that the focus of this study was on the early encoding stage of the time-course of object processing, the curves were re-fitted only to the shorter presentation durations where performance was still rising sharply (12 ms, 48 ms, 83 ms, 118 ms, and 153 ms) to see if the rate and intercept constrained model could be distinguished on the basis of their ability to account for performance when presentation duration was more limited. The curve resulting from the model in which the rate was controlled, so that it could not vary across the expert and novice groups (2 I, 1 R, 2 A), fitted equally well as the curve resulting from the full model, t(4)=.17 p=.436 (Figure 23, Table 7).

However, the reduction in the fit of the curve that assumed that the intercept was the same for car experts and novices (1I, 2 R, 2 A), relative to the curves that assumed all three parameters differed across the groups (2I, 2 R, 2 A), approached significance, t(4)=1.77 p=.076. In addition, to achieve this fit, the intercept constrained model equated the asymptote parameters for experts and novices, which is *not* consistent with the behavioral performance of the two groups. Notably, the equivalence of the asymptote parameters for the functions describing expert and novice performance is a consequence of the restriction placed on the fitting procedure that the asymptote parameter for novices could not exceed that for experts (Table 7). This was done to increase the plausibility of parameters that fit the data best. Therefore, if the fitting procedure were restricted so that the asymptote parameter for experts actually *exceeded* that for novices, consistent with the observed performance, the fit would presumably be reduced further. This could possibly even push the reduction in fit, relative to the full model, into significance. In addition, once the fits for the models were adjusted for the number of free parameters (r^2_{adjusted}), the model in which the intercept and asymptote varied across experts and novices, but not the rate, was associated with the highest r^2_{adjusted} value, even exceeding that for the full model (Table 7). This model of performance with the greatest r^2_{adjusted} suggested that the onset of performance for car experts starts approximately 53 ms earlier than that among novices. It also suggested that asymptotic level of performance among experts is 1.14 d' units greater than that among novices.

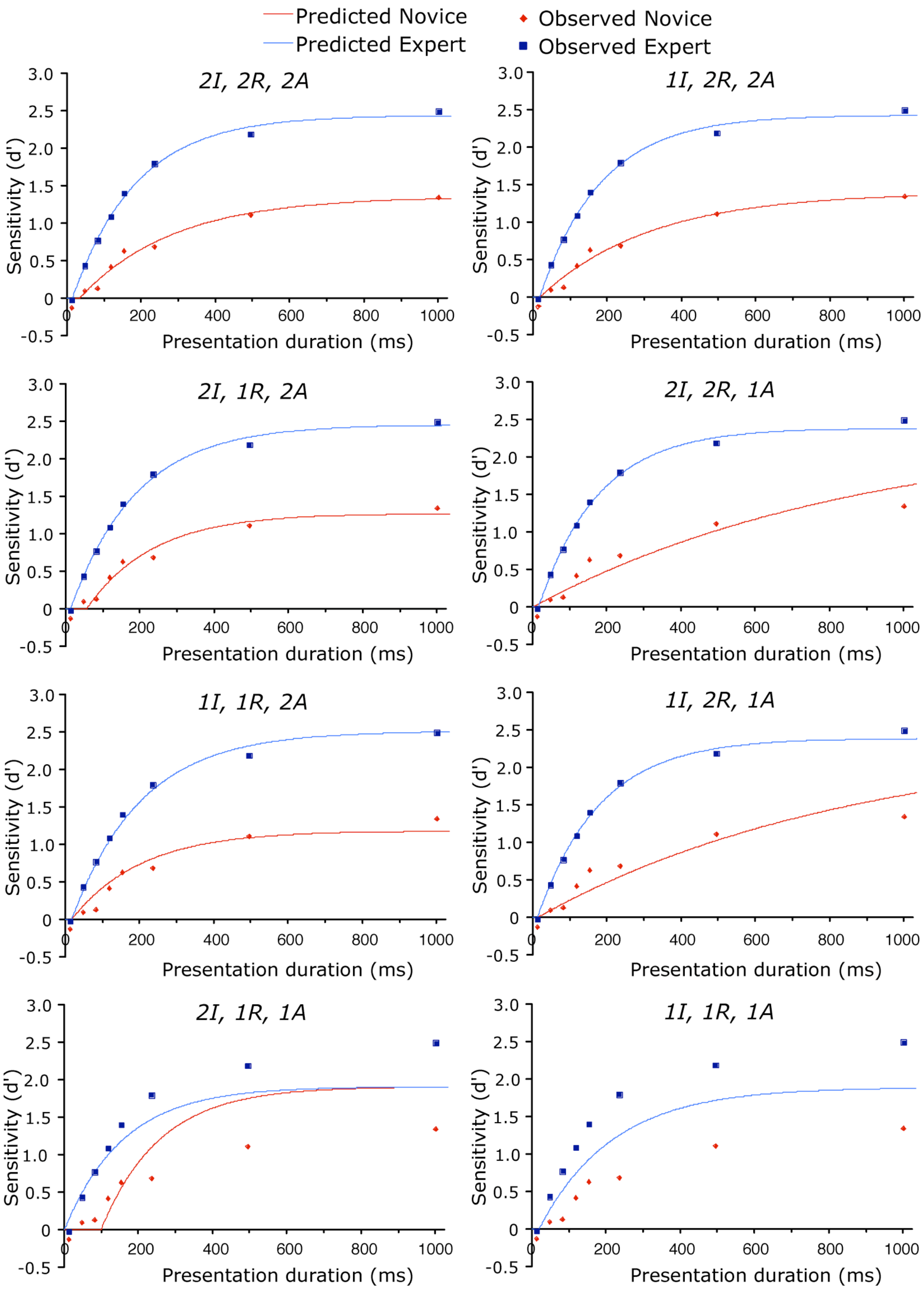


Figure 22. The curves that best fit the observed average d' values for recognition of cars among car experts and novices for each of the eight possible models derived from the shifted exponential function. The models differed in terms of whether the intercept (*I*), rate of approach to asymptote performance level (*R*), and the asymptote (*A*) parameters were fixed or free to vary (notated as a 1 or 2, respectively) for the curves fitted to the performance for experts and novices across the different presentation duration conditions.

Table 6. Best Fits of the Exponential Function in Experiment 9

Condition	Intercept	Rate	Asymptote	$r^2_{\text{(adjusted)}}$
2 I, 2 R, 2 A				
Expert	.013643	5.8004	2.4456	
Novice	.032692	3.9518	1.3608	
				.99275 (.989125)
1 I, 2 R, 2 A				
Expert	.016935	6.027	2.4354	
Novice	.016935	3.3353	1.4023	
				.99200 (.98909)
2 I, 1 R, 2 A				
Expert	.012	5.5949	2.4635	
Novice	.055204	5.5949	1.2775	
				.9912 (.98800)
2 I, 2 R, 1 A				
Expert	.015669	6.1544	2.393	
Novice	5.7412e-012	1.1273	1.1273	
				.97473 (.96554)
1 I, 1 R, 2 A				
Expert	.016085	5.3392	2.5189	
Novice	.016085	5.3392	1.1863	
				.98303 (.97879)
1 I, 2 R, 1 A				
Expert	.012	5.9316	2.3967	
Novice	.012	1.1644	2.3967	
				.97322(.966525)
2 I, 1 R, 1 A				
Expert	1.524e-012	6.3054	1.9143	
Novice	.098398	6.3054	1.9143	
				.82224 (.7778)
1 I, 1 R, 1 A				
Expert	.018276	5.0317	1.8956	
Novice	.018276	5.0317	1.8956	
				.71004(.66543)

Note: The models differed in terms of whether the intercept (*I*), rate of approach to asymptote performance level (*R*), and the asymptote (*A*) parameters were fixed or free to vary (notated as a 1 or 2, respectively) for the curves fitted to the performance for cars among car experts or novices across the different presentation duration conditions.

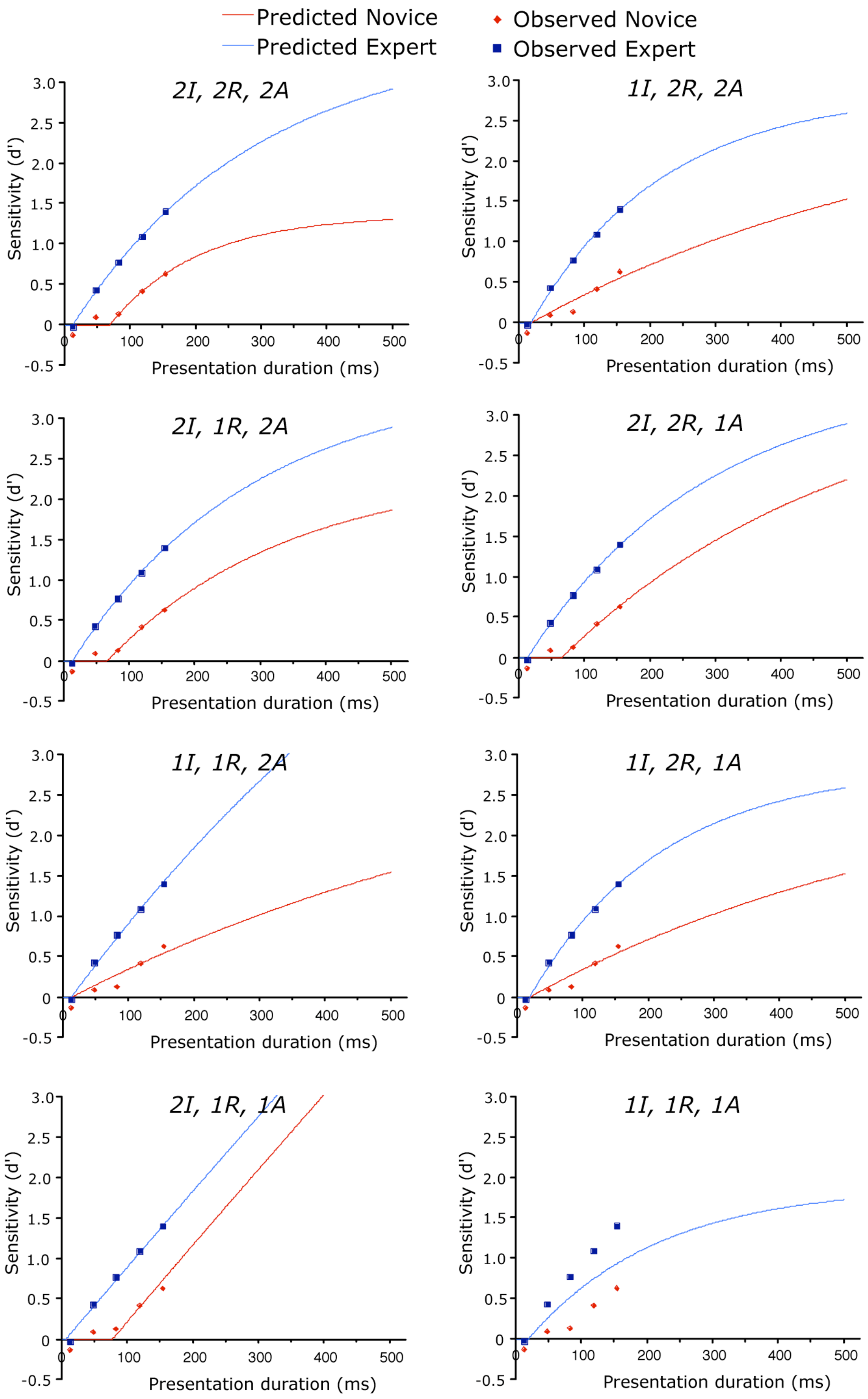


Figure 23. The curves that best fit the observed average d' values from the five shortest encoding durations for cars among car experts and novices for each of the eight possible models derived from the shifted exponential function. The models differed in terms of whether the intercept (*I*), rate of approach to asymptote performance level (*R*), and the asymptote (*A*) parameters were fixed or free to vary (notated as a 1 or 2, respectively) for the curves fitted to the performance for experts and novices across the different presentation duration conditions.

Table 7. Best fits of the Exponential Function to the Five Shortest Presentation Durations in Experiment 9.

Condition	Intercept	Rate	Asymptote	$r^2_{\text{(adjusted)}}$
2 I, 2 R, 2 A				
Expert	0.012	3.5268	3.5595	
Novice	.067961	7.4053	1.3589	
				.98782 (.972595)
1 I, 2 R, 2 A				
Expert	.018135	4.9857	2.8545	
Novice	.018135	1.5943	2.8545	
				.97832 (.96098)
2 I, 1 R, 2 A				
Expert	.012	3.6029	3.4985	
Novice	.065146	3.6029	2.3628	
				.98773(.977914)
2 I, 2 R, 1 A				
Expert	.012	3.5876	3.5105	
Novice	.063987	2.2743	3.5105	
				.98766 (.977788)
1 I, 1 R, 2 A				
Expert	.012	1.2238	9.0337	
Novice	.012	1.2238	3.4476	
				.97579(.963685)
1 I, 2 R, 1 A				
Expert	.018135	4.9857	2.8545	
Novice	.018135	1.5943	2.8545	
				.97832 (.96098)
2 I, 1 R, 1 A				
Expert	.0049202	.01305	73.3167	
Novice	.076013	.01305	73.3167	
				.97618 (.96427)
1 I, 1 R, 1 A				
Expert	.014067	.046848	155.3646	
Novice	.014067	.046848	155.3646	
				.64008(.53725)

Note: The models differed in terms of whether the intercept (*I*), rate of approach to asymptote performance level (*R*), and the asymptote (*A*) parameters were fixed or free to vary (notated as a 1 or 2, respectively) for the curves fitted to the performance for cars among car experts or novices across the different presentation duration conditions.

Discussion and combined analysis of Experiments 8 and 9

The striking similarity between the pattern of behavioral results for cars among car experts and novices in Experiment 9, and that for upright and inverted faces in Experiment 8, is consistent with an expertise account of the difference in the processing time-course between these groups (i.e. the shorter encoding duration required for recognition performance to exceed chance for objects of expertise). In neither experiment was there evidence for a difference in the rate of approach to the asymptote. Thus, expertise selectively impacted the general level of performance that can be reached (asymptote) and the onset time of performance. However, expertise did not influence the rate of performance increase that took performance between these two points in the time-course of object processing. These results are consistent with the temporal advantage among visual experts, relative to novices, previously demonstrated in subordinate-level recognition tasks by Tanaka and colleagues (Tanaka, 2001; Tanaka & Taylor, 1991) and suggest that this advantage occurs relatively early in the time-course of processing.

Distinguishing between the fits of the curves generated by alternative models of the time-course of expert and novice processing to the observed data was complicated by the large difference between the general level of performance between experts and novices. These large differences in later parts of the curves caused smaller, but important, differences in earlier parts of the curves to have a reduced impact on the overall fit of the curve to the data. In other words, detecting a difference between the onset times between the expert and novice groups was hindered by the fact that the difference in asymptote accounted for the largest portion of variability between the two groups. This made it difficult to distinguish models that constrained other important aspects of the time-course. For example, the curve resulting from the model in which experts and novices are assumed to have the same rate of approach to the asymptote, but different asymptote and onset times, could not be distinguished from the model in which experts and novices had the same performance onset time, but different rates of approach to asymptote and different asymptotes. It was only once the fit was limited to the shorter presentation durations, before performance started to plateau, that the fit of these two models started to diverge, allowing the models to be distinguished based on the plausibility of the parameter estimates and the fit of the curve adjusted for the number of free parameters.

The impact of the difference in the performance onset time among car experts, compared to car novices, may also be reduced by averaging the data across participants with a range of perceptual expertise with cars, as the level of car expertise varies considerably within the two groups broadly classified as "experts" and "novices". More specifically, if the delay in performance onset is associated with visual expertise with a particular category of objects, the degree that performance onset is delayed would be expected to vary with the level of perceptual expertise of the observer. However, since the level of expertise with cars varies considerably within the two groups this may serve to blur the sudden onset assumed by the shifted exponential function fitted to the data (Van Zandt & Ratcliff, 1995). This provides a potential explanation for the greater difficulty in distinguishing the rate and intercept constrained models in Experiment 9 compared to those in Experiment 8, as there is likely to be a clearer distinction between participants' level of expertise with upright and inverted faces.

Unfortunately, the data from Experiment 9 was too noisy at the individual level to explore the relationship between participants' level of expertise with cars and the onset time of their performance. The noise at the individual level was a result of the relatively small number of trials (20) contributing to each data point. This was a necessity of the current design, as the potential stimulus pool within a typical car experts' domain of expertise (i.e., different models of modern cars readily available in North America) is too limited to support more than 20 trials per encoding condition without repeating stimuli. Future studies could overcome these limitations to explore the relationship between performance onset time and level of expertise by either using fewer encoding conditions (thus allowing for a greater number of trials per condition) or by using a different domain of expertise with a less limited set of potential stimuli.

Combined analysis. To explore the extent of the similarity in the results from Experiments 8 and 9, an analysis was performed on the combined data. The main aim of this combined analysis was to determine if the effect of expertise (or orientation for the case of faces) differed across the recognition of faces and cars. To do this, data from Experiments 8 and 9 were pooled: participants in the inverted face condition were re-classified as novices, and those in the upright face condition were classified as experts. A 2 (Category; face, car) x 2 (Group; expert, novice) x 8 (presentation duration; 12-, 47-, 82-, 118-, 153-, 235-, 494-, or 1000 ms) ANOVA found a main effect of expertise, $\underline{F}(1,68)=83.89$, $\underline{p}\leq.0001$, and presentation duration, $\underline{F}(7,476)=77.63$, $\underline{p}\leq.0001$, but not category, $\underline{F}(1,68)=2.29$, $\underline{p}=.092$. There was an interaction between category and presentation duration, $\underline{F}(7,476)=2.26$, $\underline{p}=.029$, and expertise and presentation duration, F(7,476)=6.34, p≤.0001, but notably there was no interaction between category and expertise (F<1) or expertise, presentation duration, and category, (F<1). Therefore, although car recognition across both experts and novices benefited more from additional encoding time compared to that for faces, there was no other difference between the recognition of faces and cars. Most importantly the effect of expertise (or inversion for faces) did not differ for cars and faces. The results of this combined analysis are consistent with a common expertise-related source for the different recognition time-courses between upright and inverted faces and for cars between car experts and novices. Possible underlying mechanisms for this time-course difference between expert and novice recognition are addressed in the general discussion.

General Discussion

The studies in this chapter provide evidence that visual expertise results in a difference in the time-course of encoding. The temporal advantage demonstrated by experts is generally consistent with findings from previous behavioral and electrophysiological studies (Caldara et al., 2003; Rossion et al., 2000; Tanaka & Taylor, 1991). More specifically, the difference in the time-course of processing for expert and novice observers appears to stem from the delayed onset of novice processing relative to expert processing. The difference in the onset time of performance in the observed data for expert (upright faces, and cars among car experts) and novice (inverted faces, and cars among car novices) categories suggests that experts experience a "head start" to encoding over that for novices. Notably, this finding of a delayed onset of performance is also captured by the shifted exponential curve fitted to the data from trials with inverted faces, or cars among car novices, relative to those with upright faces, or cars among car experts. This finding extends previous work not only by providing a more thorough mapping of the time-course of expert and novice object processing, but also by demonstrating that the temporal advantage demonstrated previously in electrophysiological and response time studies is a result of an earlier onset of performance among experts, relative to novices, rather than a more general difference in processing rate between the groups. This section will address some potential limitations and also speculate about the underlying cause of the difference between the time-course of expert and novice processing.

The choice of a shifted exponential function to fit to the observed data was theory-neutral and thus the possibility remains that the observed performance may be better characterized by a different performance function. The almost perfect fits (> 98% of the variance account for) obtained with the full models in both Experiments 8 and 9, though, suggest that this function could adequately explain most of the variability in performance between experts and novices. However, given that this curve was used only to provide a quantitative description of the relationship between encoding duration and recognition performance among experts and novices, the interpretation of the data relies on both the results of the model and the statistical analyses performed directly on the behavioral data. Convergence of the results from the curve fitting procedure (Tables 3, 6, and 7) with those from the statistics performed directly on the observed data (Tables 1, 2, 4, and 5) serves to strengthen the somewhat weaker results from the curve fitting procedure, especially those in Experiment 9.

The results of the experiments in this chapter also provide some insight into the reliance of experts' greater VSTM performance on sufficient encoding time reported in the previous chapters of this dissertation. One of the goals of this chapter was to explore whether the VSTM advantage comes at a cost in the form of an additional burden on perceptual encoding mechanisms, serving to slow down the rate at which experts, relative to novices, can encode objects of expertise. This proposal was inconsistent with the faster response times of experts' for subordinate-level recognition judgments reported by Tanaka & Taylor (1991), which suggests instead that experts may process items at a *faster* rate. However, the results reported in this chapter are inconsistent with both of these possibilities, but rather suggest that expert and novice encoding proceeds at a similar rate (as inferred from the rate of increase in performance with increasing encoding duration). Therefore, the reliance on sufficient encoding time for the VSTM advantage of experts relative to novices is likely due to limitations at other processing stages involved in storing objects in VSTM (e.g., the capacity-limited mechanisms responsible for consolidating items in VSTM; Jolicoeur & Dell' Acqua, 1998).

More generally, a noteworthy implication of the results in this chapter is that the rate of approach to the asymptote with increasing encoding time is unaffected by the general costs incurred when processing objects in unfamiliar orientations (Yin, 1969). Taken together, the results of Experiments 8 and 9 also suggest that the rate of performance increase is unaffected by any benefits associated with visual expertise. Therefore, both expert and novice encoding rate may be limited by the same hard-wired physiological constraints of the visual system.

The earlier onset of performance for subordinate recognition of objects of expertise, compared to objects from categories for which one does not possess expertise, is intriguing. As discussed in the opening section of this chapter, previous studies have found evidence of a general temporal advantage for expert processing, but until now there was no clear evidence to specifically suggest that this advantage may be rooted in a difference in the encoding time required for recognition to begin. Tanaka (2001), when explaining the faster reactions times for subordinate-level categorization for objects of expertise relative to typical objects, suggested that an important consequence of visual expertise is the development of specialized “perceptual routines” that allow for the rapid analysis of domain-specific objects. Unfortunately, the nature of this “perceptual routine” and how it differs from that used for typical object processing was not elaborated. Insights might be gained by considering how such routines might relate to the differences in recognition onset time revealed here. Four alternative accounts of the effect of expertise on the onset time of recognition performance will be discussed in the remaining part of this chapter.

The first possibility is that this difference in onset reflects a cost among novices due to implementing a processing mechanism that occurs automatically among experts. Expertise may rely on highly specialized patterns of attentional allocation that allow for the extraction of the most diagnostic information, whether it be the precise configuration of features within a face or the subtle contours of a car. Recent work by Waszak and colleagues (2003) demonstrated that objects are influenced by their processing history, leading to the relatively automatic processing of items in a manner consistent with this history. Therefore, it is possible that this attentional weighting of diagnostic features may be performed automatically for objects of expertise as a result of the extensive expertise identifying these objects at a subordinate level. The automatic triggering of an optimal attentional pattern may underlie the temporal advantage of expert performance when compared to that of novices, who may have to effortfully establish an attentional set to selectively process or weight diagnostic features for each new object. Therefore, a learned pattern of attention to the information in faces and other objects of expertise may provide a temporal advantage in the form of a “head start” on encoding for processing objects within one’s domain of expertise.

Alternatively, the delay in the onset of performance for inverted faces, and for cars among car novices, may reflect a cost associated with the categorization of these items at a basic-level *before* categorization at the more subordinate level required for the task in Experiment 8 and 9 (Liu, Harris, & Kanwisher, 2002). The basic-level has been shown to be the ”entry-point” of recognition among novices, and thus novices may be obliged to first categorize the item at this level before categorizing the object at more subordinate levels. This obligatory categorization at the basic-level first may be costly as the information that is most diagnostic for categorizations at different levels of specificity is likely to vary considerably. For example, in order to recognize that an object is a car, an important feature that should be weighted heavily for this recognition judgment might be the wheels. However, in order to recognize that the object is a Honda Accord, heavily weighting the wheels will be of little value, and weighting other features such as the contours of the hood is necessary for categorization at this more subordinate level. Therefore, obligatory categorization at the basic-level by novices may result in a delay in performance in subordinate categorization tasks. In contrast, visual experts may be able to more directly encode the perceptual information required to identify objects from their domain of expertise at a subordinate-level, as these levels have been shown to be equally accessible to experts, and thus experts should not experience the same delay (Tanaka & Taylor, 1991; Tanaka, 2000). This would lead to the empirically testable prediction that the difference in the onset of performance between car experts and novices should no longer be present for basic-level categorizations. Future studies should test this prediction.

A third explanation for the difference in the time-course between expert and novice object recognition is a greater influence of top-down mechanisms on perception. More specifically, participants with considerable experience recognizing cars may benefit more from higher-level areas acting on earlier areas responsible for the encoding of visual sensory information in anticipation of the stimulus. This could be in the form of expectations serving to prime areas responsible for expert object recognition, although this would have had to occur at a very general level, as no stimulus was repeated during the experiment. Consistent with this possibility, Johnson and Olshausen (2005) suggest that top-down facilitation could lead to different onsets in performance, such as that found in Experiment 8 and 9: information may accumulate faster if an observer has a top-down template to which the incoming sensory information can be compared. The use of a template could serve to lower thresholds involved in perceptual decisions, or to amplifying the response to information in the template, thereby providing a temporal advantage to the accumulation of information for expert object recognition. However, one might expect that a top-down template would also result in a difference in the rate of performance, which was not found in Experiments 8 and 9. One aspect of the potential influence of top-down mechanism on the earlier performance onset among experts could be explored by comparing performance in conditions where trials with objects of expertise were blocked, compared to those where they were randomly intermixed with stimuli from other (non-expert) object categories, thereby reducing participants ability to anticipate the stimulus.

Finally, experience may influence the time-course of object recognition through less transient influences on the object perception system. Experience can bias population selectivity of neurons so that more cells respond to the more familiar stimuli and views (Ashbridge, Perrett, Oram, & Jellema, 2000) or influence the selectivity of neurons (Booth & Rolls, 1998; Kobatake, Wang, & Tanaka, 1998; Logothetis, Pauls, & Poggio, 1995; Sheinberg & Logothetis, 2001). This could provide an account of the results in Experiment 8 and 9– by integrating both category and orientation biases – if one assumes that greater numbers of active neurons increase the speed with which a threshold could be reached to trigger recognition. However, this account might also predict that a faster rate of increase in performance would follow the earlier onset, not found in Experiments 8 and 9. Further studies should explore the possible influence of these changes at a neural level on the time-course of object processing.

In summary, these studies provide a window into the temporal dynamics of expert visual processing. Most notably, the results reported in this chapter extend previous work by suggesting that the expert temporal advantage suggested by previous behavioral and ERP studies relies on a difference in the onset time of performance rather than a faster rate of approach to the asymptote performance level. These results also suggest that the reliance on sufficient encoding time for the expert VSTM advantage to emerge is not a result of experts' encoding objects more slowly. More specifically, the results of the two experiments reported in this chapter suggest that expert visual processing has approximately a 35-55 ms "head start" over novice processing. Four alternative explanations for this difference in the time-course of expert and novice recognition were suggested, which serve to provide a springboard for many exciting future research directions and thus they also provide a fitting end to the last empirical chapter in this dissertation.

These future directions and their more general implications will be discussed in more detail in the concluding chapter.

CHAPTER V

CONCLUDING REMARKS: IMPLICATIONS AND FUTURE DIRECTIONS FOR EXPLORING THE INFLUENCE OF PERCEPTUAL EXPERTISE ON CAPACITY LIMITATIONS

Summary and overview

The aim of this dissertation was to augment our understanding of the determinants of VSTM capacity for complex objects. In particular, the focus was on the influence of encoding biases that arise through the development of visual expertise, namely the tendency to process objects holistically rather than in a more feature-based manner. This was a unique approach as previous studies have not considered the influence of encoding strategy on VSTM capacity, and they have focused mainly on stimulus-based rather than observer-based effects on VSTM capacity. Expertise-related changes to the way objects are encoded may be especially influential because of their potential to have a cascading effect on the performance of other systems, such as VSTM, that rely on the output from early perceptual processing.

Face perception is often compared to expert perception of other objects (Diamond & Carey, 1986; Gauthier et al., 2003), both of which involve holistic encoding as well as advantages such as rapid identification of objects at the individual level (Tanaka, 2001). Holistic encoding is believed to confer an advantage to the perception of faces, making face processing less susceptible to feature changes like a new hair style or the use of disguises and increasing our sensitivity to minute configural differences between faces (Diamond & Carey, 1977; Tanaka & Sengco, 1997). The results presented here suggest that holistic processing also leads to a greater VSTM capacity as long as there is sufficient time to encode the items.

This dissertation began (CHAPTER 1) by building an important bridge between the VSTM literature and the literature on the role of holistic processing in the perception of faces and objects of expertise (Gauthier et al., 2003; Gauthier et al., 2000; Gauthier & Tarr, 1997; Tanaka & Gauthier, 1997). This bridge provided the motivation for investigating the consequence of encoding strategy on VSTM capacity. Although some previous studies have suggested that expertise influences VSTM, such effects have been thought to arise from changes in the ability to store and use information in long-term memory as well as the use of explicit mnemonic strategies (Ericsson & Kintsch, 1995). However, these strategies are limited in their utility to increase VSTM capacity, relying on the ability to “chunk” independent items together with semantic information, and thus they are generally useful only for familiar items between which meaningful connections can be made. In addition, the advantage gained by utilizing these strategies is believed to arise from the contributions of other systems and thus does not reflect a change within the VSTM system itself. The work in this dissertation differed from previous work by focusing on how the changes at *earlier* perceptual levels, specifically those that result from the development of visual expertise, impact VSTM capacity.

To explore the influence of encoding strategy on VSTM capacity, CHAPTER 2 compared VSTM for holistically encoded upright faces to that for more featurally encoded inverted faces and upright non-face objects. VSTM for faces experienced a cost due to inversion, an advantage over non-face categories, and greater sensitivity to encoding time limitations. These results provided evidence that the specialized orientation-specific encoding strategies recruited for faces provide an advantage to VSTM. In addition, the greater sensitivity to encoding time limitations for upright face VSTM raised the possibility, explored in CHAPTER 4, that these qualitatively different encoding strategies follow different processing time-courses.

To provide a stronger test of the hypothesis that perceptual encoding strategy can influence VSTM capacity, CHAPTER 3 compared VSTM for upright and inverted faces and cars among car experts and novices. The similarity between the pattern of results for cars among car experts and that for faces was notable, with car experts showing not only a VSTM advantage similar in magnitude to that found with faces, but also an inversion cost and a need for sufficient encoding time to demonstrate these effects. Importantly, car novices did not show an inversion cost for cars, providing strong evidence against the possibility that this effect on VSTM is due to a more general cost to processing objects in non-canonical orientations rather than a change in encoding strategy. In addition, a participant’s VSTM for cars was correlated with their visual expertise with this category. Therefore, the results reported in Chapter 3 not only replicated the finding of a VSTM advantage for objects of expertise; they also were consistent with the possibility of a holistic processing account of this advantage.

It must be noted that the implications of the results of Chapter 3, in terms of a holistic processing account of the expert VSTM advantage, are limited in the sense that they provide *indirect* evidence that a holistic encoding style may increase VSTM capacity. A more direct test of the role of holistic processing in increasing VSTM for objects could have manipulated the encoding strategy recruited, rather than manipulating the object (inversion) or comparing between observers (experts vs. novices). However, there are many difficulties associated with trying to manipulate a viewer's encoding strategy, especially that for faces and other objects of expertise. For example, numerous paradigms have demonstrated that expert encoding strategies tend to be impervious to task instructions. Despite asking viewers to make a same/different judgment based on only a part of a face, or car in the case of car experts, viewers are unable to ignore the rest of the object (Boutet et al., 2002; Farah et al., 1998; Gauthier et al., 2003; Young et al., 1987).

One example of an attempt to manipulate encoding strategy more directly comes from Farah, Tanaka, and Drain (1995), who attempted to manipulate an observers' processing strategy to encourage either more holistic or more featural encoding of faces. To do so they changed the format of the faces; participants studied upright faces in one of two conditions, either in an intact (holistic inducing) format or a format where the features were spatially separated to induce a more feature-based encoding style. The authors suggested that the format in which the items were studied influenced the level of holistic processing and thus influenced the size of the inversion cost in face performance. However, this study is unique in the literature and may be open to alternative explanations.

The results of Farah, Tanaka, and Drain (1995) are also, to some extent, inconsistent with data from other studies. For example, Gauthier, Tanaka, & Brown (submitted) found that even gross distortions to the format of a face (e.g. placing the top and bottom halves of faces side by side) during the initial study phase had little effect on the level of holistic processing of faces, providing that the face appeared intact during the test phase. This finding contradicts the results of Farah, Tanaka, & Drain (1995), which had suggested that study format can influence holistic processing of faces. Therefore, it is questionable whether it is truly possible to manipulate (at least in any robust manner) the holistic processing of upright faces or other object of expertise by encoding format and/or task instructions.

The expertise manipulation used throughout this dissertation also better addresses the dissertation's specific aims. One particular goal was to investigate whether the changes in processing strategy resulting from the development of visual expertise lead to a domain specific VSTM advantage. Therefore, it was necessary to recruit (or train) participants with extensive visual expertise. In addition, the use of a strategy manipulation like Farah and colleagues' requires manipulating the actual stimuli to induce the change in strategy. Given the recent studies highlighting the importance not only of object complexity (Alvarez & Cavanagh, 2004; Eng et al., in press), but also of object form (Xu, 2002a) in determining VSTM capacity for objects, it was important to hold the stimuli constant. Otherwise, any interpretation of the results in terms of encoding strategies would be confounded by the physical differences between the stimuli in the different conditions.

Comparing experts and novices, known from previous work to recruit different encoding strategies, provided numerous advantages over designs where the strategy of observers might be manipulated by changing the encoding format. Not only did it allow for interpretations of the influence of encoding strategy on VSTM to extend to real-world expertise, but it also eliminated alternative explanations based on object structure or complexity. In addition, this design also allowed for stronger testing of the hypothesized role of holistic encoding by regarding level of expertise—and thus degree of holistic encoding (Gauthier et al., 2003)—as a continuous variable among the observers within the sample. In contrast, although strategy manipulations are also likely to vary in effectiveness across individuals, it would be difficult to harness this variability in such a way that it could be used to test specific predictions about the relationship between holistic encoding and VSTM capacity. In addition, more transient strategy manipulations, when successful, are likely to be smaller and less stable compared to the strategy differences developed between experts and novices over the course of many years. Therefore, an expertise manipulation rather than a more direct manipulation of encoding strategy was not only necessary to address the specific aims of this dissertation, but was also a more reliable and powerful method of exploring the influence of encoding strategy on VSTM capacity.

Before any strong conclusions could be made about the role of encoding strategy in influencing VSTM capacity, alternative accounts for the greater VSTM capacity among experts, compared to novices, needed to be addressed. The control studies conducted in CHAPTER 3 were able to eliminate three alternative explanations; a contribution from stimulus-specific representations in long-term memory, a contribution from verbal short-term memory, and a difference in the eye movement strategies employed to encode the items in VSTM. Therefore, based on the data in this dissertation, the well-documented differences in cognitive encoding strategy employed by experts and novices appears the most likely account for the VSTM advantage demonstrated by experts.

Following up on the evidence provided in CHAPTERS 2 and 3 that the expert VSTM capacity advantage depends on sufficient encoding time, CHAPTER 4 more closely compared the time-course of object recognition for upright and inverted faces and for cars among car experts and novices. The strong reliance of expert VSTM performance on sufficient encoding time had suggested that the temporal dynamics of holistic and more feature-based encoding strategies might differ. The studies reported in this chapter revealed an initial delay of approximately 35- 55 ms before performance rose above chance-level for more featurally processed items (i.e. cars among car novices or inverted faces) compared to more holistically processed items (i.e. cars among car experts and upright faces). This finding suggests that experts experience a "head start" over novices. This head start may underlie the generally superior performance of visual experts in many perceptual tasks, however it cannot account for the dependence on sufficient encoding time to demonstrate a VSTM advantage.

It is possible that the emergence of the expert VSTM advantage *only* with sufficient encoding time may reflect a bottleneck at another stage of processing. For example, Jolicoeur and Dell' Acqua (1998) demonstrated that the process of consolidating items into VSTM is also capacity-limited. If visual expertise selectively benefits the efficiency of perceptual representations for objects of expertise, but not other capacity-limited processes such as those responsible for consolidating information into VSTM, experts may experience a greater bottleneck at consolidation due to the increased number of objects of expertise that need to be consolidated into VSTM. Further studies should explore this possibility.

Implications, speculations, and future directions

Visual short-term memory

The question explored in this dissertation—whether visual expertise and the resulting change in encoding strategy influences VSTM capacity—has important consequences not only for our understanding of the nature of expertise, but also for our understanding of VSTM more generally. A potential influence of encoding strategy on VSTM capacity has far reaching consequences. Previous accounts of VSTM have suggested that VSTM capacity is determined by a hard-wired limit, with the variability in this limit arising from inherent differences between individuals (Luck & Vogel, 1997; Vogel & Machizawa, 2004). In contrast, the demonstration that experience, possibly through a change in encoding strategy, can impact VSTM suggests that VSTM is flexible depending partly on the outputs of other systems. Therefore, VSTM capacity not only differs across categories (Alvarez & Cavanagh, 2004), individuals (Todd & Marois, 2005; Vogel & Machizawa, 2004), and with task constraints (Eng et al., in press), but also with experience-based perceptual skill.

One of the implications of these findings is that there is potential for training to benefit this central cognitive function. Although practice on VSTM tasks has not been found to improve VSTM capacity substantially (Chen et al., in press, as discussed in CHAPTER I), the results in this dissertation suggest that expertise training procedures that have been shown to increase holistic processing (Gauthier & Tarr, 1997) are more likely to impact our VSTM capacity within the trained domain. Thus, future studies should explore whether VSTM capacity changes, along with markers of holistic processing, over the course of lab-based expertise training programs.

One possibility suggested by the results of these studies is that the development of visual expertise may, in effect, reduce the perceived perceptual complexity of objects of expertise. In this way, visual experts may be able to circumvent the limitations involved in encoding and storing complex objects in VSTM. The lower VSTM capacity for faces compared to non-faces objects (i.e. cars and watches) at short encoding times, which was demonstrated in Experiment 2 (CHAPTER 2), suggests that the faces had a greater information load than the cars or watches (Alvarez & Cavanagh, 2004). Consistent with demonstrations that extending encoding duration can, in part, reduce the effect of perceptual complexity for more complex items (Eng et al., in press), additional encoding time eliminated the disadvantage for faces relative to the other categories. When encoding time was extended even further (Experiment 3, CHAPTER 2), VSTM for faces exceeded that for the other categories of objects. Notably, the capacity for faces was in the range of that reported for very simple objects, such as colored dots. Consistent with the suggestion that expertise may serve to reduce the impact of perceptual complexity on VSTM capacity, the benefits of expertise for VSTM capacity never allowed it to exceed that for simple objects, for which VSTM capacity is presumably not limited by perceptual complexity. This finding suggests that expertise provides an *additional* benefit over that experienced with extended encoding time for objects outside one's domain of expertise. This allows VSTM for complex objects of expertise to approach that for more simple objects, such as colored dots.

An important consideration is that the information load or complexity of an object is at least in part determined by the context within which it appears. Increased homogeneity between items in a search array increases visual search rate and thus the information load of the items within the array (Duncan & Humphreys, 1989). Therefore, expertise may serve to reduce the information load/complexity of objects of expertise by decreasing the perceived level of homogeneity between the items within the memory array. This possibility could be explored in further studies by manipulating the level of homogeneity between items within the study array and measuring the resulting impact on expert and novice performance.

It is important to acknowledge that although these results have promising implications for the flexibility of VSTM and even suggest specific ways to increase VSTM capacity, they do not provide a "magic bullet" for VSTM capacity limitations. Expertise training paradigms require hours of practice and the changes that result are not only domain-specific, but also orientation-specific. In addition, such training would be expected to reduce the effects of object complexity, and thus only allow an observer to maximize the use of their inherent VSTM capacity. Hence, training would be expected to have a reduced effect on VSTM for very simple objects that can easily be distinguished, as capacity for these objects should not be limited by object complexity (Alvarez & Cavanagh, 2004). Therefore, the inherent object or 'slot' based limit of VSTM likely provides an upper limit to the benefits of a holistic encoding strategy (Luck & Vogel, 1997).

The studies in this dissertation extend insights about VSTM for complex objects. Much of the work exploring VSTM capacity has used very simple objects such as colored dots (Luck & Vogel, 1997; Todd & Marois, 2004), and it has been unclear how well such findings extend to VSTM for more complex items, such as faces or other real-world objects. Such limitations were partly addressed by Alvarez and Cavanagh (2004), who used line drawings of familiar objects as stimuli. However, they only used brief stimulus presentation durations (500 ms) and thus did not consider the importance of sufficient presentation time for allowing encoding of complex information.

Recently, Eng and colleagues (in press) included complex objects, e.g. pictures of three-dimensional faces, as stimuli in their study comparing VSTM across categories. These researchers focused on the role of presentation time in reducing the effect of object complexity on VSTM capacity reported by Alvarez and Cavanagh (2004), but they did not consider the role of encoding strategy in influencing this effect. The studies in the current dissertation demonstrate that VSTM not only differs across orientations of the same object category, but also within orientations and object categories depending on the experience of, and thus the encoding strategy used by, the observer. Therefore, they extend the findings of previous studies by demonstrating that differences in VSTM are not entirely determined by perceptual complexity; they also differ due to experience.

The results reported in this dissertation also 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 & Poom (2005) found that with 500 ms of encoding time, participants had a VSTM capacity for intra-categorical geometric shapes (e.g., ovals with varying aspect ratios) of only a single item. Based on this finding, they suggest that the performance in previous VSTM studies reporting a VSTM capacity of 3-4 objects (e.g. Luck & Vogel, 1997; Vogel et al., 2001) was facilitated by categorical structures in LTM. Specifically, they suggest that such a benefit arises from the use of stimuli in such studies that cross category boundaries (e.g. a red and a yellow square cross a color boundary). One factor that was not considered in this study is that, when multiple objects need to be retained, objects within the same category may require more time to be encoded into VSTM than objects that cross category boundaries. Notably, the studies in this dissertation not only used a match-to-sample paradigm, but the faces used were unfamiliar and thus had no known labels and shared the same category boundary. Therefore, according to Olsson and Poom (2005), 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 (2005) study; the findings of this dissertation and those of Eng et al. (in press) suggest that VSTM capacity for complex objects is underestimated with 500 ms of encoding time because of perceptual encoding limitations. Consistent with this possibility, in this dissertation, VSTM capacity for faces with 500 ms encoding time was found to be a little over one. It is possible that capacity for the geometrical objects used by Olsson & Poom (2005) could reach that reported for watches, for instance, given enough encoding time.

This dissertation has important implications for developmental studies of VSTM capacity. In particular, the results offer an explanation for conflicting findings in the literature with respect to developmental changes in VSTM capacity. While a number of studies have demonstrated adult-like VSTM capacity (3-4 objects) in infants by the time they reach 10-12 months old (Rose et al., 2001; Ross Sheehy, Oakes, & Luck, 2003), others report changes in VSTM in school age children (Meyler & Breznitz, 1998). It is possible that these different studies may be measuring different aspects of VSTM capacity. The adult-like limit reported in infants may reflect the hard-wired limit of VSTM, whereas developmental differences in VSTM capacity among school-aged children might reflect the benefits of experience for the encoding and storage of complex information. Consistent with this suggestion, studies reporting adult-like VSTM capacity in infants either used very simple objects, such as colored squares (e.g. Ross Sheehy et al., 2003), or real-world objects from different categories that differed along many dimensions and thus typically require encoding of only a single dimension to be distinguish (Rose et al., 2001). In contrast, studies reporting developmental changes in VSTM performance used more complex tasks such as the Bead Memory Subtest of the Standford-Binet Intelligence Scale (Thorndike, Hagen, & Sattler, 1986), which requires participants to reproduce bead patterns from memory. The use of a more complex task to measure VSTM performance may allow for a greater role of developmental change and/or experience to influence the efficiency of encoding. According to this account, VSTM for faces may also be expected to change over the first decade or so of life because markers of holistic processing, such as the effect of inversion or configural effects, are not at adult like-levels until up to 14 years of age (Carey et al., 1980; Mondloch, Le Grand, & Maurer, 2003). Although experience may not change the absolute capacity of VSTM, the development of more efficient encoding strategies could result in the smaller influence of stimulus complexity or homogeneity seen among visual experts, thus accounting for the developmental changes reported in school age children. Therefore, the apparent inconsistency between the different developmental studies may arise because the studies are actually tapping into different aspects of VSTM capacity limitations, the hard-wired limit versus the flexibility in the manner in which complex information can be stored in VSTM.

The demonstration of an influence of encoding style on VSTM capacity opens up many possibilities. For example, how much do other encoding strategies influence VSTM? Is there an encoding strategy that could produce a domain general change to VSTM capacity? For example, could a contextual encoding strategy, that is, a strategy in which individual items are encoded predominantly in terms of their relations with their environment, impact VSTM? It is unlikely that holistic processing is unique in its ability to influence VSTM capacity, and other changes to the way information is encoded in VSTM could potentially impact VSTM capacity. Consistent with this possibility, it has been reported that there is a bi-directional relationship between learning to read and VSTM: not only is learning to read facilitated by good pre-reading VSTM scores, but it also helps VSTM to develop (Meyler & Breznitz, 1998). These findings provide further evidence for the role of learning and experience in facilitating VSTM capacity.

In summary, the results presented in this dissertation suggest that although VSTM has an inherent upper limit, within these bounds encoding time, perceptual complexity, and visual expertise act in concert to determine the capacity of VSTM. Most importantly these findings highlight the dependency of VSTM on earlier perceptual processes, which in themselves can be influenced by visual expertise. Expertise might reduce the effects of perceptual complexity on VSTM by increasing the efficiency of storage of information; such mechanisms would presumably provide less of a benefit for more simple items that can be easily distinguished, as there would be minimal flexibility for such information to be further compressed.

Visual expertise and the time-course of object recognition

The results reported in this dissertation are consistent with the notion that face processing and the processing of non-face objects of expertise have many properties in common (Diamond & Carey, 1986; Gauthier et al., 2003; Gauthier et al., 2000; Tanaka, 2001; Tanaka & Curran, 2001; Tanaka & Taylor, 1991; Xu, 2005). For example, VSTM for upright faces, relative to inverted faces, and cars among car experts relative to car novices, not only demonstrated a VSTM capacity advantage that was similar in magnitude, but this advantage also emerged only with sufficient encoding time (4000 ms) in both cases. The results support the possibility of a cascading effect, where expertise-related changes at a perceptual level impact other cognitive functions involving objects of expertise, namely VSTM.

Despite the relationship between the VSTM advantage among experts and level of visual skill at identifying cars, it is important to acknowledge that such findings cannot speak to the issue of causality. In addition, even though previous studies have found a relationship between the same measure of skill at recognizing cars and holistic processing measures for cars, this can only be interpreted as suggesting that they are somehow related. It is possible that they both have a common underlying cause rather than a causal relationship (Gauthier et al., 2003; Gauthier et al., 2000). There are also some obstacles to linking holistic processing for cars and the expert VSTM advantage even though both are correlated with visual expertise with cars, as it is possible that they account for different aspects of an individual's expertise with cars. Therefore, a link between these two aspects of expert performance is hypothesized with some caution. Futures studies could examine this proposed relationship by more directly measuring the correlation between the level of holistic processing of cars and the VSTM capacity for cars.

The results of this dissertation do, however, provide direct evidence relevant to a major current debate in the literature. Specifically, McKone and colleagues (McKone & Kanwisher, 2005; McKone & Robbins, 2005) have argued against the expertise account of face processing, particularly against evidence for an inversion effect for objects of expertise. They suggest that previous studies demonstrating such an effect were either flawed (e.g. Diamond & Carey, 1986) or that the inversion was too small compared to that for faces to warrant the supposition of a common underlying mechanism (Gauthier et al., 2000; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002). In addition, a recent study by McKone and Robbins (2005) failed to replicate Diamond and Carey's (1986) original finding of a larger inversion effect for dogs among dog experts compared to dog novices. However, in CHAPTER 3, a large cost of inversion to the recognition for cars among car experts, but not novices, was demonstrated and replicated. Importantly, the effect of inversion differed reliably across categories for novices but not experts in both Experiments 4 and 5.[10] Therefore, the studies in this dissertation provide evidence for the existence of a large inversion effect similar in magnitude to that demonstrated for faces, refuting McKone and Kanwisher's (2005) suggestion that such an effect does not exist for objects of expertise.

The work in this dissertation not only provides evidence relevant to this major debate, but it also provides additional insights into the nature of face and non-face object expertise. Most notably, CHAPTER 4 provides evidence that expertise selectively influences certain aspects of the time-course of object processing: whereas the onset of recognition performance occurred earlier for experts than novices, the rate of increase in performance was unaffected by a participant's expertise. This potentially provides important constraints for explanations of the temporal advantage previously demonstrated for the processing of objects of expertise.

One interesting possibility is that the difference in the time-course of processing for objects of expertise, relative to those for which a participant does not possess expertise, is rooted

---

[10] Experiment 4; novices, $F(1,17)= 5.2868$, $p = .0344$, experts, $F < 1$; Experiment 5; novices, $F(1,16) = 13.848$, $p = .0019$, experts, $F<1$)

in fundamental differences in the nature of visual information processing. More, specifically, there is evidence that processing different spatial frequency information follows different time-courses. The lower resolution visual information carried by low spatial frequencies is believed to be processed more rapidly relative to the higher resolution visual information carried by high spatial frequencies. For example, under speeded conditions, recognition performance is better for images containing only low spatial frequency information than for those containing only high spatial frequency information (Hughes, Nozawa, & Kitterle, 1996; Robertson, 1996; Shulman, Sullivan, Gish, & Sakoda, 1986). Notably, face recognition relies more on relatively low spatial frequencies (8-16 cycles per face) (Parker & Costen, 1999; Schyns & Oliva, 1999). Conceivably, the difference between the performance onset times for experts and novices may reflect a difference in the ability to utilize information from lower spatial frequencies, which may dominate the output from earlier processing stages. The time-course of novice processing may be handicapped due to novices' relative inability to use this coarser information available in the early stages of processing. Although speculative, this notion raises intriguing possibilities for future research

Indeed, a recent study provided some insight into the role of high (fine-detail) and low (coarse) spatial frequency information in contributing to expert performance. This study explored the relative response in functionally defined face-selective areas of cortex in the fusiform gyrus to images of faces and cars containing only high and low spatial frequency information (Gauthier et al., 2005). The participants in this study had a range of visual expertise with cars, ranging from very little to extensive. Results failed to show a bias for high or low spatial frequency filtered images, but they suggested that both types of information contribute to expert performance. Consistent with previous studies showing an increase in FFA for cars among car experts, activation in the FFA for cars relative to faces for both high and low spatial frequency filtered images was correlated with a participants level of visual expertise with cars. Yet, these activation measures for the high and low spatial frequency filtered images were not correlated with each other. These results suggest that the processing of both low and high spatial frequency information make important, but independent, contributions to expert performance. One possibility is that the independence of the contribution of high and low spatial frequency to expert performance may arise because they influence different stages of encoding. Further studies should explore the differential contributions of high and low spatial frequency information and their relation to the temporal processing advantage for objects of expertise.

The earlier onset for subordinate-level categorizations for objects of expertise might also be a result of privileged access to processing resources, whether it be directly though the presence of more neurons tuned to objects of expertise or indirectly through automatic attentional allocation. A sub-cortical area of the brain known as the amygdala is believed to play an important role in allocating attention to highly arousing visual information (Anderson & Phelps, 2001), particularly when the motivational significance of such information has been learned (Holland, Han, & Gallagher, 2000). It is possible that objects of expertise preferentially receive attention due to signals from this sub-cortical area, as objects of expertise—such as cars for car experts—are likely to be more arousing than typical objects, such as chairs. Consistent with this possibility, the amygdala is believed to be crucial for the development of visual expertise with faces; abnormalities in the function of this area are believed to contribute to the failure of individuals with autism to develop expertise with faces (Schultz, 2005). Providing further support for the role of this area in the development of expertise is the presence of amygdala activation for “Digimon” cartoon characters, but not faces, in a boy with autism who has considerable skill at identifying these creatures (Grelotti, Gauthier, & Schultz, 2002). Interestingly, the amygdala has also been shown to respond more strongly to the low spatial frequency information in faces, once again suggesting that this coarse level information available first to perceptual processing mechanisms may play an important role in expertise. Future studies should explore the role of the amygdala in the development of visual expertise among normal observers.

Final Conclusions

The studies in this dissertation highlight the potential of perceptual expertise to impact numerous cognitive functions that rely on output from specialized perceptual mechanisms. In particular, face and non-face expert encoding appear to share common underlying mechanisms that, in turn, affect VSTM. The potential for experience, possibly through a change in encoding strategy, to influence VSTM reveals plasticity in this system. Results from prior literature suggest that VSTM capacity is determined by a number of factors, including stimulus complexity and inherent individual differences. The results of this dissertation add to the list of such factors by revealing important roles of task constraints and observers' level of experience with a given object category. Taken together, such demonstrations challenge accounts of VSTM that appeal only to hard-wired capacity limits (Luck & Vogel, 1997). This dissertation also provides intriguing insights into the differences between expert and novice encoding, such as the initial "head-start" that characterizes expert recognition performance. This raises interesting and empirically testable hypotheses about the mechanisms underlying the typically superior performance of experts more generally. The results of this dissertation can serve as a catalyst for studies probing the nature of holistic representations and the impact of such representations on higher-level cognitive functions.

APPENDIX A

Comparison of response times from Experiment 1

A 2 (orientation; upright, inverted) x 3 (duration; 500 ms, 1500 ms, 2500 ms) x 5 (set size; 1, 2, 3, 4, 5) found main a effect of orientation, $F(1,23) = 20.240$, $p=.0002$, with response times longer for trials with inverted, compared to upright, faces. There was also a main effect of set size, $F(4,92) = 119.21$, $p<.0002$, with response times longer for larger set sizes. The main effect of duration, $F(2,46) = 1.9824$, $p=.1493$, nor the interactions between any of the factors reached significance (all *p*s >.2522). Therefore, the larger VSTM for upright, relative to inverted faces, cannot be explained by a trade-off with response time as this account would have predicted that responses to trials with upright faces would be slower, rather than faster, than those to trials with inverted faces.

Table A. Comparison of Response Times for Upright and Inverted Faces in Experiment 1

Condition	Set Size 1	Set Size 2	Set Size 3	Set Size 4	Set Size 5	*Mean*
Upright						
500 ms	754(29)	888(30)	959(34)	1007(42)	1074(40)	936
1500 ms	753(31)	895(27)	988(33)	1018(36)	1073(37)	946
2500 ms	742(34)	916(36)	962(29)	1047(40)	1061(36)	946
Mean	750	900	970	1024	1069	943
Inverted						
500 ms	803(36)	953(29)	994(40)	1077(59)	1089(51)	983
1500 ms	808(43)	946(33)	1030(41)	1061(53)	1133(59)	995
2500 ms	826(35)	973(36)	1064(38)	1076(45)	1094(44)	1007
Mean	812	957	1029	1071	1105	995

APPENDIX B

Comparison of response times from Experiment 2

A 2 (category; face, watch, car) x 3 (duration; 500 ms, 1500 ms, 2500 ms) x 5 (set size; 1, 2, 3, 4, 5) found main a effect of set size, $F(4,80) = 158.45$, $p \leq .0001$, with response times longer for trials with a larger number of items. No other main effects or interactions between any of the factors reached significance (all *p*s >.2995). Therefore, the larger benefit of additional encoding time for VSTM capacity for faces, relative to that for watches or cars, cannot be explained by a trade-off with response time as this account would have predicted that responses to trials with upright faces would be slower, particularly in trials with a longer presentation duration for the study array, which is not supported by the results of this analysis of the response time data.

Table B. Comparison of Response Times for Faces, Watches, and Cars in Experiment 2

Condition	Set Size 1	Set Size 2	Set Size 3	Set Size 4	Set Size 5	*Mean*
Faces						
500 ms	731(40)	873(44)	964(41)	997(49)	1014(40)	916
1500 ms	762(42)	845(40)	945(43)	1025(52)	1046(47)	924
2500 ms	746(44)	861(41)	943(39)	994(53)	1011(44)	911
Mean	746	859	951	1005	1024	917
Watches						
500 ms	748(42)	865(43)	940(57)	987(47)	1017(50)	911
1500 ms	762(45)	865(49)	916(51)	973(47)	1044(57)	912
2500 ms	746(43)	875(48)	923(52)	1005(60)	1047(52)	919
Mean	752	868	926	988	1036	914
Cars						
500 ms	710(49)	841(55)	890(58)	932(52)	999(57)	875
1500 ms	696(46)	831(48)	887(50)	970(52)	986(55)	874
2500 ms	684(49)	838(52)	895(50)	924(52)	974(58)	863
Mean	697	837	891	942	987	871

APPENDIX C

Comparison of response times from Experiment 3

A 2 (category; face, watch, car) x 3 (set size; 3, 4, 5) found main a effect of set size, $F(2,70) = 16.555$, $p \leq .0001$, with response times longer for trials with a larger number of items. No other main effects or interactions between any of the factors reached significance (all *p*s >.5685). Therefore, the larger VSTM capacity for faces, relative to that for watches or cars, cannot be explained by a trade-off with response time as this account would have predicted that responses to trials with upright faces would be slower than that for watches and cars.

Table C. Comparison of Response Times for Faces, Watches, and Cars in Experiment 3

Condition	Set Size 3	Set Size 4	Set Size 5	*Mean*
Faces	1085(42)	1112(40)	1151(41)	1116
Watches	1082(43)	1115(47)	1152(52)	1116
Cars	1071(45)	1072(45)	1139(50)	1094
Mean	1079	1100	1147	1109

APPENDIX D

Description of the car expertise measure

Participants matched sequentially presented (256 x 256) grayscale images of cars and birds on the basis of their model or species, respectively. One-hundred and twelve images of cars and birds in viewpoints varying from front to profile were used. Each participant performed 224 trials, with 112 from each category. The first image was presented for 1000 ms, followed by a mask for 500 ms, and then the second image was presented and remained on the screen until participants made a key press to indicate of the two cars were the same or different model or the two birds with the same or different species, or 5000 ms had passed. The trials were blocked by category, with four blocks of each bird and car trials. Matching stimuli were not physically identical, but were different exemplars of the same bird species or the same make/model of car from different years.

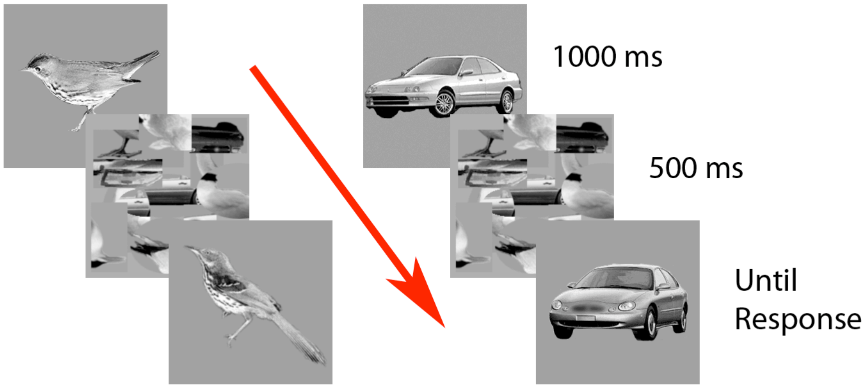


Figure D. Diagram illustrating the trial sequence in the car expert measure.

APPENDIX E

Comparison of response times from Experiment 4

A 2 (category; faces, cars) x 3 (duration; 500 ms, 2500 ms, 4000 ms) x 2 (Category; face, car) x 2 (orientation; upright, inverted) x 2 (expertise group; expert, novice) found a main effect of expertise, with car experts responding more slowly in general compared to car novices, F(1,34)=4.801, p=.0354. Notably, there was no interaction between expertise group and any other factor (all ps>.2046). There was also main effect of orientation, F(1,34)=31.979, p≤.0001, and duration, F(2,68)=9.9319, p=.0002, with slower response for trials with inverted items and also longer presentation durations. There was also an interaction between orientation and duration, F(2,68)=3.5366, p=.0346, with responses slowing more with longer durations for inverted than upright stimuli. There was also an interaction between category and duration, F(2,68)=4.6943, p=.0123, with responses slowing more with longer durations for cars than faces. No other main effects or interactions were significant (all ps>.1241). Importantly, the general slowing of responses among car experts cannot account for the larger VSTM capacity for upright cars among experts but not novices. A sensitivity (d') / response time trade-off account of the expert VSTM advantage would have predicted that experts would be slower only for upright cars, but not inverted cars or upright and inverted faces, relative to car novices.

Table E. Mean Response Times for the Conditions in Experiment 4

Condition	Expert	Novice
Face		
Upright		
500 ms	1042(39)	956(36)
2500 ms	1049(36)	924(32)
4000 ms	1044(38)	928(29)
Mean	1045	936
Inverted		
500 ms	1066(36)	962(33)
2500 ms	1072(32)	987(35)
4000 ms	1105(37)	998(34)
Mean	1081	982
Car		
Upright		
500 ms	1085(33)	931(35)
2500 ms	1094(30)	955(35)
4000 ms	1144(38)	973(37)
Mean	1108	953
Inverted		
500 ms	1116(33)	979(41)
2500 ms	1120(28)	994(37)
4000 ms	1193(36)	1014(38)
Mean	1143	996

APPENDIX F

Comparison of response times from Experiment 5

A 2 (category; faces, cars) x 3 (duration; 500 ms, 2500 ms, 4000 ms) x 2 (Category; face, car) x 2 (orientation; upright, inverted) x 2 (expertise group; expert, novice) found main effects of orientation, F(1,29)=79.130, p≤.0001, and duration, F(2,58)=4.6984, p=.0128, similar to that found in Experiment 5, with slower response for trials with inverted items and also longer presentation durations. There was an interaction between category and orientation, F(1,29)=4.6666, p=.0392, with faster responses for upright, but not inverted, faces relative to cars. There was also a trend for generally slower response times among car experts compared to car novices, F(1,29)=3.1232, p=.0877, and for an interaction between expertise group, orientation and duration, F(1,58)=2.5258, p=.0888, with car experts tending to show slower response times with longer presentations for inverted items but not for upright items. Finally, there was a 4-way interaction between orientation, duration, expertise group, and category, F(2,58)=4.5586, p=.0145. This interaction was due to a greater increase in response times for longer presentation durations for inverted cars among cars experts. No other interactions between any of the factors were significant (all ps>.2046). Importantly, the trend for a general slowing of responses for among car experts and also the slower response times among experts for trials with inverted cars presented for longer presentation durations cannot explain the larger VSTM capacity for upright cars among experts but not novices. A sensitivity (d') / response time trade-off account of the expert VSTM advantage would have predicted that experts would be slower for *upright* cars, but *not* inverted cars or upright and inverted faces, relative to car novices.

Table F. Mean Response Times for the Conditions in Experiment 5

Condition	Expert	Novice
Face		
Upright		
500 ms	1018(52)	880(31)
2500 ms	1012(47)	881(30)
4000 ms	1020(50)	910(36)
Mean	1016	890
Inverted		
500 ms	1078(51)	942(36)
2500 ms	1091(52)	966(32)
4000 ms	1132(52)	966(38)
Mean	1100	958
Car		
Upright		
500 ms	1063(45)	926(31)
2500 ms	1050(45)	961(31)
4000 ms	1098(55)	935(30)
Mean	1070	941
Inverted		
500 ms	1081(47)	984(34)
2500 ms	1120(46)	962(32)
4000 ms	1153(51)	999(35)
Mean	1118	982

APPENDIX G

Comparison of response times from Experiment 6

A 2 (category; faces, cars) x 2 (duration; 500 ms, 4000 ms) x 2 (Stimulus set; pre-switch, post-switch) x 2 (expertise group; expert, novice) found no main effects or interaction with expertise group (all ps>.2083). There was however a main effect of stimulus set, F(1,34)=32.462, p ≤.0001, with generally faster response times for the post-switch stimulus set. This is consistent with a generalized practice effect with participants responding faster towards the end of the experiment. There was also an interaction between stimulus set and presentation duration, F(1,34)=7.7019, p =.0089, with a larger drop in response time between the pre- and post-switch stimulus sets for shorter presentation durations. There was a trend for response times for trials with cars to be lower than those with faces, F(1,34)=2.5718, p =.1180. No other effects were significant (all p>.2001).

Table G. Mean Response Times for the Conditions in Experiment 6

Condition	Stimulus Set	Expert	Novice
Face			
500 ms	Pre-switch	1242(59)	1204(67)
	Post-switch	1080(63)	1037(43)
4000 ms	Pre-switch	1227(58)	1155(56)
	Post-switch	1148(69)	1058(49)
Car			
500 ms	Pre-switch	1222(61)	1261(91)
	Post-switch	1151(52)	1057(59)
4000 ms	Pre-switch	1253(45)	1255(77)
	Post-switch	1183(50)	1138(60)

APPENDIX H

Comparison of response times from Experiment 7

A 2 (category; faces, cars) x 3 (duration; 1000 ms, 2000 ms, 4000 ms) x 3 (Presentation format; simultaneous, sequential 1, sequential 2) ANOVA performed on the response time data found no main effects or interactions between any of the variables (all ps>.2761).

Table H. Mean Response Times for the Conditions in Experiment 7

Condition	Presentation Format	Response Times (SE)
Face		
1000 ms	Simultaneous	1207(44)
	Sequential 1	1175(51)
	Sequential 2	
2000 ms	Simultaneous	1170(38)
	Sequential 1	1168(44)
	Sequential 2	1171(44)
4000 ms	Simultaneous	1178(38)
	Sequential 1	1179(42)
	Sequential 2	1166(35)
Car		
1000 ms	Simultaneous	1201(54)
	Sequential 1	1207(59)
	Sequential 2	
2000 ms	Simultaneous	1202(51)
	Sequential 1	1208(57)
	Sequential 2	1222(59)
4000 ms	Simultaneous	1215(56)
	Sequential 1	1220(57)
	Sequential 2	1235(48)

REFERENCES

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological Studies of Human Face Perception. I: Potentials Generated in Occipitotemporal Cortex by Face and Non-face Stimuli. *Cerebral Cortex, 9*, 415-430.

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.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411*(6835), 305-309.

Ashbridge, E., Perrett, D. I., Oram, M. W., & Jellema, T. (2000). Effect of image orientation and size on object recognition: Responses of single units in the macaque monkey temporal cortex. *Cognitive Neuropsychology, 17*(13), 13-34.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225*(2), 82-90.

Baddeley, A. D. (1986). *Working Memory*. New York: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *In G. H. Bower (Ed.), The psychology of learning and motivation, 8*, (pages 47-89) New York: Academic Press.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104*(1), 54-75.

Barnard, P. J. (1999). Modelling working memory pphenomena within a multiprocessor architecture. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active control and executive maintenance* (pp. 298-339). New York, NY, US: Cambridge University Press.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*(3), 281-316.

Bentin, S., Allison, T., Puce, A., Perez, E., & et al. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*(6), 551-565.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Bindemann, M., Burton, A., M., & Jenkins, R. (in press). Capacity limits for face processing.

Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex, 8*(6), 510-523.

Boutet, I., Gentes-Hawn, A., & Chaudhuri, A. (2002). The influence of attention on holistic face encoding. *Cognition, 84*, 321-341.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433-436.

Brandeis, D., & Lehmann, D. (1986). Event-related potentials of the brain and cognitive processes: Approaches and applications. *Neuropsychologia, 24*(1), 151-168.

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition, 38*(2), 109-144.

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

Caldara, R., Thut, G., Servoir, P., Michel, C. M., Bovet, P., & Renault, B. (2003). Face versus non-face object perception and the 'other-race' effect: a spatio-temporal event-related potential study. *Clinical Neurophysiology, 114*(3), 515-528.

Carey, S. (1981). The development of face perception. In G. Davies & H. Ellis & J. Shephard (Eds.), *Perceiving and Remembering Faces* (pp. 9-38). New York: Academic Press.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology, 16*(4), 257-269.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory, 2*(6), 641-653.

Chase, W. G., & Ericsson, K. A. (1981). Skilled Memory, *Cognitive skills and their acquisition* (pp. 141-189). NJ, Erlbaum: Hillsdale.

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

Chen, D., Eng, H. Y., & Jiang, Y. (in press). Visual working memory for trained and novel polygons. *Visual Cognition.*

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 109-127.

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.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature, 386*, 608-611.

Cowan, N. (1999). An Embedded-Processes Model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active control and executive maintenance* (pp. 62-101). New York, NY, US: Cambridge University Press.

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.

Curby, K. M., & Gauthier, I. (2001). Interference between car and face expertise. *Paper presented at the OPAM workshop, Orlando, Florida, November*.

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

Diamond, R., & Carey, S. (1977). Developmental Changes in the Representation of Faces. *Journal of Experimental Child Psychology, 23*(1), 1-22.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107-117.

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.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433-458.

Eng, H. Y., Chen, D., & Jiang, Y. (in press). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin and Review*.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active control and executive maintenance* (pp. 102-134). New York, NY, US: Cambridge University Press.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. *American Scientist, 70*(6), 607-615.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*(4448), 1181-1182.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active control and executive maintenance* (pp. 257-297). New York, NY, US: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.

Ericsson, K. A., & Oliver, W. L. (Eds.). (1989). *A methodology for assessing the detailed structure of memory skills*: Oxford, England: John Wiley and Sons.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13*(2), 171-180.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 628-634.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "Special" about Face Perception? *Psychological Review, 105*(3), 482-498.

Finke, R. A. (1989). *Principles of Mental Imagery*. Cambridge, MA: The MIT Press.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception, 29*, 159-170.

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. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673-1682.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 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.

Grelotti, D. J., Gauthier, I., & Schultz, R. T. (2002). Social interest and the development of cortical face specialization: What autism teaches us about face processing. *Developmental Psychobiology, 40*(3), 213-225.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: as soon as you know it is there, you know what it is. *Psychological Science, 16*(2), 152-160.

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception, 13*(5), 505-512.

Holland, P. C., Han, J. S., & Gallagher, M. (2000). Lesions of the amygdala central nucleus alter performance on a selective attention task. *Journal of Neuroscience, 20*(17), 6701-6706.

Hosie, J. A., Ellis, H. D., & Haig, N. D. (1988). The effect of feature displacement on the perception of well-known faces. *Perception, 17*(4), 461-474.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience, 8*(3), 197-230.

James, W. (1890). *The principles of psychology*. Oxford, England: Holt.

Johnson, J. S., & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision, 5*(4), 299-312.

Jolicoeur, P. (1990). Identification of disoriented objects: A dual-systems theory. *Mind & Language, 5*(4), 387-410.

Jolicoeur, P., & Dell' Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36*(2), 138-202.

Jolicoeur, P., Gluck, M., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16*, 243-275.

Kanwisher, N. (2000). Domain specificity in face perception. *Nat Neurosci, 3*(8), 759-763.

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

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

Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active control and executive maintenance* (pp. 183-223). New York, NY, US: Cambridge University Press.

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology, 80*(1), 324-330.

Kundel, H. L., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*(3), 527-532.

Lawson, R., & Jolicoeur, P. (1998). The effects of plane rotation on the recognition of brief masked pictures of familiar objects. *Memory and Cognition, 26*(4), 791-803.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology, A: Human Experimental Psychology, 51*(3), 449-473.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *Quarterly Journal of Experimental Psychology, A: Human Experimental Psychology, 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.

Leder, H., & Carbon, C. C. (2005). When context hinders! Learn-test compatibility in face recognition. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58A*(2), 235-250.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience,* 5(9), 910-916.

Logie, R. H. (1995). *Visuo-spatial working memory*: Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5*(5), 552-563.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working* *memory: mechanisms of active control and executive maintenance* (pp. 135-182). New York, NY, US: Cambridge University Press.

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

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. B, 200*, 269-294.

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*(2), 336-349.

McKone, E., & Kanwisher, N. (2005). Does the human brain process objects of expertise like faces? A review of the evidence. In S. Dehaene & J.-R. Duhamel & M. D. Hauser & G. Rizzolatti (Eds.), *From Monkey Brain to Human Brain*: MIT Press.

McKone, E., & Robbins, R. (2005). *No face-like processing for objects-of-expertise in three behavioral tasks.* Paper presented at the Meeting of the Cognitive Neuroscience Society, New York, N.Y.

Meyler, A., & Breznitz, Z. (1998). Developmental associations between verbal and visual short-term memory and the acquisition of decoding skill. *Reading and Writing, 10*(6), 519-540.

Michel, C. M., Henggeler, B., & Lehmann, D. (1992). 42-Channel potential map series to visual contrast and stereo stimuli: Perceptual and cognitive event-related segments. *International Journal of Psychophysiology, 12*(2), 133-145.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2003). Early visual experience is necessary for the development of some--but not all--aspects of face processing. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood* (pp. 99-117). Hauppauge, NY, US: Nova Science Publishers, Inc.

Mruczek, R. E. B., & Sheinberg, D. L. (in press). Distractor familiarity leads to more efficient visual search for complex stimuli. *Perception and Psychophysics*.

Murray, J., Yong, E., & Rhodes, G. (2000). Revisiting the Perception of Upside-Down Faces. *Psychological Science, 11*(6), 492-496.

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The Influence of Expertise on X-Ray Image Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 553-557.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75*, 522-536.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active control and executive maintenance* (pp. 375-411). New York, NY, US: Cambridge University Press.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception and Psychophysics, 64*(7), 1055-1067.

Olson, I. R., & Jiang, Y. (in press). Visual short-term memory is not improved by training. *Memory and Cognition*.

Olson, I. R., Jiang, Y., & Moore, K. S. (2005). Associative learning improves visual working memory performance. *Journal of Experimental Psychology: Human Perception & Performance, 31*(5), 889-900.

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

Otani, H., Landau, J. D., Libkuman, T. M., St Louis, J. P., Kazen, J. K., & Throne, G. W. (1997). Prospective memory and divided attention. *Memory, 5*(3), 343-360.

Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition, 82*, 225-257.

Palmeri, T. J., Wong, A. C., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Science, 8*(8), 378-386.

Parker, D. M., & Costen, N. P. (1999). One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face processing. *Current Psychology, 18*(1), 118-127.

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics, 44*(4), 369-378.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*, 283-290.

Postle, B. R., Druzgal, T. J., & D'Esposito, M. (2003). Seeking the neural substrates of visual working memory storage. *Cortex, 39*(4-5), 927-946.

Ranganath, C., Cohen, M. X., Dam, C., & D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *Journal of Neuroscience, 24*(16), 3917-3925.

Ranganath, C., DeGutis, J., & D'Esposito, M. (2004). Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. *Brain Research: Cognitive Brain Research, 20*(1), 37-45.

Reed, A. V. (1976). List length and the time course of recognition in immediate memory. *Memory and Cognition, 4*(1), 16-30.

Reed, S. K. (1974). Structural descriptions and the limitations of visual images. *Memory & Cognition, 2*, 329-336.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition, 47*(1), 25-57.

Robertson, L. C. (1996). Attentional persistence for features of hierarchical patterns. *Journal of Experimental Psychology: General, 125*(3), 227-249.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

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*(4), 539-549.

Ross Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development, 74*(6), 1807-1822.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Reviews, 1*(1), 63-75.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left lateralized face-like electrophysiological responses. *Psychological Science, 13*(3), 250-257.

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), 69-74.

Rossion, B., Kung, C. C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Science,s U S A, 101*(40), 14521-14526.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Science, 8*(8), 363-370.

Schneider, W. (1999). Working memory in a multilevel hybrid connectionist control architecture (CAP2). In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of* *active control and executive maintenance* (pp. 340-374). New York, NY, US: Cambridge University Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.

Schultz, R. T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience, 23*(2-3), 125-141.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition, 69*(3), 243-265.

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*(4), 904-915.

Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *Journal of Neuroscience, 21*(4), 1340-1350.

Shulman, G. L., Sullivan, M. A., Gish, K., & Sakoda, W. J. (1986). The role of spatial-frequency channels in the perception of local and global structure. *Perception, 15*(3), 259-273.

Smith, E. E., & Jonides, J. (1997). Working memory: a view from neuroimaging. *Cognitive Psychology, 33*(1), 5-42.

Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences, U S A, 95*(20), 12061-12068.

Sperling, G. (1960). The information available in brief presentations. *Psychological Monographs, 74*.

Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance, 21*(4), 901-913.

Tanaka, J., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *NeuroReport, 10*(4), 829-835.

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *Journal of Experimental Psychology: General, 130*(3), 534-543.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science, 12*(1), 43-47.

Tanaka, J. W., & Farah, M. J. (1991). Second-order relational properties and the inversion effect: Testing a theory of face perception. *Perception & Psychophysics, 50*(4), 367-372.

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

Tanaka, J. W., & Gauthier, I. (1997). Expertise in object and face recognition. In R. L. Goldstone & P. G. Schyns (Eds.), *Mechanisms of Perceptual Learning* (Vol. 36, pp. 83-125). San Diego, CA: Academic Press.

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.

Theeuwes, J., & Van der Stigchel, S. (in press). Faces capture attention: Evidence from inhibition-of-return. *Visual Cognition*.

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception, 9*(4), 483-484.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale* (4th ed.). Chicago: Riverside.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*(6984), 751-754.

Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive Affective Behavioral Neuroscience, 5*(2), 144-155.

Tong, F., & Nakayama, K. (1999). Robust representations for faces: evidence from visual search. *Journal of Experimental Psychology: Human Perception & Performance, 25*(4), 1016-1035.

Treisman, A. M. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 194-214.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin and Review, 2*(1), 20-54.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*(6984), 748-751.

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*(1), 92-114.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (in press). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*.

Vuilleumier, P. (2000). Faces call for attention: evidence from patients with visual extinction. *Neuropsychologia, 38*, 693-700.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cognitive Affective Behavioral Neuroscience, 3*(4), 255-274.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception and Psychophysics, 56*(5), 495-500.

Ward, R., Duncan, J., & Shapiro, K. (1996). The Slow Time-Course of Visual Attention. *Cognitive Psychology, 30*(1), 79-109.

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology, 46*(4), 361-413.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*(1), 48-64.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*(1), 67-85.

Wickelgren, W. A., & Corbett, A. T. (1977). Associative interference and retrieval dynamics in yes-no recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 3*(2), 189-202.

Xu, Y. (2002a). Encoding color and shape from different parts of an object in visual short-term memory. *Perception and Psychophysics, 64*(8), 1260-1280.

Xu, Y. (2002b). Feature integration across parts in visual search. *Perception, 31*(11), 1335-1347.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex, 15*(1234-1242).

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141-145.

Young, A. W., Hellawell, D., & Hay, D. (1987). Configural information in face perception. *Perception, 10*, 747-759.

Young, R. M., & Lewis, R. L. (1999). The Soar cognitive architecture and human working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active control and executive maintenance* (pp. 224-256). New York, NY, US: Cambridge University Press.