THE EFFECT OF DIFFERENT TRAINING EXPERIENCES ON

OBJECT RECOGNITION IN THE VISUAL SYSTEM

By

Alan Chun-Nang Wong

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

December, 2007

Nashville, Tennessee

Approved:

Professor Isabel Gauthier

Professor Thomas H. Carr

Professor Donald L. Compton

Professor Gordon D. Logan

Professor Thomas J. Palmeri

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Chapter

LIST OF TABLES

LIST OF FIGURES

PREFACE

OVERVIEW OF THE DISSERTATION

This dissertation is aimed at comparing the effects of different experiences on human visual object recognition. Functional specialization occurs in the visual object recognition system in terms of selectivity for certain categories (e.g., faces, letters, scenes, and body parts) over others in different regions of the ventral occipitotemporal cortex, as well as specific behaviors associated with different categories. A number of factors have been proposed to explain the functional specialization, yet few empirical studies addressed the unique contribution of different factors. In this dissertation I investigated the role of recognition experience by comparing the effects of two types of categorization training. The use of novel objects allowed me to control for stimulus properties and higher-level factors (emotional, social, and linguistic processing). I found that different training experiences resulted in different changes in behavior and neural activity.

This dissertation begins (CHAPTER I) with a review of the functional specialization for object categories, with an emphasis on faces and letters. I describe the different proposed factors affecting category selectivity in the brain, and discuss the difficulties associated with testing the contribution of individual factors using real-world objects. I then discuss the use of novel object training as a useful tool in answering this research question, and the study of more than one type of training as a valuable way to extend the tool.

Following the review of this literature, CHAPTER II reports a novel-object training paradigm inspired by Gauthier *et al.* (1997, 1998, 1999, 2002) to address one facet of the problem: can different recognition experiences with one set of novel objects lead to qualitatively different behavioral and neural consequences? Two groups of participants went through training procedures designed to mimic experience with either faces or letters. The face-like group learned to categorize novel objects at a subordinate level, similar to how people discriminate among faces in person identification. The letter-like group learned to recognize, at the basic level, multiple novel objects presented in spatially organized clusters with coherent font-like styles, similar to how people process letters when reading a text. The two training procedures led to different behavioral phenomena associated with face perception (e.g., a smaller basic-level advantage and holistic and configural processing effects) and letter perception (e.g., a larger basic-level advantage and faster recognition of objects in a string).

CHAPTER III reports the effects of the two types of training on fMRI activity. The face-like group showed enhanced activations around the right fusiform face area. The degree of change also correlated with the magnitude of the configural processing effect found in the behavioral post-training test. The letter-like group did not show enhanced activations in the letter-selective region as predicted. Instead, activations for this group changed in a more distributed manner. A greater emphasis was put on the medial parts of the ventral occipitotemporal cortex after training. These medial parts also showed a preference for object presentation in the periphery of the visual field, and for basic-level categorization relative to subordinate-level categorization.

I conclude my dissertation with CHAPTER IV in which I discuss the implications

concerning the role of experience in forming functional specialization in the visual object

recognition system. I also discuss how the comparison of multiple types of training can

help answer broader questions concerning perceptual expertise.

CHAPTER I

ACCOUNTING FOR CATEGORY SELECTIVITY IN THE VISUAL OBJECT
RECOGNITION SYSTEM

One of the most well-known and difficult questions in vision science is how the

human visual system recognizes a wide range of objects so efficiently. One recipe for

such achievement may be functional specialization, in that different neural substrates are

best at processing different aspects of the visual input. Perhaps the most well-known

example is the separation between the "what" and "where" pathways, i.e., the ventral

visual stream focusing on form and object recognition, and the dorsal visual stream

concentrating on spatial vision (Goodale *et al.*, 1991, p.457; Haxby *et al.*, 1991). Apart

from this distinction, research has further revealed that within the ventral visual pathway,

different regions are responsible for different categories of objects (see review in Grill-

Spector & Malach, 2004). The functional organization of the object recognition system

underlying such category selectivity remains a matter of intense debate.

This chapter will first review the findings of category selectivity in the visual

object recognition system, with an emphasis on face and letter perception, two heavily

studied cases of expert object perception. The chapter will then evaluate theories of

category selectivity and discuss the difficulties in testing the theories. This chapter will

conclude by stating that one promising way to approach the question is to compare the

effects of different novel object training procedures, a basis of this dissertation.

Category selectivity in visual object recognition system

The suggestion of functional specialization within the visual recognition system first came from neuropsychological findings of double dissociations between, for instance, deficits in face vs. non-face perception, scene vs. non-scene perception, and word vs. non-word perception (see review in Farah, 2004). With the advance of neurophysiological and neuroimaging research, it has been revealed that within the object recognition system located in the ventral visual pathway, different substrates respond with some selectivity to different objects over others, including faces (Allison *et al.*, 1994a; Bentin *et al.*, 1996; Kanwisher *et al.*, 1997; Puce *et al.*, 1996; Xu, 2005), scenes and buildings (Epstein *et al.*, 1999; Epstein & Kanwisher, 1998), body parts (Downing *et al.*, 2001), tools (Chao *et al.*, 1999; Noppeney *et al.*, 2006), and letters and words (Cohen *et al.*, 2000; Gauthier *et al.*, 2000b; Hasson *et al.*, 2002; James *et al.*, 2005; Polk *et al.*, 2002; Puce *et al.*, 1996).

Face and letter selectivity probably represent the best examples of object selectivity as they are frequently studied and contrasted. In recent years, a growing number of studies have highlighted behavioral and neural differences between the brain regions recruited by the two categories, as reviewed below.

Selectivity for face perception

The earliest indication of face selectivity in the object recognition system comes from studies of patients with impaired face perception but less impaired or normal ability to recognize non-face objects (see reviews in Farah, 2004; McNeil & Warrington, 1991; Young, 1992). EEG studies have shown that visual presentation of a face elicits more

activity than other objects (e.g., flowers, cars, hands, furniture) at posterior channels at about 170 ms after stimulus onset (Bentin *et al.*, 1996; Rossion *et al.*, 2000). Intracranial recordings and brain imaging studies have confirmed that a right (sometimes bilateral) ventral occipitotemporal area, including parts of the fusiform gyrus, is more strongly activated by faces than over a dozen other object categories (Allison *et al.*, 1994a; Downing *et al.*, 2006; Kanwisher *et al.*, 1997; Puce *et al.*, 1996; Tarkiainen *et al.*, 2002). This area has been named the "fusiform face area" (FFA) for its preference for face processing.

A few behavioral signatures have also been suggested to differentiate between processing for faces and non-face objects. Face recognition is associated with a much smaller basic-level advantage than object perception. For instance, people can identify a face (e.g., as "Elvis Presley") as fast as they can say it is a face (Tanaka, 2001). In contrast, objects are generally categorized faster at the basic level (e.g., bird) than at a subordinate level (e.g., robin) (Rosch *et al.*, 1976; Tanaka, 2001). Another phenomenon is the inversion effect, where recognition performance suffers more from inverting a face than a non-face object (Yin, 1969). Besides, faces seem to be processed in a more holistic manner. One demonstration is the part-whole effect (Tanaka & Farah, 1993), where judging which of two face parts appeared in a previously presented face is easier when the face parts are in the context of a whole face than when they are presented in isolation. Another indicator is the composite effect (Young *et al.*, 1987), where identifying one half of a composite face (formed with top and bottom halves of different faces) is harder when the other, inconsistent half is aligned with the relevant half than when the two halves are misaligned. Subsequent studies that filtered out the effect of response bias confirmed the

effect (Gauthier *et al.*, submitted). It is also found that the probability of identifying a whole face is greater than the sum of the probabilities of identifying its two halves (Yovel *et al.*, 2005), again suggestive of the holistic nature in face processing.

Selectivity for letter perception

The indication of selectivity for letter perception began with a case study of a patient suffering from a left inferior occipitotemporal lesion and the inability to recognize letters and words, while having no trouble speaking, writing or recognizing other visual material (Bub *et al.*, 1993). This deficit, called pure alexia, was interpreted as a specific deficit in processing orthographic codes, or word forms. Since then word forms have been regarded as perceptual units with distinct representations, and a visual word form system has been proposed for the parsing of letter strings into familiar units for further analyses (Carr & Pollatsek, 1985; Warrington & Shallice, 1980). Recent neuroimaging studies have identified a visual word form area (VWFA) in the left inferior occipitotemporal region, including parts of the left fusiform gyrus, which may be responsible for such processes (Cohen *et al.*, 2000). This area has been suggested to be responsible for reading-specific processing of an ordered string of letters or graphemes, invariant to changes in location, case, and font (Cohen & Dehaene, 2004; Cohen *et al.*, 2002; Dehaene *et al.*, 2004; Dehaene *et al.*, 2002; McCandliss *et al.*, 2003), although alternative interpretations exist (Price & Devlin, 2003; Vigneau *et al.*, 2005; Vinckier *et al.*, 2007).

Meanwhile, there is evidence for the role of a left inferior occipitotemporal region in processing letters outside of the word context. For example, intracranial recordings in bilateral posterior fusiform gyrus have found a larger N200 component for letterstrings

(in the form of words, pronounceable pseudowords, or consonant strings) than for cars and butterflies (Allison *et al.*, 1994b; Nobre *et al.*, 1994). Neuroimaging studies have revealed higher activations for consonant strings compared with textures and faces (Puce *et al.*, 1996), with faces and buildings (Hasson *et al.*, 2002), and with digit strings (Polk *et al.*, 2002). Some findings further suggest that the left occipitotemporal region also plays a crucial role in single letter processing. An EEG study showed that at posterior channels an early negative component occurring at about 170 ms after stimulus onset was enhanced with single, familiar characters compared with unknown characters or pseudoletters (Wong *et al.*, 2005). An MEG study has also shown an early component at about 150-200 ms that is larger for single letters than geometric shapes (Tarkiainen *et al.*, 1999). A PET study revealed a correlation between the activity level in the left occipitotemporal region and the performance of discrimination between letters and symbols (Garrett *et al.*, 2000). In addition, fMRI studies have shown higher activations in the left fusiform gyrus for letters than for oblique lines (Longcamp *et al.*, 2003), Chinese characters (James *et al.*, 2005), and objects and faces (James & Gauthier, 2006). Further, higher activations have been found for letters compared with faces in a bilateral occipitotemporal region (Gauthier *et al.*, 2000b) and in the left occipitotemporal region when attention was paid to letters compared with colors and symbols (Flowers *et al.*, 2004). Finally, higher activations have been shown for letters and words in one's familiar than unfamiliar writing system, indicating the role of experience in forming letter selectivity (Baker *et al.*, 2007; Wong *et al.*, submitted).

Importantly, single letter processing has been shown to recruit separate neural substrates from word and letterstring processing. A recent study found that a fusiform

region anterior to the VWFA showed selectivity for single letters but not letter strings, while a region overlapping with the VWFA showed selectivity for strings but not single letters (James *et al.*, 2005). Overall, the finding of letter-selective areas with different visual controls provides strong evidence for the pivotal role of the left inferior occipitotemporal region in visual processing of letters.

Finally, work conducted in our own and other labs has also found a number of behavioral phenomena specific for letter processing. One phenomenon is that expert letter perception is associated with an increase in the basic-level advantage. While all observers show a basic-level advantage with novel characters (i.e., faster at recognizing a character as a "B" than as a "B in Arial") just like other common objects, this basic-level advantage becomes even larger for those who are experts with the characters in the writing system (Wong & Gauthier, in press). Letter perception has also been shown to be sensitive to regularities in font (Gauthier *et al.*, 2006; Sanocki, 1987, 1988, 1991) and orientation (Jolicoeur, 1990), such that recognition of letters in a string is faster when the letters are in the same than different fonts or orientations.

Contrasting face and letter selectivity

In summary, face and letter selectivity are associated with very different neural activity patterns and behavioral signatures. Face perception tends to activate the middle fusiform region with a right preponderance, whereas letter perception tends to activate a more lateral region in the inferior temporal gyrus with a left preponderance. Face perception has a smaller basic-level advantage than common objects, whereas expertise with letter perception is related to an increase in such an advantage. Face perception is

suggested to proceed in a holistic manner. Letter perception involves rapid recognition of multiple instances and is affected by contextual factors like regularity in orientation and fonts.

Why would there be different recognition behaviors and neural selectivity patterns for different categories such as faces and letters? One would think that this question would be easy to answer given our understanding of the organization of the visual object recognition system. Theoretical accounts of the organization of the ventral occipitotemporal cortex have suggested factors that affect object selectivity. However, as discussed later, there is no satisfactory answer to the question yet, due to the insufficient empirical work that tests the unique contributions of different factors. The explanation of different face- and letter-selective areas, therefore, would be an important step in our pursuit of understanding object selectivity and the organization of object recognition system in general.

Theories about object selectivity

In recent years different theoretical views have emerged to explain existing data on object selectivity. They can be summarized as module-based, feature-based, and process-based views. These theories focus on different aspects of the issue and offer some ideas on how to characterize the different areas selective for faces and letters. Two questions will be emphasized. First, why is there selectivity for faces and letters? Second, why are the areas selective for faces and letters different?

According to this view, there are in the ventral stream several domain-specific modules for important categories, while the rest of the region is a general area responsible for the recognition of all other objects (Downing *et al.*, 2006; Kanwisher, 2000; Spiridon & Kanwisher, 2002; Yovel & Kanwisher, 2004). Support for this view comes from the finding that certain ventral areas show a high degree of category selectivity. For example, areas selective for faces, scenes, and body parts respond significantly higher to their preferred category than to each of 18 other categories (Downing *et al.*, 2006). It should be noted, however, that even in these areas, responses are not all or none. Instead, there is a gradation from the highest to the lowest responses across categories, suggesting that these areas are highly selective but not completely domain-specific.

Concerning the finding of graded and overlapping activity for different categories, it has been argued that standard imaging studies have a low spatial resolution, such that the measured activity is a result of pooling across distinct neural populations that respond exclusively to a category (Peelen & Downing, 2005). On this account, fMRI techniques with a higher resolution are expected to be able to reveal the activity of 'true' modules. This idea gains initial support from the finding that the areas selective for faces and body parts, though close to each other, can be dissociated using high-resolution fMRI techniques (Schwarzlose *et al.*, 2005). A recent high-resolution fMRI study showed that within the FFA defined by standard-resolution fMRI, some voxels showed a higher selectivity for faces than other voxels (Kalanit Grill-Spector *et al.*, 2007). It should be noted that, even with such a high resolution ($1 \times 1 \times 1$-mm voxel size), the most face-

selective voxels still show some activity to other categories relative to scrambled images. It remains to be seen if the high selectivity will still be observed with a larger number of control categories included. Also, it may be difficult to explain why in that study some voxels are selective for categories like novel sculptures for which a module would not be expected.

It is speculated that objects that recruit domain-specific modules do so because they meet three special conditions: (1) they are critical for survival over the course of evolution; (2) they are frequently encountered during one's life span; and (3) they are highly constrained in their geometry, i.e., "the visual appearance of exemplars of the stimulus class is highly constrained" (Downing *et al.*, 2006, p.7).

How well does the module-based view explain the existence of face-selective and letter-selective areas? Face perception meets the above-mentioned conditions well. Letter perception, however, probably does not meet the first condition. Writing appears too recent a human accomplishment (since about 5400 years ago) for evolutionary forces to cause development of a letter-selective area from scratch. It could be that evolution modified existing object recognition devices to create a letter area in 5400 years (or 270 generations). But to argue for such a case one has to show that the ability to read was critical for natural selection, and significantly increased the chance of survival over the majority of human population throughout the 5400 years. It is doubtful whether evolutionary significance is involved in the emergence of a letter-selective area.

Regarding why different areas are selective for faces and letters, the module-based view apparently does not offer a specific answer. Nevertheless, one can infer from the conditions listed above that the reason may be related to the differences in our

experience with faces and letters and to their physical properties. Exactly what differences in experience and physical properties are involved requires further investigation.

Feature-based view

This account suggests that the ventral stream is organized in terms of continuous object-form topography, such that neighboring neural substrates represent similar features or dimensions (Carlson *et al.*, 2003; Cox & Savoy, 2003; Haxby *et al.*, 2001; Ishai *et al.*, 1999; O'Toole *et al.*, 2005; K. Tanaka, 1996, 2003). Contrary to the module-based view, the feature-based view stresses that there is no absolute neural specialization for certain object categories. Instead, object representations are widely distributed and overlapping. An object may cause large responses in certain areas and low responses in others, but both the large and small responses are important parts of the population coding for the object. Support comes from findings that response patterns across different regions of the ventral temporal cortex contain diagnostic information for discriminating among different object categories (e.g., faces, houses, cats, chairs, shoes etc.). Even regions with high selectivity for certain objects (e.g., faces in the FFA) contain diagnostic information for categorization among non-preferred objects (Haxby *et al.*, 2001). It has also been found that how much object categories are confused with each other in terms of physical properties is correlated with how much the categories are confused in terms of fMRI response patterns in a wide region of the ventral temporal cortex (O'Toole *et al.*, 2005). From these results it is claimed that all parts of the ventral temporal cortex are equally important in representing a particular object category.

If indeed different parts of the object recognition system are equally useful for representing different objects, irrespective of activity level, then it would not be very meaningful to examine the origin and importance of the category-selective areas. However, regions within the ventral temporal cortex that prefer certain categories (e.g., FFA and the parahippocampal place area, PPA) have greater discriminatory power for categorization involving the preferred objects compared with other objects (O'Toole *et al.*, 2005; Spiridon *et al.*, 2006). Also, the activity in focal areas like the FFA is correlated with behavioral measures of expertise (Gauthier & Tarr, 2002). These all indicate that representations of objects are not fully distributed.

Similarly to the module-based view, the feature-based view does not readily explain why strong selectivity occurs for categories like faces and letters but not for many other objects. One can speculate, however, that since faces and letters are much more frequently seen than many other objects, neurons that prefer relevant features may gradually be tuned to respond more to faces and letters. And concerning why areas selective for faces and letters are different, the most straightforward explanation would be that these two categories have very different physical properties.

Process-based view

This view argues for a process map in the ventral temporal cortex, in which different substrates support different computations and are thus best suited for different recognition demands (Bukach *et al.*, 2006; Gauthier, 2000; Tarr & Gauthier, 2000). The central notion in this theory is that it is "recognition demand" and "experience" that determine which neural substrates are selectively recruited for a certain object. Different

object categories are associated with different recognition demands, and through experience, perhaps through feedback from parietal and frontal cortices (Freedman *et al.*, 2001), the system learns to recruit the best neural substrates to fulfill the different demands associated with an object category. Recognition demand in this context refers to the task one needs to accomplish for successful recognition. Examples of such demands can include distinguishing among similar objects, finding similarities among dissimilar objects, encoding objects in terms of spatial layout, etc. The strongest support for a role of experience in accounting for category selectivity comes from studies of face selectivity. The process-based view suggests that the FFA is involved in fine-grained, subordinate-level discrimination, and faces tend to selectively activate this area because faces are among the few categories that require subordinate-level discrimination. Accordingly, other object categories can also activate the FFA to a high extent when subordinate-level discrimination is involved. Indeed, it has been found that the FFA shows higher activity with subordinate-level identification of non-face objects compared with basic-level categorization (Gauthier *et al.*, 1997). In addition, selectivity has also been shown in the FFA for objects in domains of real-world expertise that involve subordinate-level categorization, such as birds for bird watchers and cars for car experts (Gauthier *et al.*, 2000a; Xu, 2005; but see Grill-Spector, Knouf, & Kanwisher, 2004). Furthermore, after one learned to identify computer-generated, novel objects in fine detail, the FFA began to develop selectivity for the novel objects, with correlations between the activity level and behavioral measures of holistic processing (Gauthier & Tarr, 2002; Gauthier *et al.*, 1999; Rossion *et al.*, 2004; but see Kanwisher, 2000). In addition, interference has also been shown between car and face processing for car

experts but not car novices, providing further evidence for the use of overlapping substrates for faces and objects that share a similar recognition demand of subordinate-level discrimination (Gauthier *et al.*, 2003).

Recently a separate line of research has examined one dimension along which recognition demand varies. It was revealed that in the ventral occipitotemporal cortex areas selective for different object categories correspond to the distribution of areas preferring foveal vs. peripheral presentation (Hasson *et al.*, 2003; Hasson *et al.*, 2002; Levy *et al.*, 2001; Levy *et al.*, 2004). For instance, face- and letter-selective areas are situated at more lateral regions (closer to the left and right boundaries of the brain). These same regions also have a foveal bias, i.e., a preference for object presentation at the fovea. On the contrary, scene-selective areas are in the more medial regions (closer to the midline of the brain). These regions respond more to object presentation in the periphery of the visual field. Based on this finding, a continuum was suggested such that at one extreme substrates mainly represent foveal regions with higher cortical magnification, enabling high-resolution object analyses; at the other extreme substrates represent peripheral regions with a lower cortical magnification and thus suitable for coarse analyses and large-scale integration (Malach *et al.*, 2002). Accordingly, recognition of object categories with different resolution demands should be associated with different cortical areas through experience (Hasson *et al.*, 2003; Malach *et al.*, 2002).

According to process-based view experience plays an important role in the development of object selectivity. This is similar to one of the conditions for the development of a domain-specific module suggested by the module-based view: faces and letters develop selectivity because we have extensive experience with both categories

during our lifetime. The process-based view goes further to distinguish between different types of experience. Faces and letters recruit different areas *because* our experiences with these categories are different. Our recent work has begun to compare face and letter perception. For example, while face perception typically demands categorization at a subordinate level (discriminating among faces with similar parts in the same general configuration), letter perception requires basic-level categorization (distinguishing between letters with different parts from each other) (Wong & Gauthier, in press). It has been suggested that because of the demand for subordinate-level categorization, face perception and similar types of object expertise may thus rely on mechanisms that are "holistic", in the sense that experts tend to process all the parts of a face, even when instructed to selectively attend to only one part (Gauthier *et al.*, submitted; Young *et al.*, 1987). In addition, face experts appear to rely on configural or relational information (e.g., the distance between parts) more than they do for other objects (Tanaka & Sengco, 1997). On the other hand, experience with letters is unique in that letters mostly occur in spatially organized clusters (as in a text where letters are well aligned and have a regular size and style). Such regularities in font can facilitate the recognition of letters at the basic level (Sanocki, 1987, 1988, 1991), in a similar way as voice regularity can facilitate the recognition of speech (Dupoux & Green, 1997; Mullennix *et al.*, 1989). In our recent work, this "font-tuning" effect was shown to be associated with expert but not novice perception of Roman letters and Chinese characters (Gauthier *et al.*, 2006). Hence, the different recognition demands and experiences associated with face and letter perception can be important factors for the different areas found to be selective for faces and letters.

Difficulty in studying individual factors

The theories and the body of empirical studies have improved our understanding of object selectivity in human visual system. Whether the category selectivity found provides sufficient support for domain specificity in the ventral stream is still under debate, but it is believed that object representations in the ventral occipitotemporal cortex are at least not completely distributed. Category-selective areas not only have higher activity but also contain more diagnostic information for their preferred categories. Different factors have been suggested to explain the recruitment of different areas for different object categories. Physical properties of the objects, recognition demand, and experience are some of the important factors.

So, going back to the question raised in the beginning, do we now understand why the areas selective for faces and letters are different? There are several possible answers. It can be because faces and letters have very different physical properties. Another possibility is that these differences are caused by the different recognition demands that characterize face and letter recognition. Another important factor not addressed by the theories above is that faces and letters are associated with different higher-level functions. Face perception has close links to emotional processing and social cognition, while letter perception is tightly tied with different language-related processes like reading, speaking, and writing. It is plausible that the need to connect to the higher-level areas may have constrained the locations of the face- and letter-selective areas. It is unknown whether all these factors are necessary to account for the different face- and letter-selective areas, or which factor is the most important. Such a problem also applies to comparisons involving other areas such as those selective for body parts and scenes.

The difficulty in evaluating the unique contribution of a particular type of factors on object selectivity comes from the fact that real-world object categories differ in multiple ways. It is interesting and informative to know that physical properties of objects correlate with neural activity patterns, and locations of category-selective areas seem to overlap with the eccentricity map. But these studies are all correlational, with a number of other confounding factors not controlled.

The study of the effects of experience with object training

Training with novel objects

Recently there are a number of fMRI training studies using novel objects. They have demonstrated a way to study the unique contributions of particular factors in object selectivity. The training studies focused on the effect of experience, and the use of novel objects allowed a test of the effect of experience, with less confounds from processing beyond visual object processing. The "Greeble" training studies are among the first of them (Gauthier & Tarr, 2002; Gauthier *et al.*, 1999; Rossion *et al.*, 2004).

The first fMRI training study of object expertise used novel, complex objects to test the hypothesis that specialization of the FFA stems from expertise with faces (Gauthier *et al.*, 1999). Participants learned to categorize 20 novel objects (Greebles – see Figure 1A) at an individual level in naming and name-verification tasks with feedback. Participants also learned labels to categorize the Greebles at higher level of abstraction, for example at the family level (defined by the central body shape of each object) or their gender (defined by the orientation of the four parts, all up or down). Greeble stimuli are highly distinctive and include highly redundant information diagnostic for each level of

16

categorization. That is, Greeble family can be determined by the shape of the central body, which can be captured in a relatively local part of the contour for this shape. Greeble gender can be perfectly determined by the upward or downward pointing orientation of any of the four protruding parts. Finally, individual Greebles can be individuated on the basis of any of several local features, as each part of each Greeble is unique within the set. Participants were scanned in repeated fMRI sessions before, during, and after the training with entirely novel Greebles. Training led to a decrease in the basic-level advantage (i.e., individuation became as fast as family categorization). Neurally, training increased activity in the right fusiform face area (FFA) during matching of upright Greebles compared to inverted Greebles. Passive viewing of Greebles also showed increases in FFA activity, compared to common objects. Behaviorally, Greeble experts showed modest increases in configural and holistic processing of untrained Greeble exemplars, and holistic processing was correlated with activity in the FFA (Gauthier & Tarr, 2002). Only the temporal lobe was imaged, and the only other areas showing increases in expertise were the left and right occipital gyrus, in area LOC (Talairach Coordinates: 48, -72, -6 and -43, -78, -4), and a face-selective region of the anterior right temporal lobe (41, -30, -16). The results are consistent with several studies of real-world expertise (e.g., with cars and birds) where objects of expertise recruit the FFA, and the FFA response is strongly correlated with behavioral measures of expertise (Gauthier *et al.*, 2005; Gauthier *et al.*, 2000a; Grelotti *et al.*, 2005; Xu, 2005). This similarity may not be surprising, as Greeble training, unlike any of the other trainings reviewed below, was designed to model real-world expertise (Gauthier & Tarr, 1997; Gauthier *et al.*, 1998), for example by using discrete stimuli with unique

features and by using a training criterion (decrease in basic-level advantage) that is a hallmark of real-world expertise (Tanaka & Taylor, 1991). It was concluded that face selectivity is a result of prolonged experience in performing fine-detailed discrimination with an object category, and thus similar behavioral and neural patterns can be developed for other objects (Bukach *et al.*, 2006; Gauthier, 2000).
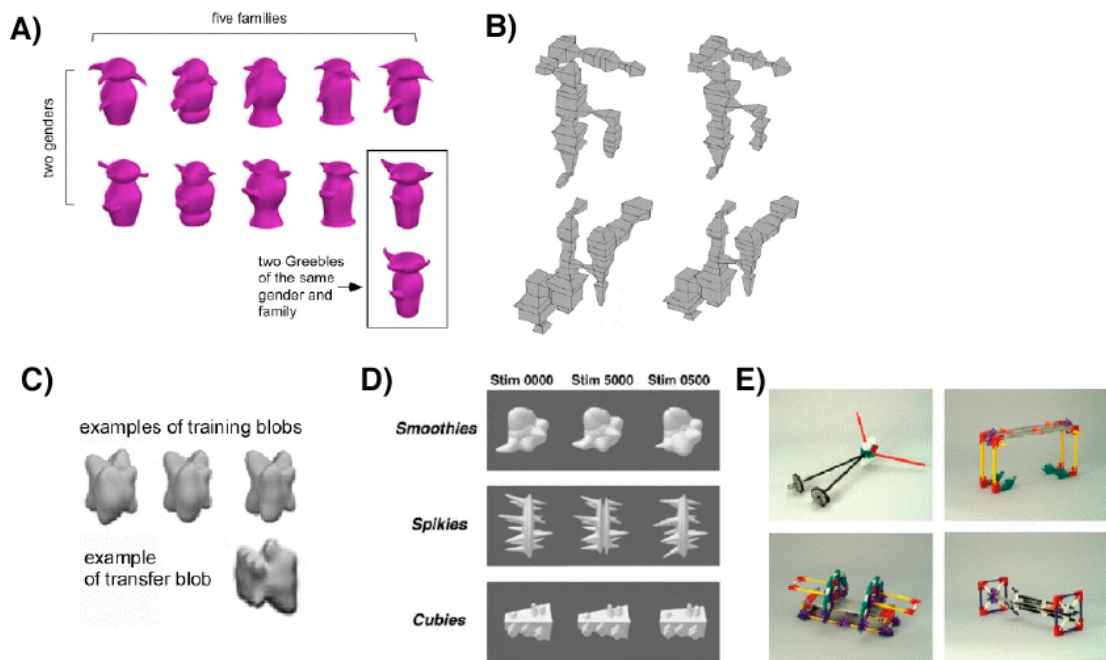
Figure 1. Stimuli used in different object training studies. Adapted from (A) Gauthier *et al.*, 1998, (B) Moore *et al.*, 2006, (C) Yue *et al.*, 2006, (D) Op de Beeck *et al.*, 2006, and (E) Weisburg *et al.*, 2006.

Similar training-induced changes in the face-selective regions have been found with a different object set and training procedure (Moore *et al.*, 2006). Participants trained for 10.5 hours with one of two categories of complex 3D novel objects (Figure 1B). Training used both categorization and discrimination tasks, and practice holding objects in memory for delayed matching. Two object categories (trained and untrained) were counterbalanced across participants. Following training, participants performed better in a working memory task with trained than untrained objects, and showed an inversion effect for trained objects only. During fMRI, participants matched two objects shown sequentially. The trained category led to higher activations than the untrained category in dorsolateral prefrontal cortex (DLPFC), posterior occipitotemporal cortex (OTC – Talairach coordinates: -35, -67, -4), and intraparietal sulcus (IPS) during working memory encoding and maintenance. Region of interest analyses also showed expertise effects of the same magnitude in the LOC and FFA (although more variable in the FFA), only during encoding. However, the changes did not correlate with behavioral measures of expertise.

Yue *et al.* (2006) also attempted to induce face-like training effects by training participants to discriminate 'blobs' (Figure 1C) in a match-to-sample task for 8 hours. A very specific behavioral marker of expertise was used, based on the hypothesis that face (but not object) representations retain the original spatial frequency and orientation information. Behaviorally, sensitivity to spatial frequency content was observed for faces but not for blobs, even in blob experts. Neurally, the FFA showed sensitivity to spatial frequency content of both faces and blobs, regardless of expertise. The right LOC was the only region reported to show a larger response to blobs with expertise, although it

showed no sensitivity to spatial frequency content. A potential problem with this study is that participants were trained with blobs from one part of a shape space and tested for expertise effects on a completely non-overlapping part of that space (see Fig1b). Thus, transfer across trained and tested objects could have been mainly domain-general rather than domain-specific. Whereas transfer to untrained exemplars is a hallmark of perceptual expertise, it is useful to test experts with transfer exemplars that are part of the trained shape space, and to use control stimuli not expected to show domain-specific effects. At least two studies of extant expertise do not report *any* expertise effect in the visual system, and in both cases test objects were from an unfamiliar domain (antique cars shown to modern car experts, Grill-Spector *et al.*, 2004; see also Rhodes *et al.*, 2004).

A recent discrimination training study (Op de Beeck *et al.*, 2006) was perhaps more successful in creating categories of novels objects (smoothies, spikies and cubies – see Figure 1D) that should reduce transfer to control objects. Participants performed a match-to-sample task with increasingly similar objects for 10 hours with only one of these categories. Although there was some learning for untrained categories, there was reliable training-specific improvement in shape discrimination. fMRI scans were performed before and after training, using a color change detection task with trained and untrained objects. More voxels preferred the trained to untrained classes after training. While area LO and posterior fusiform showed a preference for trained than untrained classes after training, the right FFA and V1/V2 did not. Interestingly, analyses of changes in distributed patterns of selectivity showed that experience produced qualitative changes in the spatial distribution of responses, rather than simply increasing the signal in voxels

that showed stimulus selectivity in novices. This suggests that pre-existing biases for object geometry may not play a large role in determining the end point of expert responses.

Finally, one more training study is reviewed here, although its design departs even more from those which use identity, discrimination or categorization training. In this study, learning involved real interactions with novel objects (Weisberg *et al.*, 2006). Participants were trained for 4 ½ hours to use novel objects made of construction toys (Figure 1E) so as to manipulate other items (e.g., crushing paper cup, lifting and transferring items between containers, etc.). Performance (in terms of time spent to finish each action task) improved within and across sessions. Before and after training, participants matched objects with different views of both trained and untrained objects during fMRI. Activity increased after training for both trained and untrained objects in the left middle temporal gyrus, IPS, and premotor cortex. In the anterior part of each of these regions however, increases were specific to trained objects. Also, activity increased (for both trained and untrained objects) in the medial part of the fusiform gyrus (thought to respond more to tools) but decreased in the lateral part (known to prefer faces and animate objects). Thus, like most other studies, this work shows some generalization from trained to untrained objects, although in this case different trained objects were as structurally different as different objects from the trained and untrained groups. This work reveals that experience with objects that are from the start easily discriminated, and which does not involve discrimination training but rather associations of functional information with each object can nonetheless lead to changes across the brain including in visual areas.

The need to go beyond one type of training

The fMRI training studies with novel objects have offered a promising way in showing the effect of experience on object recognition. These studies have shown, for example, that face-selective regions are not recruited exclusively for face processing only. Instead, given the right kind of experiences (like those in the Greeble studies), behaviors and neural activity associated with face-like perception can also be shown with non-face objects. In addition, higher-level processing (like processing of functional information of objects) can also be associated with an object category and affect what neural substrates are recruited for them.

Despite the success of these studies, I would argue that to better illuminate the unique contributions of a single type of factors, it is necessary to compare two or more types of training in a single study. First, these different training studies produced different results and it is hard to pinpoint the reason given the many differences in the stimuli (see in Figure 1) and procedure used. For instance, subordinate-level training results in behavioral expertise effects correlated with FFA activity increase in the Greeble study (Gauthier *et al.*, 1999) but did not work as well in other studies (Moore *et al.*, 2006; Op de Beeck *et al.*, 2006; Yue *et al.*, 2006). It could be that only objects that look animate like Greebles can result in FFA activations after training. Or the requirement to identify individuals with names may be the key. While Gauthier *et al.* (1999) and Moore *et al.* (2006) used names and found training effects in the FFA, Yue *et al.*, (2006) and Op de Beeck (2006) used a match-to-sample task with no names introduced and did not obtain any training effect in the FFA. It is also possible that the 'family' and 'gender' levels in the Greebles study created a "human" context, making one treat the Greebles more like

persons. These are just a few possible reasons among many. This demonstrates that even in novel object training, multiple factors (in terms of stimulus properties, recognition demand, semantic context, etc.) can account for a specific training result. With over one type of training in a study, one can specifically manipulate one type of factors while better controlling possible confounds. For example, it would be good to introduce different experiences using the exact same objects with no semantic context to probe the effect of task demand more directly, with stimulus properties and semantic association controlled.

Another reason for the need to introduce more than one type of training in a study is that to understand the functional specialization of the whole object recognition system one cannot just study one type of object selectivity. Because Gauthier *et al.*, (1999) found Greeble expertise to engage the FFA, several studies have queried training effects in this region. With more understanding of selectivity for other categories, it is good to start studying regions other than the face-selective areas. Weisberg (2006) made a good step in studying the selectivity for tools, but again it would be more fruitful to be able to show how in one study different types of training lead to different selectivity patterns.

Little imaging research has been conducted to contrast two types of training with novel objects. The closest study is one probing the neurophysiological correlates of categorization training at the basic and subordinate levels (Scott *et al.*, 2006). Fifteen participants learned in six sessions to categorize one of two bird types (owls and wading birds) at the basic level (e.g., owl) and the other type at a subordinate level (species, e.g., snowy owl) such that categorization speed at the subordinate level approached that at the basic level. Before and after training they performed subordinate-level matching with the

birds while an EEG was recorded. Accuracy improved only for the bird type trained at the subordinate level, and learning generalized to new exemplars of the same species as well as new species of the same type. Whereas the N170 component was enhanced regardless of training level, a later N250 component increased only for the bird type trained at the subordinate level. These results showed both similarities and differences in categorization training at different levels of abstraction. The study demonstrated a good way to study the effect of a specific factor (levels of categorization).

Summary and implications

The origins of object selectivity in the visual object recognition system have been a matter of intense debate. Although different types of factors have been suggested to account for such selectivity, it is difficulty to pinpoint the unique contribution of any single type of factors. Training studies with novel objects represents a promising approach to the research question, especially if the effects of different training procedures can be directly compared within a study, something that has not been commonplace.

This dissertation will capitalize on the success of previous Greeble training studies and go beyond them by examining the effects of different types of training on activity in the visual object recognition system. It has been shown that different object categories (e.g., faces, cars, birds and Greebles) can activate a common area (the FFA) apparently because of a common recognition demand (subordinate-level categorization). The current study will extend to other parts of the object recognition system by testing whether the same set of novel objects will recruit different areas with different recognition experiences during training. The specific question of interest is: Is

recognition experience alone sufficient to account for the different areas selective for faces and letters? CHAPTER II reports the specifics of the training paradigm and the behavioral effects, and CHAPTER III reports the neural training effects.

CHAPTER II

BEHAVIORAL CONSEQUENCES OF OBJECT RECOGNITION EXPERIENCES
MIRRORING FACE AND LETTER RECOGNITION

Introduction

Face and letter perception probably represent the most frequently studied and

contrasted cases of expert object recognition. As described in CHAPTER I, face and

letter perception have been associated with different behavioral and neural signatures of

expertise. One theory about the organization of the ventral visual object pathway posits

that the different neural substrates recruited for different object categories is the result of

prolonged experience in fulfilling different task demands associated with different

categories (Bukach *et al.*, 2006; Gauthier, 2000). Accordingly, the different experiences

associated with letters and faces may be sufficient to explain the different behaviors for

letter and face perception. Specifically, face perception requires fine-detail discrimination

among individuals within a homogeneous set, whereas letter perception requires one to

rapidly perform less detailed basic-level categorization for multiple instances

simultaneously (Wong & Gauthier, in press). However, faces and letters are also

drastically different in other aspects. Faces and letters contain completely different

stimulus properties, and are associated with different higher-level processes (e.g., social

and emotional processing for faces, phonological and linguistic processing for letters).

To directly test the contribution of different recognition experiences to the

different selectivity patterns found for faces and letters, I adopted a training paradigm

inspired by the Greeble training studies (Gauthier & Tarr, 1997; Gauthier *et al.*, 1999;

Gauthier *et al.*, 1998) and extended it. Two types of training experiences, mirroring face and letter perception respectively, were introduced to two groups of participants on the same set of objects. A set of novel, computer-generated objects ('Ziggerins') were presented during one of the two training procedures that mirrored the experience of face and letter recognition. With the same set of novel objects used across the two training procedures, the study allowed a relatively pure test of the effect of recognition experience on neural activity, without the confounds caused by differences in physical properties and higher-level processing associated with familiar objects.

"Face-like" training was defined as learning to discriminate objects that share a common part structure. This sort of training was designed to model real-world subordinate-level object and face expertise (Tanaka, 2001; Tanaka & Taylor, 1991) and has been used before and shown to result in a variety of behavioral and neural effects that are typically elicited by faces (Gauthier & Tarr, 2002; Gauthier *et al.*, 1999; Rossion *et al.*, 2004; Scott *et al.*, 2006).

"Letter-like" training was defined as learning to categorize objects based on common part structures (at the basic level). It is expected that successful training should result in an enhancement of the basic-level advantage (based on Wong & Gauthier, in press) and higher ability in making use of regularities at the subordinate level (Gauthier *et al.*, 2006; Sanocki, 1987, 1988, 1991). Thus, while the goal of letter perception, in the context of reading expertise, requires distinguishing objects at the basic level, the letter input differs from typical object input in that it offers regularities in style (font, shape, size) that can be used for more efficient basic-level access (Gauthier *et al.*, 2006; Wong & Gauthier, in press).

Letter-like training of this sort has not been implemented before, although related behavioral and neural expertise effects have been demonstrated with Chinese characters (Gauthier *et al.*, 2006; Wong & Gauthier, in press; Wong *et al.*, submitted). Following the strategy of Greeble training that caused recruitment of face-like processing without modeling all aspects of our rich experience with faces, the letter-like training implemented in this study did not try to model all aspects of reading experience either. For instance, mapping of objects onto a phonological system is purposefully not part of this training. In that sense, the operationalization of this training represents a hypothesis about whether some, but not all, aspects of our experience with letters are important for specialization of the letter- and string-selective areas in the ventral system. Importantly, we would expect that other areas engaged by words, such as the visual word form area and the brain-wide network involved in reading, require more "reading-like" training. While this study was inspired by the two different types of expertise that are face and letter perception, it was unclear to what extent we could get all of the training details right so as to achieve behavioral and neural outcomes which are truly "face-like" and "letter-like". Indeed, one lesson from the literature review is that it remains very difficult to define what aspect of the training is sufficient or necessary for a given training outcome. In that spirit, a more realistic goal is to create two training experiences with the same category of objects that can result in different processing strategies and ultimately, result in the recruitment of different neural substrates.

As discussed before, existing work has not conclusively answered why faces and letters activate selectively different areas in the ventral occipitotemporal cortex. Some theories also do not predict whether the areas selective for faces and letters would be

recruited respectively for the two types of training. For instance, the module-based view notes the importance of experience on the recruitment of a module, but does not explicitly state whether and why different types of recognition experience results in different modules. The feature-based view would be consistent with different results. If feature representations in the ventral occipitotemporal cortex are fixed, then the two types of training would result in the same area recruited because of the same object set used. Alternatively, if different types of recognition experience can change the features extracted even for the same objects, then activation of two areas would be expected. The process-based view would predict recruitment of different areas as long as recognition demand is one of the organizing principles of the ventral occipitotemporal cortex. Results of the proposed study, therefore, can offer important constraints for existing theories of the organization of the visual object recognition system.

Overview of the Study

The object set

One main challenge of the current study was to create a set of stimuli suitable for both face-like and letter-like training. Face-like training (or Greeble-like training) requires structurally similar objects to be discriminated. Whereas different Greeble families have different body shapes (Figure 1A), they share a part structure. In earlier work (Wong & Gauthier, in press) we have argued that letter-expertise develops at the basic level, and can make use of regularities in style, which requires objects that can be grouped into different basic levels but share stylistic properties. The main challenge was thus to preserve important constraints of the Greeble training while building into the

object set stylistic properties as well as more variability at the family level than there was with Greebles, as necessary for letter-like training.

The stimuli used were novel objects ("Ziggerins") composed of a limited set of volumes similar to the "geons" described in Biederman's recognition-by-components theory (Biederman, 1987). The asymmetric, inanimate, three-dimensional Ziggerins were designed in order to reduce their visual similarity to faces and letters, so that they would less likely be associated with faces or with letters a priori. They were created in such a way to allow categorization at different levels of abstraction (see Figure 2). The set contained six 'classes' that differ from each other in the composing parts, representing six different basic-level categories. Within each class, Ziggerins shared a similar basic structure and differed in the detailed manifestations of the parts ('styles'). The styles applied to all classes such that there were 12 styles each with the six classes of Ziggerins. The inclusion of styles was important for the letter-like training as described below.

Training

The letter-like training was designed to mirror the experience of letter perception. Letter-like training required discrimination among different classes of Ziggerins and thus basic-level categorization. The different classes and styles in the Ziggerin set were analogous to the different letter identities in different fonts in letter perception. As seen in Figure 2, each row represents Ziggerins in different styles within a class, similar to the different fonts for a particular letter. Each column represents the different classes with the same style, corresponding to different letters in the same font. On the other hand, a large part of letter-like training involved presentation of different classes of Ziggerins in the

same style presented in a matrix, similar to how letters are presented during reading. Importantly, this was not word recognition or orthography training. Even though the Ziggerins were presented in clusters, recognition was always required for individual Ziggerins only. The learning of Ziggerin sequences should be minimized, as there were no names given to any sequence of Ziggerins and the order of Ziggerins were randomized during training.

The design of face-like training was based on previous Greeble training studies (Gauthier & Tarr, 1997, 2002; Gauthier *et al.*, 1999; Gauthier *et al.*, 1998; Rossion *et al.*, 2004), with a few differences. First, the current study used a different set of objects from the Greeble study. Second, the Ziggerins in the current study, with different parts across classes, formed a more heterogeneous set than the Greebles that share the same first-order configuration of parts (compare Figures 1A and 2). This was because in the current study, the same objects were used also for letter-like training and thus had to include several classes. The heterogeneity of the Ziggerin set could lower the demand for subordinate-level categorization and reduce the chance of obtaining face-like training effects as seen in previous Greeble studies. Therefore, to impose a greater demand for subordinate-level categorization in face-like training, special arrangements were made to the non-targets in different training tasks (as described in more detail later). For example, in one of the training tasks where participants had to choose one out of two alternative Ziggerins that match a label, the non-target could be another Ziggerin within the same class as the target, or the exact same target Ziggerin with just a random part of it slightly rotated or moved upwards or downwards. This was expected to raise the demand for detailed categorization, since the targets and non-targets would be more similar and the

31

participants would have to attend to all parts of a Ziggerin. Finally, an additional task was

added to face-like training to further emphasize the demand for within-class

discrimination. The task involved selecting one out of two Ziggerins that matched with a

label (an individual Ziggerin's name). This task has been used successfully in previous

subordinate-level training with other objects like birds and cars (Tanaka *et al.*, 2005).

Overall, the face-like training was designed to match previous Greeble training as closely

as possible while allowing for comparisons with the letter-like training procedure in this

study.

      To summarize, the two types of training differed in three main ways. The first

difference was the level of categorization. Letter-like training required discrimination

among different classes of Ziggerins and thus basic-level categorization. Face-like

training emphasized more detailed, subordinate-level discrimination among Ziggerins of

different styles within each class. Secondly, to mirror the fact that letter recognition often

requires rapid recognition of spatially organized clusters of letters presented, letter-like

training also required participants to recognize multiple Ziggerins presented within

spatially organized clusters in the same style. Certainly, during face perception there are

situations when individual persons have to be picked out from the crowd. But this is

relatively infrequent compared with the frequency at which letters occur in words and

texts. Therefore during face-like training Ziggerins did not appear in clusters. Thirdly,

during letter-like training when the Ziggerins appear in clusters they were always in the

same style. This style regularity did not occur during face-like training as the Ziggerins

did not occur in clusters.

      Overall, as a first study contrasting face-like and letter-like training, the current

study introduced differences in the recognition experience between the two types of training, while minimizing the effects of stimulus differences and higher-level cognitive factors. Note that the goal of this study was to recreate face-like and letter-like expertise by manipulating recognition experiences, rather than to study the effect of one particular factor (e.g., level of categorization) in the consequent recruitment of neural substrates. Compared with previous studies that contrast face and letter expertise, this study went a large step forward by excluding the effects of stimulus differences and high-level factors. Therefore, instead of trying to equate everything but one factor between the face-like and letter-like training, a number of differences in experience was allowed.

Pre- and post-training tests

The behavioral pre- and post-training tests were used to assess the effects of the two types of training. The critical task for assessing the effect of training was the sequential matching task, which measured categorization performance at the basic and subordinate levels. While objects are typically recognized more efficiently at the basic than subordinate level, it has been shown that performance at the two levels are not different for perception of faces (Tanaka, 2001). Also, Greeble training has been shown to lead to a reduction of the basic-level advantage over the subordinate level (Gauthier & Tarr, 1997; Gauthier *et al.*, 1998). In contrast, a larger basic-level advantage over the subordinate level has been shown for familiar vs. unfamiliar characters (Wong & Gauthier, in press). It was therefore expected that whereas the face-like training would lead to a reduction of the basic-level advantage for post-test compared with pre-test, the letter-like training would lead to an increase of the basic-level advantage. In other words,

the letter-like training should lead to a larger improvement in basic-level performance than the face-like training, whereas the face-like training should lead to a larger improvement in subordinate-level performance than the letter-like training. This task was the only post-test task introduced before training.

Two other tasks were used to tap into other possible effects of face-like and letter-like training. A part-matching task was used to measure any changes in holistic processing for the two groups. The task has been used in previous studies and revealed greater holistic processing for face perception and subordinate-level expertise (Gauthier & Tarr, 2002). The part-matching task required selective attention to parts of an object, providing a measure of the effect of the irrelevant part on matching of the target part. It has been shown that after Greeble training, there is an increase in the influence of the irrelevant part on target part matching, as indicated by the better performance when both the target and bottom parts led to the same response (both same or both different) than when the parts led to different responses (Gauthier & Tarr, 2002). This congruence effect has been used as an operational definition of holistic processing, because even when the irrelevant part should be ignored it nonetheless affects target part matching. Also, a larger congruence effect has been found with faces when the top and bottom halves were aligned than misaligned. This alignment effect indicates sensitivity to configurations of parts in an object.

Finally, a triplet naming task was introduced to measure the degree in which regularity in styles facilitates object recognition at a basic level, which has been suggested as one behavioral marker of expert letter perception (Gauthier *et al.*, 2006). The triplet-naming task is devised to measure the facilitation of style regularity for object

recognition and how it changes after training. It has been shown that recognition of letters in a string is more efficient when the letters are of the same font as opposed to mixed fonts (Sanocki, 1987, 1988, 1991). Our previous work has shown that such font regularity effect is greater for someone who is familiar with the characters in a writing system than someone who is not (Gauthier *et al.*, 2006).

It should be noted that these two tasks were introduced after the training. The main reason was that including them before training would have resulted in a very long pre-training test period (5.5-6 hours). Participants would have been likely to learn a lot during this period and this could have greatly reduced the differences between the two training groups.

Methods

Participants

Participants were 36 undergraduate students, graduate students, or employees of Vanderbilt University. Eighteen participants (12 females, age M = 24.06, SD = 5.92) were assigned to the face-like training group, and 18 (ten females, age M = 23.33, SD = 5.63) to the letter-like training group. All reported normal or corrected-to-normal vision, and none have been exposed to the set of Ziggerins before. They were each compensated for $12 per hour for their time in behavioral training and testing.

Stimuli and material

All testing was conducted on Mac computers using Matlab (MathWorks, Natick MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). Seventy-

two novel objects (Ziggerins) were created (six classes each with 12 styles) using Carrara 5 software (MetaCreations). The six classes differed from each other in the parts involved and the general spatial relations among the parts. The 12 styles were formed by variations of parts along three shape dimensions (Figure 2 & Table 1). Note that the same style variations applied to all classes. For example, Ziggerins in the first column of Figure 2 belonged to different classes but were in the same style.

While the Ziggerins were designed to contain classes corresponding to basic-level categories and styles analogous to font variations of each letter, it is important to ensure that the class structure and style variations are readily recognizable. For this purpose a card-sorting task was conducted. Thirteen people (who did not participate in subsequent training) were asked to sort 72 cards (each with one Ziggerin) in three ways. First, they were allowed to freely sort the cards into categories with no restriction on the number of categories or the number of cards in each category. Four out of the 13 people sorted the Ziggerins into six categories corresponding exactly to the six classes during free sorting, and others broke the Ziggerins down into more categories (from 7 to 18) such that each category contained Ziggerins of different styles within one class. Secondly, they were told that the Ziggerin set was supposed to contain six classes each with 12 members, and were asked to sort the cards accordingly. Twelve out of 13 people identified the class structure perfectly, and the remaining person made a mistake in two out of the 72 Ziggerins. Thirdly and lastly, they were asked to sort the 12 members within each class such that the same style across different classes were aligned (i.e., to display the cards like Figure 2). Twelve out of 13 people identified the styles perfectly, with the remaining person making a mistake in two of the Ziggerins. Several people were asked on what

basis they performed the sorting, and they reported the difference in the general structure

as the basis for class variations and could list the dimensions at which cross section

shapes differed as the basis for style variations. Overall, a person seeing the Ziggerins for

the first time can readily recognize the class structure and style variations.
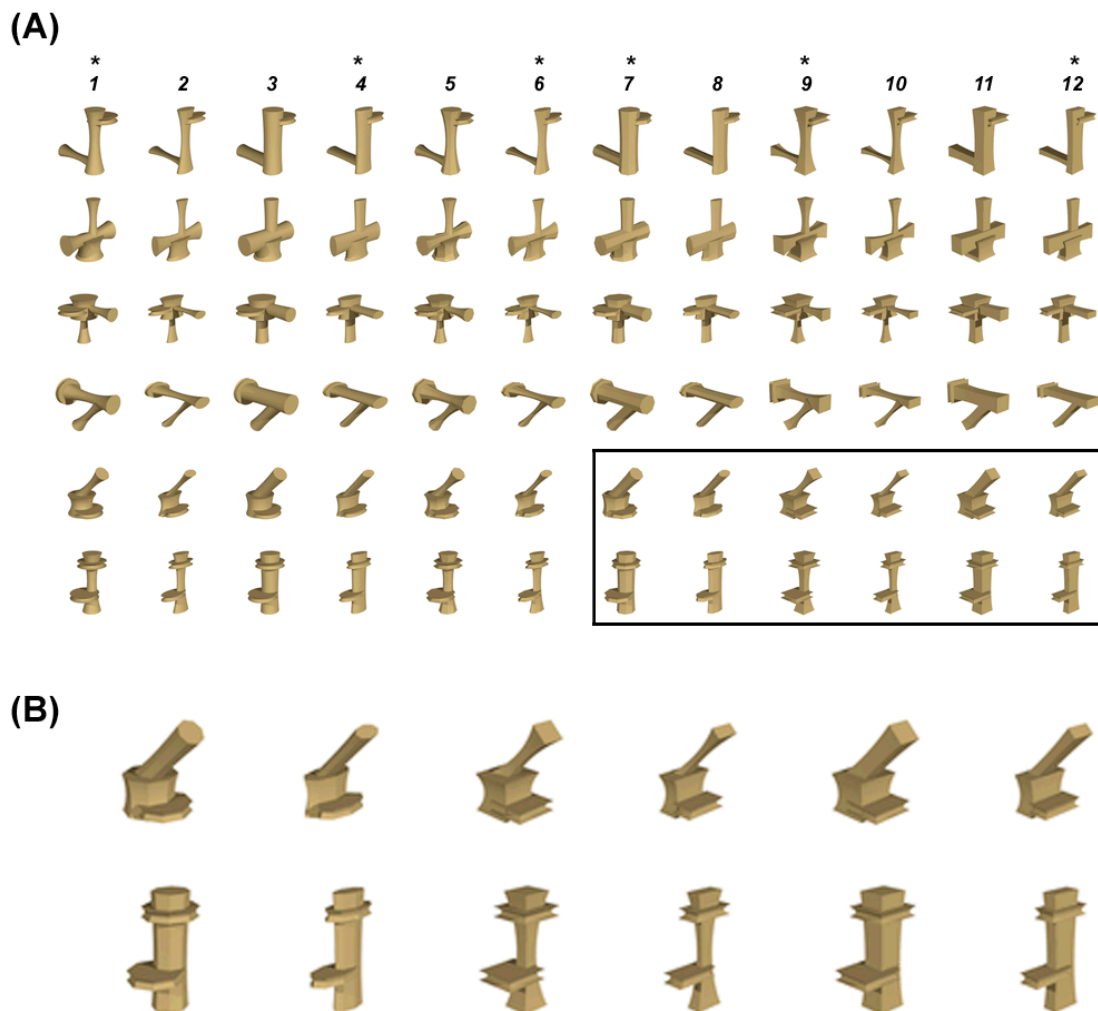
Figure 2. The Ziggerins. (A) The whole set, with the rows representing the six classes and the columns representing the 12 styles. The stars indicate the six styles selected for training for half of the participants. The unselected styles were the transfer set for this participant. The Ziggerin images at the lower right corner are enlarged in (B).

Table 1. Formation of the 12 styles based on differences in cross section shapes. The style numbers correspond to the 12 columns (from left to right) in Figure 2.

Style	Shape (1 = round; 2 = octagon; 3 = square)	Size (1 = varying; 2 = constant)	Aspect ratio (1 = thick; 2 = thin)
1	1	1	1
2	1	1	2
3	1	2	1
4	1	2	2
5	2	1	1
6	2	1	2
7	2	2	1
8	2	2	2
9	3	1	1
10	3	1	2
11	3	2	1
12	3	2	2

During training each participant only half of the Ziggerins (six styles, with styles 1, 4, 6, 7, 9, 12 for half of the participants and styles 2, 3, 5, 8, 10, 11 for the other half) were used during training, while only the transfer Ziggerins (those not used in training) were used during pre- and post-training tests, and both trained and transfer Ziggerins were used during fMRI scans. The training set and the transfer set was assigned such that different levels of each dimension were equally represented within each set.

The Ziggerins were rendered with realistic lighting and shading, and given a yellow plastic texture on a white background. Each object image was about 4 × 4 cm large (a visual angle of about 4 degrees from a viewing distance of 60 cm).

Names (e.g., Xedo, Kimo) were assigned for the classes and individuals randomly for each participant. The letter-like group learned six class names for the 36 Ziggerins in

the training set (Figure 3B). The face-like group learned individual names for 18 of the 36 Ziggerins (Figure 3A). The 18 Ziggerins with individual names were selected from six classes, each with three individuals. The selection was counterbalanced across participants.
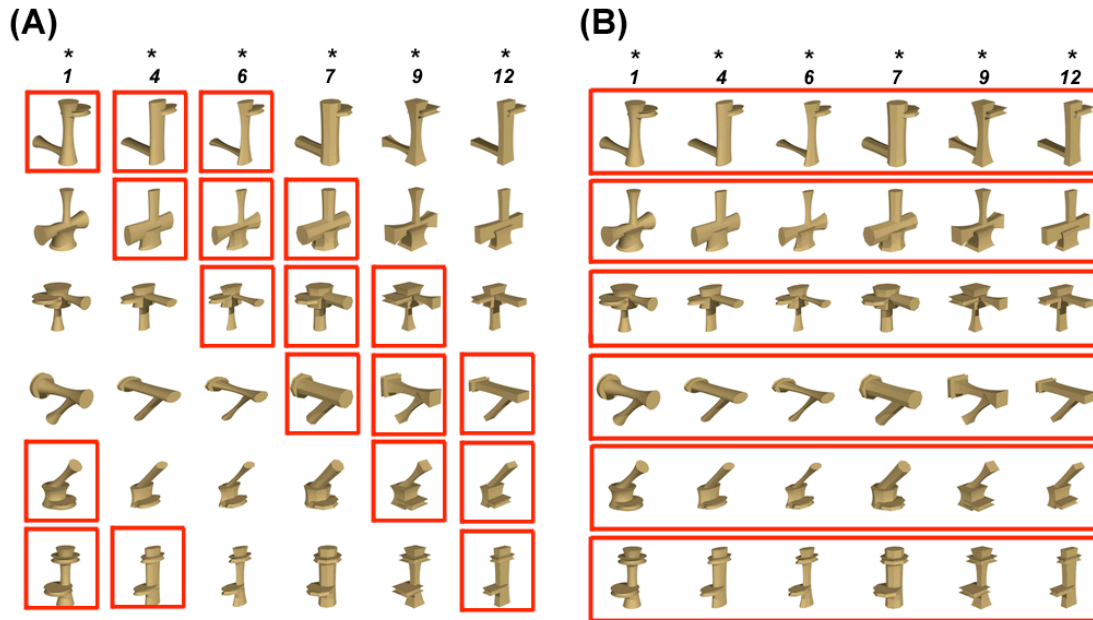
Figure 3. An example training set for each type of training. (A) Face-like group learned 18 names, three for each class. (B) Letter-like group learned to categorize the 36 objects into six classes.

Face-like Training

In ten hour-long training sessions, the face-like training required the participants to recognize the Ziggerins at an individual level by learning the names for 18 of the 36 Ziggerins in the training set. The remaining 18 Ziggerins served as distractors in different tasks as described later. Participants started with two classes in the first training session,

with two more classes added in the second session, and the final two added in the third

session. Each session included three tasks: naming, verification, and matching (Figure 4).

In all tasks both accuracy and speed were stressed. At the end of each block, the average

accuracy and speed for the block were shown to inform the participants of their progress.

To motivate the participants, from session 4 onwards participants also saw, after each

block, a rank table showing performances for the 10 best blocks for each task.

Participants were encouraged to break their record by improving both their speed while

maintaining near-perfect accuracy.
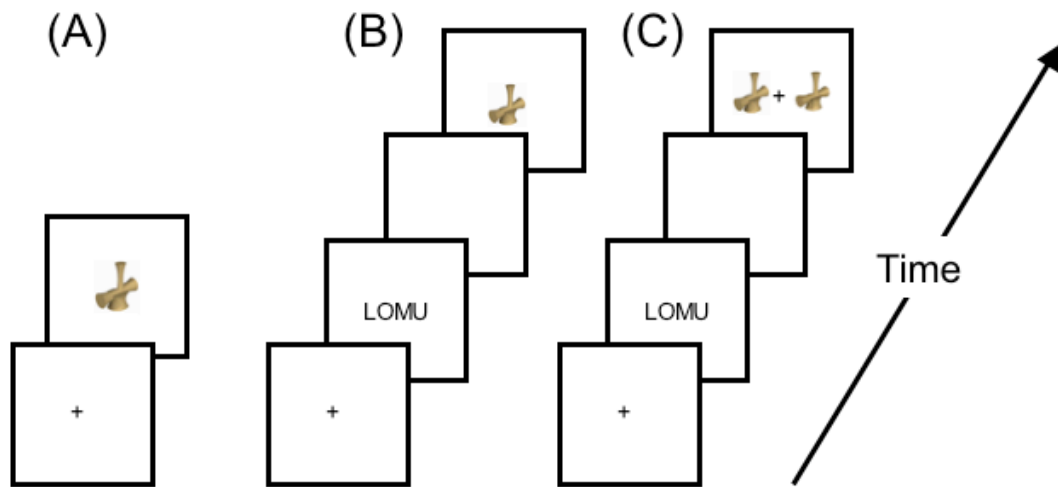
Figure 4. The tasks in face-like training: (A) In the naming task, participants had to press the key for the first letter of a Ziggerin's name. (B) In a verification task, partcipants had to judge if a Ziggerin matched with a label. (C) In a matching task, participants had to judge which one of the two Ziggerins matched with the label.

Naming

Participants learned to recognize the Ziggerins with individual names. In each trial a fixation cross appeared at the centre of the screen for 250 ms, followed by a Ziggerin image until response (Figure 4A). Participants had to type the first letter of the name (e.g., "P" for "pimo"). Labels of 18 Ziggerins (in six classes each with three styles) were learned. The remaining 18 Ziggerins in the trained set were also shown as distractors in 10% of the trials and participants had to type the space bar to indicate that they had no names. Upon an incorrect response both the Ziggerin image and its name were shown until participants pressed the space bar to continue. In the first block of each session both the Ziggerin and its name were shown and participants were encouraged to take as much time as necessary to study the Ziggerin and the name before responding. There were 360 trials in each session. In session 1, six Ziggerins with names were shown for 54 times each, and the other six Ziggerin distractors within the same classes were shown for 6 times each. In session 2, 12 Ziggerins with names were presented for 27 times each, and the other 12 Ziggerin distractors within the same classes were shown for 3 times each. In sessions 3-10, 18 Ziggerins with names were shown for 18 times each, and the other 18 Ziggerin distractors within the same classes were shown for 2 times each.

Verification

Participants judged if a label matched the image of a Ziggerin. Each trial began with a fixation for 500 ms and then a label presented for 1000 ms, followed by a blank screen for 200 ms and a Ziggerin image until response (Figure 4B). Participants pressed

the "1" or "2" key on the number pad to indicate a "same" or "different" judgment. In "different" trials, the Ziggerin shown was either within the same class as the target Ziggerin, or one of those extra Ziggerins created by taking the target Ziggerin and moving a part of it to create a slight configural change. Upon an incorrect response participants would see the Ziggerin image and the correct answer until participants pressed the space bar to continue. There were 288 trials in each session. In session 1, six Ziggerin names appeared in 24 "same" and 24 "different" trials each. In session 2, 12 Ziggerin names appeared in 12 "same" and 12 "different" trials each. In sessions 3-10, 18 Ziggerin names appeared in eight "same" and eight "different" trials each.

Matching

Participants had to choose the target Ziggerin (out of two alternatives) that matched with a label. In each trial, a fixation cross appeared for 500 ms. Then a label appeared for 1000 ms, followed by a blank screen for 200 ms and then two Ziggerin images side by side until response (Figure 4C). Participants had to indicate the position of the target Ziggerin that matched the label by pressing "1" (for the left) or "2" (for the right). In 25% of the trials neither Ziggerins were the target and participants had to press the space bar. A distractor that did not match the label was either another Ziggerin within the same class, or the same as the target with just a part of it moved slightly. Upon an incorrect response both the Ziggerin images and the correct answer were shown as feedback until participants pressed the space bar to continue. There were 288 trials in each session. In session 1, six Ziggerins appeared as the target in 36 trials each. In session 2, 12 Ziggerins appeared as the target in 18 trials each. In sessions 3-10, 18 Ziggerins

appeared as the target for 12 trials each. In all sessions there were 72 trials where both Ziggerins were distractors.

Letter-like Training

During the ten hour-long training sessions, the letter-like training participants learned to rapidly recognize a series of Ziggerins (six classes in six styles) at the class level. To match with the number of objects introduced in the first three sessions during face-like training, the letter-like training group started with two styles of each class of objects in the first training session, with two more styles added in the second session, and the final two styles added in the third session. Each session contained three different tasks: naming, verification, and matrix scanning. Similar to face-like training, in all tasks both accuracy and speed were stressed. At the end of each block, the average accuracy and speed for the block were shown to inform the participants of their progress. To motivate the participants, from session 4 onwards participants also saw, after each block, a rank table showing performances for the 10 best blocks for each task. Participants were encouraged to break their record by improving both their speed while maintaining near-perfect accuracy.

Naming

This was the same task as the naming in face-like training, except that participants in the letter-like group learned to name the Ziggerins at the class level. The presentation sequence was the same as the naming task in the face-like group. There were 360 trials in the first three sessions. As it was easy to learn six names, from session 4 onwards there

were only 216 trials per session, to leave more training time for the harder matrix

scanning task (as described later). In session 1, 12 Ziggerins appeared for 30 times each.

In session 2, 24 Ziggerins appeared for 15 times each. In session 3, 36 Ziggerins appeared

for ten times each. In sessions 4-10, 36 Ziggerins appeared for 6 times each.

Verification

This was the same task as the verification in face-like training, except that the

labels were class names instead of individual names. There were 288 trials in each of the

first three sessions, and the number was reduced to 216 trials from session 4 onwards to

leave more training time for the matrix scanning task. In session 1, 12 Ziggerins appeared

in 12 "same" and 12 "different" trials each. In session 2, 24 Ziggerins appeared in six

"same" and six "different trials" each. In session 3, 36 Ziggerins appeared in four "same"

and four "different" trials each. In sessions 4-10, 36 Ziggerins appeared in three "same"

and three "different" trials each.

Matrix scanning

Participants learned in this task to rapidly perform class-level judgment on a large

number of Ziggerins with a coherent style, similar to what often occurs for letter

perception (Figure 5). Participants pressed the space bar to begin each trial. A matrix of

Ziggerins then appeared. In order to provide a scanning experience similar to left-to-right,

top-to-bottom letter perception in reading, participants were told to do the following:

"Each time you will see a 6-row-by-10-column array of objects. The upper left

object will be your first target. Scan the matrix from left to right, top to bottom until you

find your target. The object that immediately follows it will be your new target. Keep

scanning the matrix until you find this new target. And the object that immediately

follows it will be your next target. Continue this process until you get to the end of the

matrix. Press the space bar as soon as you get to the end. After pressing the space bar you

should type in the first letter of the last target you are searching for."
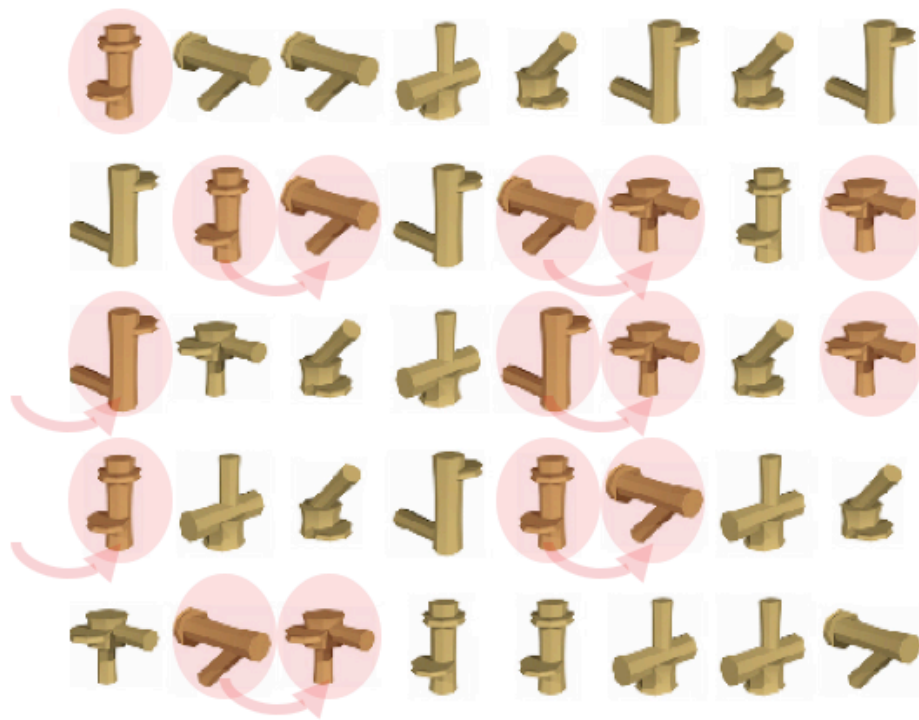
Figure 5. An example trial in the matrix scanning task in letter-like training. Participants
started with the top left Ziggerin as the target and searched through the matrix from left
to right, top to bottom for that target. Once they found the target they would have to
change their target to the Ziggerin immediately after. The pink ovals represent the
targets and the arrows represent the target shifts.

There were five to seven target shifts on each trial. There were 84 trials in each of the first three sessions, and 112 trials for sessions 4-10. The 1036 matrices used in the whole training process were carefully generated with the following considerations: (1) Each duplet and triplet combination should occur as frequently as every other combination to avoid learning of particular sequences. (2) Each Ziggerin should have an equal chance of becoming the final target. (3) Each Ziggerin should occur as frequently as every other Ziggerin. (4) The same number of matrices should be devoted to each style. To meet these criteria 1000 sets of 1036 matrices were generated in random and the set of 1036 matrices that best matched the above criteria were selected for training for all participants. The order of matrices presented were pseudo-randomized for each participants, such that only two styles were shown in the first session, with two more added in the second session and two more added in the third session.

Pre- and post-training tests

In these tests the transfer set of 36 Ziggerins not shown during training were used to assess training effects.

Sequential matching

Participant judged if two sequentially presented images contained the same or different individual Ziggerins (individual judgment), or if they contained Ziggerins in the same or different class (class judgment). To demonstrate the meaning of a "class", before the task the participants were shown a sheet containing all the Ziggerins (similar to Figure 2), and were told that objects within each row formed a class. In each trial

46

participants viewed a fixation cross for 500 ms and then the first image for 800 ms,

followed by a pattern mask for 500 ms and then a second image (Figure 6). Participants

had to press "1" and "2" for "same" and "different" responses respectively. Depending on

the block participants made class or individual judgment. For class judgment, a "same"

trial would have two Ziggerins in the same class presented and a "different" trial would

show two Ziggerins in different classes. For individual judgment, a "same" trial would

present two identical Ziggerins while a "different" trial would show two Ziggerins from

the same class but with different style. To encourage matching of objects and not images,

one Ziggerin was always larger than the other (three sizes were used: 3.2 × 3.2, 4 × 4, 4.8

× 4.8 cm), and participants were asked to ignore size differences. Speed and accuracy

were both emphasized. No feedback was given. There were 432 class-level and 432
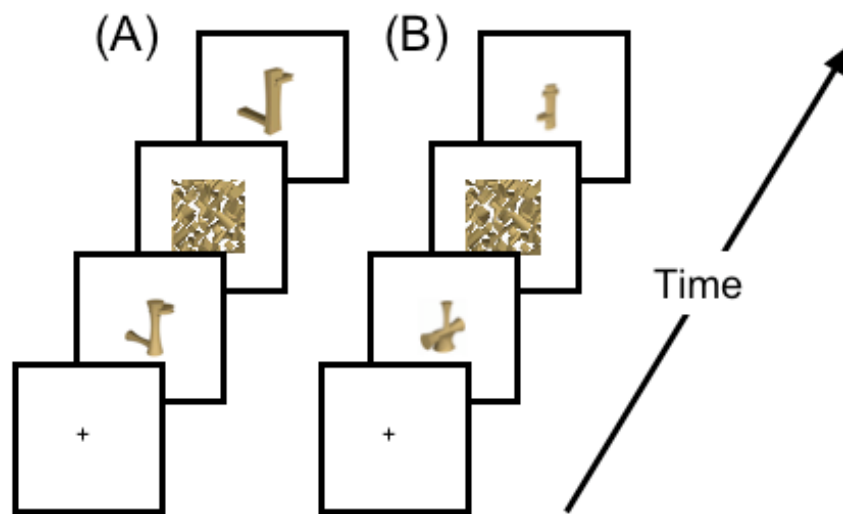
Figure 6. An example trial in sequential matching task. The two images in a trial were always different in size. (A) An individual-level trial where one had to make a "different" response. (B) A class-level trial where one had to make a "different" response. Note that the two Ziggerins in a trial were always of different sizes.

individual-level trials (half same half different), separated into 12 blocks of 72 trials, with class-level blocks alternating with individual-level blocks. Participants were notified at the beginning of each block which task would be required, and the words "same class?" or "same individual?" were presented above the images throughout a trial to remind them of the task.

Part-matching

This task was used to measure the holistic and configural effects after training. In this task two composites (each with the top half of one Ziggerin and the bottom of another) were presented sequentially and participants had to judge if the target parts (tops or bottoms) of the two composites were the same or different. Composites were made of the tops and bottoms from Ziggerins in the same class but different styles (Figure 7A) or from Ziggerins in different classes (Figure 7C). Participants viewed a fixation cross for 500 ms and then the first composite for 400 ms. A mask then appeared for 3000 ms, followed by the second composite. A cue on the top (or bottom) appeared at the last 500 ms of the mask and during the second composite. Participants had to indicate by key press if the top (or bottom) halves of the two composites are the same within 1000 ms. No feedback was given.

Apart from the between-participant Group variable (face-like, letter-like), there were three main within-participant variables. One variable was distractor difference level. In half of the trials, the distractor parts could come from Ziggerins in the same class with different styles (within-class distractor changes, see Figures 7A and 7B). In other trials, the distractor parts could come from Ziggerins in different classes (between-class
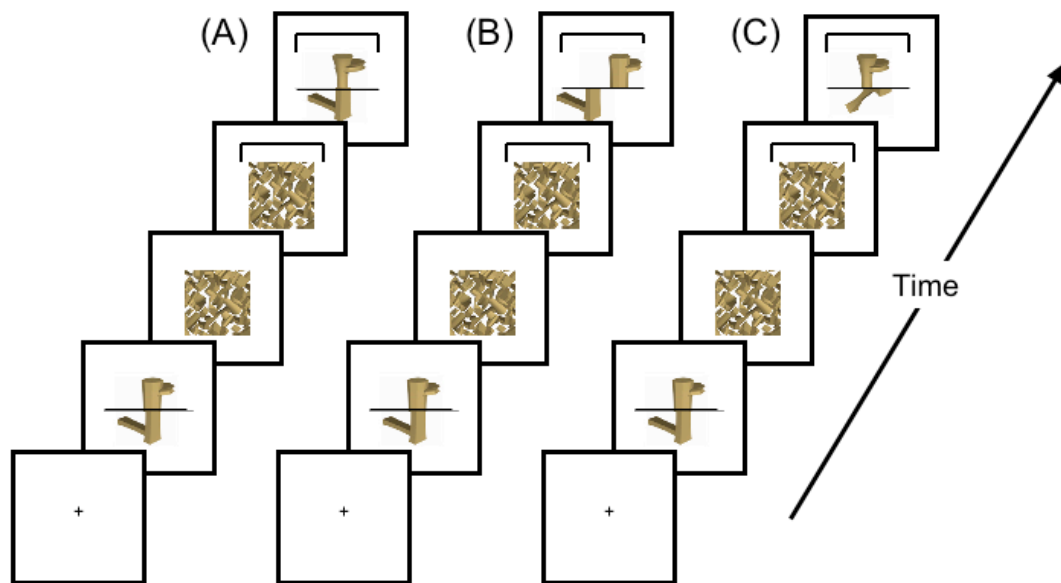
Figure 7. Three example trials of the part matching task, all requiring top matching. (A) A trial where the top and bottom of the second composite were aligned. It was an incongruent trial as the tops differed but the bottom parts were the same. (B) A trial where the top and bottom of the second composite were misaligned. It was a congruent trial as both the tops and the bottoms were the same. (C) A trial where the distractor parts (bottoms) of the two composites differed in class. It was a congruent trial as both the tops and the bottoms were different.

Ziggerin changes, see Figure 7C). Another variable was alignment, i.e., the halves of the

second composite could be either aligned (Figures 7A and 7C) or misaligned (Figure 7B).

The third variable was congruency, i.e., whether the target parts and the distractor parts of

the two composites would lead to compatible responses (e.g., both same or both different,

see Figures 7B and 7C) or not (e.g., target parts same and distractor parts different, or

target parts different and distractor parts same, see Figure 7A). Half of the trials had top

target matching and half bottom target matching. Half of the trials were "same" trials and

half were "different" trials.

The Distractor Difference Level × Alignment × Congruence × Target Part × Same/Different design resulted in 32 conditions, each with 18 trials, leading to a total of 576 trials. The different types of trials were intermixed, and presented in 8 blocks of 72 trials.

Triplet recognition

This is for testing if style regularity facilitates Ziggerin perception after the two types of training (Figure 8). Participants first viewed a mask for 1 sec and then three Ziggerins simultaneously for a certain duration (calibrated for each participant to avoid ceiling and floor performance) followed by a 200-ms mask. Then six Ziggerins were shown, two at each position, until response. Participants had to judge which one of the two alternatives was the Ziggerin that appeared in that location in the first display. They had to select the upper or lower alternative by typing the upper or lower arrow keys, and they had to respond to the leftmost location first, followed by the middle and the rightmost positions. Accuracy was emphasized and participants could take as much time as needed. No feedback was given. The three objects were always in different classes with either the same or mixed styles.

There were two phases in this task: (1) Calibration phase – A staircase procedure was introduced with 10 blocks of 12 trials to find the presentation time of the three Ziggerins at which average performance lingered around 2.25 Ziggerins recognized. The presentation time started at 600 ms for all participants, and the step size changed from 220 ms at 600ms or above to 20 ms at 100 ms or below. (2) Experimental phase – In this

phase the three Ziggerins were presented at a constant presentation time decided in the calibration phase for 360 trials. In both phases, half of the trials had the three Ziggerins in the same style (Figure 7A), while half had them in three different styles (Figure 7B).
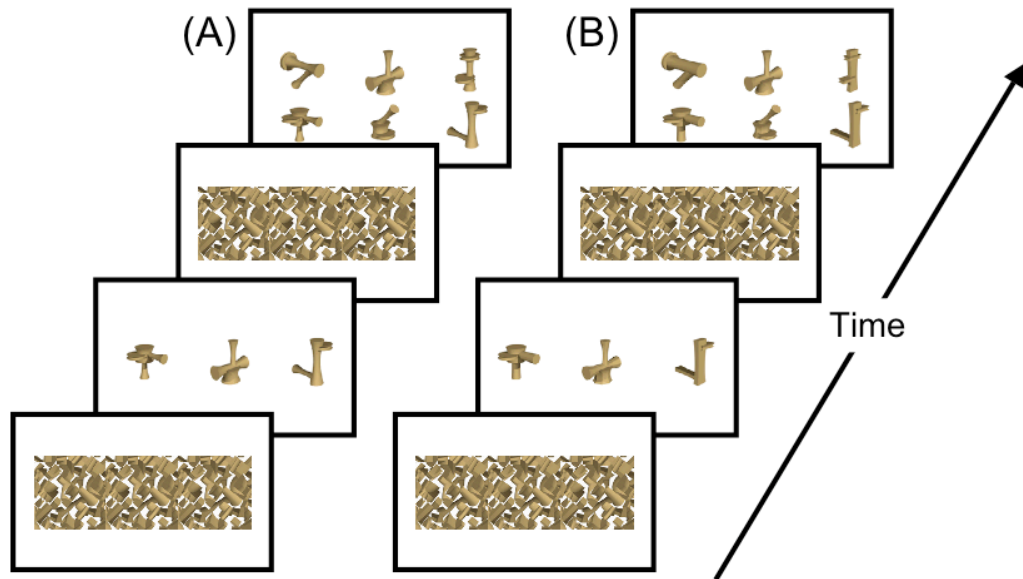
Figure 8. The triplet recognition task. (A) An example trial with all the Ziggerins in the same style. (B) A trial with the three Ziggerins in three different styles.

Predictions

It was expected that both face-like and letter-like groups would improve in terms of response time and accuracy in all tasks during training.

However, different effects were predicted for the groups in the pre- and post-training tasks. For the sequential matching task, it was expected that before training, both groups would show a comparable basic-level advantage, i.e., matching should be faster and more accurate for the basic than subordinate level. After training, the face-like group

should show a smaller basic-level advantage, as the group was trained at the subordinate level and thus should improve more at that level. The letter-like group, in contrast, should show a larger basic-level advantage, since the group was trained at the basic level and should improve more at that level.

The part matching task was performed only after training. The face-like group was expected to show a larger holistic effect and a larger configural effect than the letter-like group. The face-like group should have a higher tendency to process all the Ziggerin parts as well as the configurations between them. Therefore, part matching should be faster and more accurate when the distractor part led to the same response as the target part. And this congruency effect should be larger when the target and distractor parts were aligned as opposed to when they were misaligned. The letter-like group should not develop as strong a tendency to process all Ziggerin parts. Neither should the group develop sensitivity to configurations between the Ziggeirn parts. So the letter-like group should show less or no holistic or configural effect.

In the triplet recognition task, the letter-like group was expected to show their sensitivity to the regularity in style, i.e., they should recognize the Ziggerin triplets more accurately when the Ziggerins were in the same style as opposed to mixed styles. The face-like group was not expected to develop this sensitivity and thus style regularity should have no effect for them.

Face-like training

Participants improved in both accuracy and response time as training proceeded in all three tasks as detailed below (Figure 9, Appendices A to C).

In the naming task (Figure 9A), a Session × Ziggerin Type (with / without names) ANOVA showed a main effect of Session in RT [$F(9,153)=55.67$, $p≤.0001$]. No interaction was found. Accuracy was very high (>90%) throughout the whole training.

In the verification task (Figure 9B), an ANOVA showed a main effect of Session [RT: $F(9,153)=36.73$, $p≤.$; d': $F(9,153)=5.19$, $p≤.0001$]. The effect of Session was also significant in bias [$F(9,153)=4.45$, $p≤.0001$]. At the beginning, participants were biased to respond "same" and became less and less so as training proceeded.

In the matching task (Figure 9C), performance improved in terms of both speed and accuracy, A Session × Target Presence ANOVA revealed a significant main effect of Session [RT: $F(9,153)=29.41$, $p≤.0001$; Proportion correct: $F(9,153)=4.55$, $p≤.0001$]. No effect of and interaction with Target Presence was found.

Figure 9. Group performance during face-like training. In general responses became faster as training proceded. Error bars represent 95% confidence intervals for the Session effect.

Letter-like training

Participants improved steadily in all three tasks during training (Figure 10, Appendices D to F).

In the naming task, responses became faster as training proceeded, as shown by a significant Session effect in ANOVA on RT data [$F(9,153)=27.06$, $p\leq.0001$]. Accuracy was high (>95%) for all sessions.

In the verification task, participants also got faster with more sessions completed, as shown by a significant Session effect [$F(9,153)=18.83$, $p\leq.0001$]. Accuracy was high (d'>3.6) in all sessions. No change in bias was observed [$F(9,153)=1.19$, $p>.30$].

In the matrix scanning task, participants improved in both speed and accuracy, as indicated by the one-way ANOVA results [RT: $F(9,153)=92.72$, $p\leq.0001$; Proportion correct: $F(9,153)=4.57$, $p\leq.0001$].
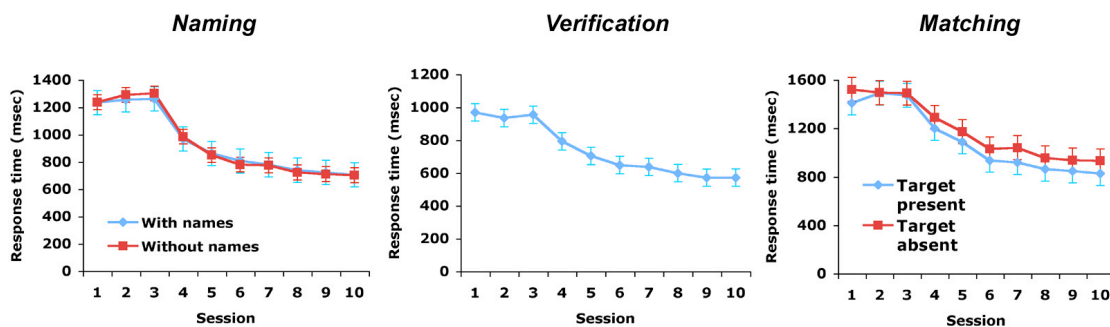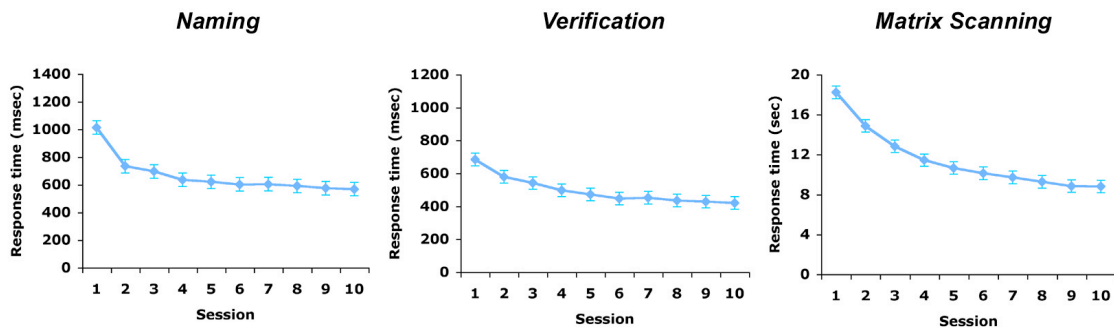

Figure 10. Group performance during letter-like training. In general responses became faster as training proceded. Error bars represent 95% confidence intervals for the Session effect.

Sequential Matching

The first two blocks (one class-level and one individual-level) were for practice and only trials after them were included in the analyses.
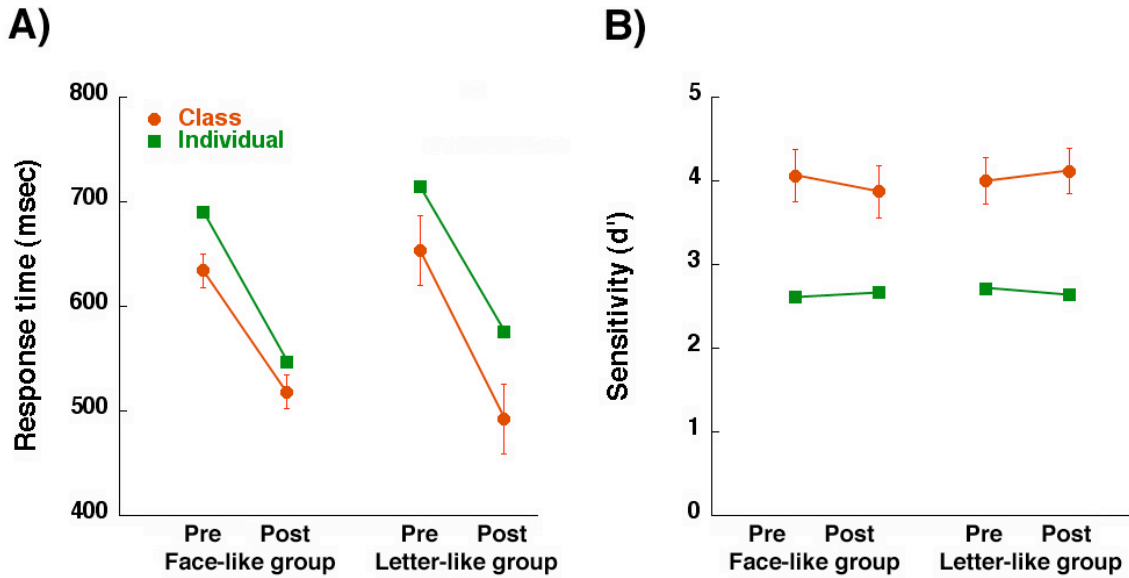
Figure 11. Response times (A) and sensitivity measures (B) for class and individual judgments. Both face- and letter-like groups responded faster in class than individual judgments before training. After training, the face-like group showed a reduction in the advantage of class over individual judgments, while the letter-like group showed the opposite trend. Sensitivity measures also showed a similar numerical pattern. Error bars represent the 95% confidence interval for the Pre/Post × Task interactions in each group.

In the sequential matching task, both groups performed better in class than individual judgment before training, but they showed opposite changes after training, as indicated by the correct response time and sensitivity data (Figure 11). For the face-like group, the performance difference between class and individual judgments shrank after training. For the letter-like group, however, this difference rose numerically with training.

For the response time data, a Group (face-like, letter-like) × Pre/Post × 2 Level

(class, individual) ANOVA showed a main effect of Pre/Post [$F(1,34)=94.13$, $p\leq.0001$] and a main effect of Level [$F(1,34)=41.43$, $p\leq.0001$]. Most importantly, there was a three-way interaction, confirming that the two groups displayed different changes in performance across judgments after training [$F(1,34)=4.00$, $p=.054$]. Separate ANOVAs revealed a Pre/Post × Level interaction for the face group, indicating a reduction of the class-level advantage over the individual level after training [$F(1,17)=6.34$, $p=.022$]. The letter-like group did not show a significant interaction, despite a numerical increase in the class-level advantage with training [$F(1,34)=1.02$, $p=.326$]. Scheffé tests ($p<.05$) showed that responses were faster at the class than individual levels for both groups during both pre- and post-training tests. Also, the letter-like group was faster at the class level than the face-like group after training but not before.

Sensitivity (d') measures showed a similar trend, although the Group × Pre/Post Level interaction did not reach significance [$F(1,34)=2.21$, $p=.146$]. Only the main effect of Level was significant [$F(1,34)=166.05$, $p\leq.0001$]. Note that sensitivity was already close to ceiling before training.

Part matching

Part matching was only performed after training. Data from two participants in the face-like group and four in the letter-like group were discarded because of low overall accuracy (<57%). Sixteen additional participants who had no experience with the Ziggerins were recruited to perform this task as control participants. Separate analyses were performed for trials in which distractors could change at an individual level and for trials in which distractors could change at a class level.

Distractor changes at the individual level. As seen in Figure 12, all groups showed a

congruency effect, but while the face-like group showed it in both response time and

sensitivity, the letter-like and control groups showed it only in sensitivity. In addition, the

congruency effect of the face-like group was sensitive to the alignment of Ziggerin parts,

with a larger congruency effect when the parts were aligned than misaligned.

Figure 12. Response times (A) and sensitivity measures (B) for part maching where
distractors changed at the individual level. Note that this task was performed only after
training. The face-like group showed a congruency effect only when the parts were
aligned. The letter-like group and control group did not have any congruency effect.
Sensitivity measures revealed a main effect of congruency. Error bars represent the
95% confidence intervals for the congruency effect in each group.

An Alignment (aligned or misaligned halves) × Congruency (congruent or incongruent halves) ANOVA on response time was conducted for each group. There was an Alignment × Congruency interaction for the face-like group [$F(1,15)=6.131$, $p=.026$], indicating the presence of a congruency effect only for aligned but not misaligned parts. The letter-like and control groups, however, showed a main effect of Alignment [Letter-like: $F(1,13)=5.711$, $p=.033$; Control: $F(1,15)=3.710$, $p=.073$] but no effect of congruency nor interaction between alignment and congruency [ps>.246 and .698 for the two groups respectively]. However, a Group × Alignment × Congruency ANOVA only revealed a moderately significant three-way interaction [$F(2,43)=2.478$, $p=.0958$]. Similar pattern of results were obtained when we split the congruent and incongruent conditions into different target-distractor matching trials (Appendix G).

Sensitivity data showed a Congruency effect in each of the three groups [Face-like: $F(1,15)=5.658$, $p=.031$; Letter-like: $F(1,13)=5.992$, $p=.029$; Control: $F(1,15)=18.845$, $p=.0006$]. The control group also showed a significant Alignment effect [$F(1,15)=4.213$, $p=.058$]. No Alignment × Congruency interaction was found in any group [ps>.22]. A Group × Alignment × Congruency ANOVA revealed no interaction effects with Group.

Distractor changes at the class level. As seen in Figure 13, there was no congruency effect for either response time or sensitivity for the face-like and letter-like group. The control group, however, showed a congruency effect in both aligned and misaligned trials (in terms of sensitivity and response time respectively).
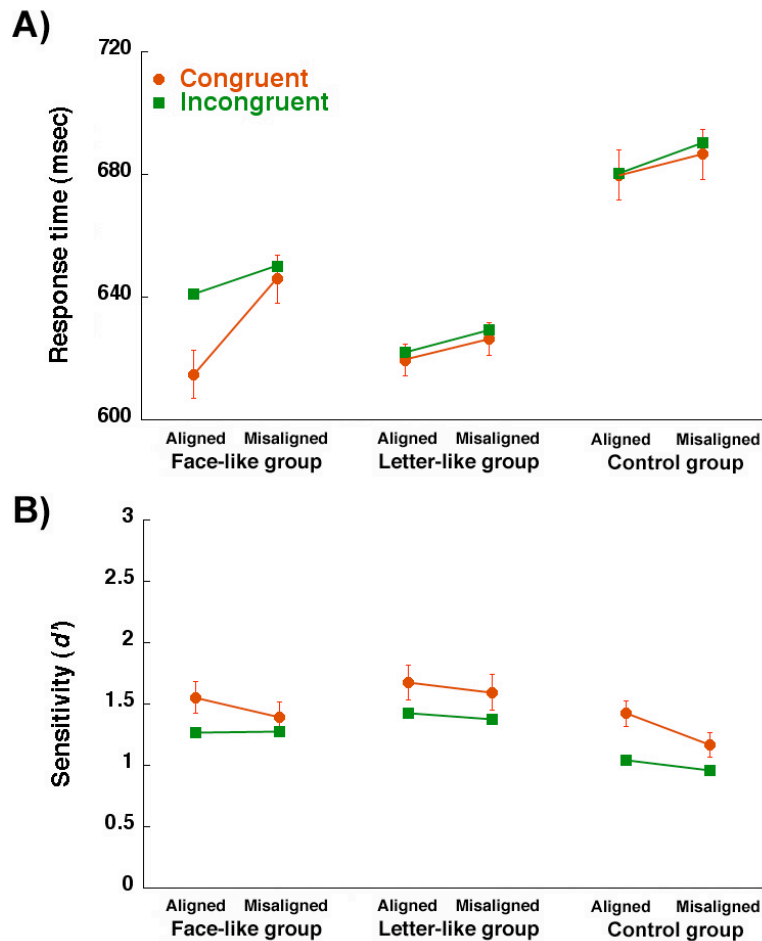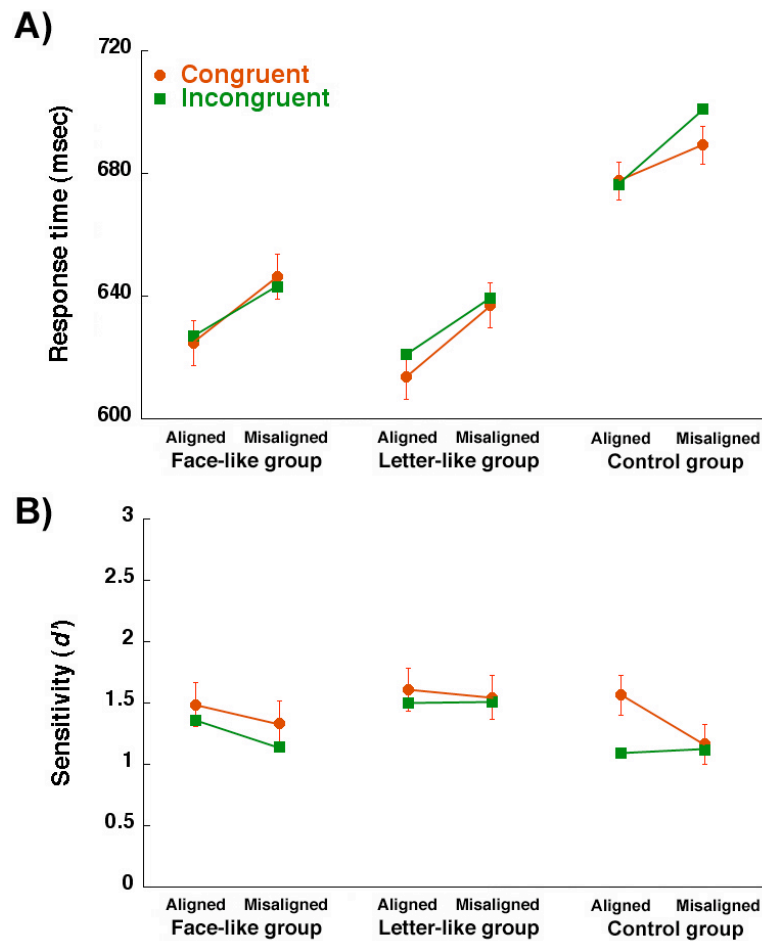
Figure 13. Response times (A) and sensitivity measures (B) for part matching where distractors changed at the class level. Responses were faster for aligned than misaligned trials for all groups. The control group showed a congruency effect for algined trials in terms of response time and a congruency effect for misaligned trials in terms of sensitivity. Error bars represent the 95% confidence intervals for the congruency effect in each group.

Alignment × Congruency ANOVAs on response times showed significant

Alignment effects for all groups [Face-like: $F(1,15)=19.357$, $p=.0005$; Letter-like:

$F(1,13)=17.258$, $p=.001$; Control: $F(1,15)=20.246$, $p=.0004$]. The control group also

showed a moderately significant Alignment × Congruency interaction [$F(1,15)=2.812$,

p=.114], with a larger congruency effect for misaligned trials. However, the sensitivity data for the control group showed the opposite results, with a larger congruency effect for aligned trials [Alignment × Congruency interaction: $F(1,15)$=4.742, p=.0458]. Therefore, it is hard to conclude whether the congruency effect was larger for aligned than misaligned trials. Sensitivity data showed no significant effects for the face-like and letter-like groups. Group × Alignment × Congruency ANOVAs revealed no interaction effects with Group.

Triplet recognition

The letter-like group recognized the triplets better than the face-like group, as shown by the shorter triplet duration required to attain a general performance level of around 2.25 Ziggerins correct per trial during the calibration blocks (Figure 14A). A one-way ANOVA confirmed this observation, showing a significant Group effect [$F(1,35)$=6.929, p=.013]. This was, however, the only difference between the two groups. During the experimental blocks, neither group showed any difference in performance between trials with a regular style and trials with mixed style (Figure 14B). A Group (face-like, letter-like) × Regularity (regular or mixed style) ANOVA showed no main effect or interaction.
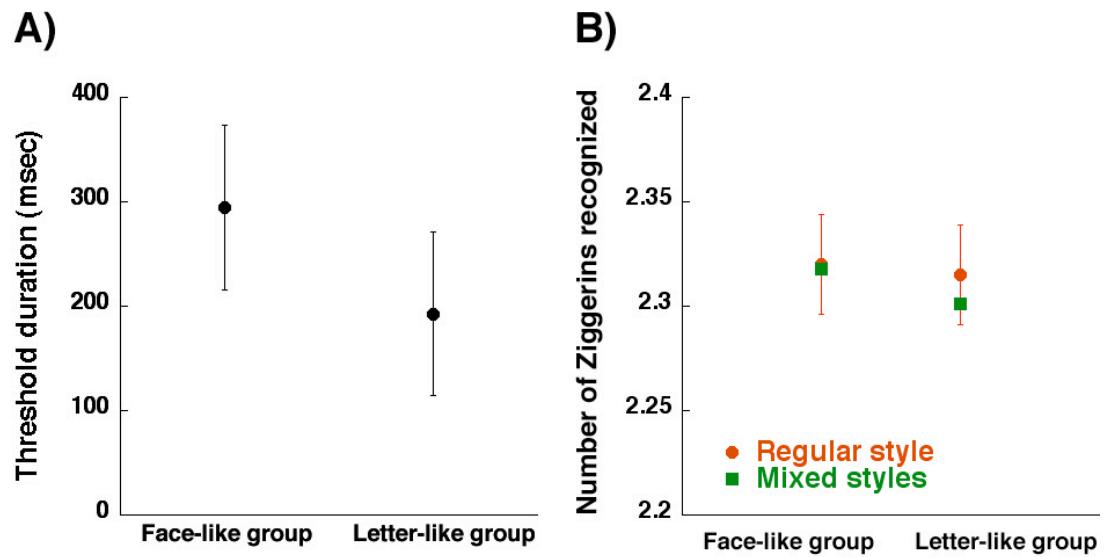
Figure 14. Performance in the triplet recognition task. (A) The threshold duration set during the calibration block. The letter-like group needed a shorter duration to recognize 2.25 objects than the face-like group. Error bars represeent the 95% confidence interval for the group difference. (B). The averaged number of Ziggerins recognized per trial for trials with regular vs. mixed styles. Neither group showed any difference between regular and mixed styles. Error bars represent the 95% confidence interal for style regularity differences.

General Discussion

These results confirm that it is possible to train people to become experts in two qualitatively different ways for the same set of objects.

Face-like training resulted in effects similar to face processing (Gauthier *et al.*, submitted; Tanaka, 2001; Tanaka & Farah, 1993). It also extended effects obtained before with Greeble training (Gauthier & Tarr, 1997, 2002; Gauthier *et al.*, 1998) to objects that are definitely not face-like in geometry, being asymmetric and inanimate looking. There was a reduced basic-level advantage, with more improvement in individuation than in class categorization. Note that the basic-level advantage still existed

after training, i.e., performance was still better at the class than individual level, contrary to the equivalent performance at both levels after Greeble training. This is likely because participants had to learn individuation for a heterogeneous Ziggerin set. Due to the need to accommodate letter-like training with the same set of objects, the training set contained six distinct classes. Still it is impressive that individuation training can significantly reduce the basic-level advantage with 10 hours of training in total for six distinct basic levels.

Another effect of the face-like training was an increase in holistic and configural processing. In the part matching task in which one had to perform matching on half of an object, the face-like group found it harder to selectively attend to the target part without being affected by the other, distractor half. This holistic effect, the tendency to consider all the parts of an object despite selective attention instructions, was larger when the target and distractor parts were aligned than misaligned, indicating a development of sensitivity to configuration of Ziggerin parts. This confirms the idea that experience in individuating a set of objects, be it faces, cars, birds, or Greebles, causes the objects to be processed in a more holistic and configural manner. Importantly, the letter-like training group had a smaller holistic effect with a magnitude similar to that of the control group, and did not show any configural effect at all. Therefore, it is not the exposure to objects with the geometry of the Ziggerins that led to the face-like training effects. Instead, the demand for individuation may be the crucial factor for the similar training effects across Greeble and Ziggerin training studies.

Letter-like training showed a trend for an increased basic-level advantage, opposite to face-like training effects, an effect characteristic of letter expertise (Wong &

Gauthier, in press). Letter-like training also led to a speed advantage, compared to face-like training, in recognizing triplets of Ziggerins, similar to what has been found for characters in a familiar writing system (Gauthier *et al.*, 2006). An intuitive explanation is that the effect came from domain-specific changes due to basic-level categorization training with the Ziggerins. Alternatively, because this task was not performed with control objects from an untrained category, the training effect could be solely due to the transfer of domain-general strategies from the matrix scanning task during training to the triplet recognition task. However, this appears unlikely, as the letter-like group performed class judgment on individual Ziggerins in the sequential matching task faster and more accurately than the face-like group after training (Figure 11).

Another effect expected after letter-like training was style tuning. During training participants scanned through over 1000 matrices each with 40 Ziggerins presented simultaneously in the same style. With this experience, participants were expected to develop some kind of sensitivity to the regularity of styles, and make use of that to facilitate performance in the triplet recognition task when the three Ziggerins were in the same styles compared with different styles. This prediction was drawn based on the font tuning effects shown before with characters in a familiar writing system (Gauthier *et al.*, 2006). However, no style tuning effect was found after letter-like training. There are at least two probable reasons. First, the triplet recognition task may not be sensitive enough. Since there were only three Ziggerins in each trial, the cost of disrupting style regularity may not be sizeable enough to be observable, especially given that on average a participant took just 200 ms to recognize over 2.3 Ziggerins. Second, the styles applied to the Ziggerins may not be the 'right' kind of styles that the visual system would tune to.

Note that in Gauthier *et al.* (2006), experts did not just show tuning effects for any kind of font changes. Tuning was found for changes in the aspect ratio of character parts (e.g., the size of the loop relative to the pole in 'b'), but not for visually obvious changes in fill and slant. It is unclear why different types of font changes can lead to different degrees of tuning effects.

The results of the letter-like training are encouraging, as letter-like training has never been implemented before. It was unclear how much training would be needed to produce sizeable effects. Besides, since the same set of objects had to be used for face-like training also, the number of classes was restricted to a small number (six in this case). In this sense the letter-like training introduced here was less demanding than the basic-level categorization demand with letters as there are far more than six characters in natural writing systems.

Overall, the current findings demonstrate that, with careful control of object geometry and processing beyond object recognition, one can produce differential behavioral consequences by introducing different training experiences.

CHAPTER III

NEURAL CONSEQUENCES OF OBJECT RECOGNITION EXPERIENCES
MIRRORING FACE AND LETTER RECOGNITION

Introduction

It has been shown that Ziggerin training modeling some aspects of experience with faces and letters resulted in behavioral markers for face-like and letter-like expertise respectively. This chapter focuses on the changes in neural activity as a result of the two types of training. The main questions are: would the two types of training result in different neural changes in the occipitotemporal regions, and if so, would the activity patterns correspond to neural signatures for face and letter selectivity?

Half of the training participants who have participated in the behavioral testing were also scanned using fMRI both before and after training. To know whether face-like and letter-like training would lead to enhanced fMRI activity in the face- and letter-selective regions respectively, the fusiform face area (FFA), letter area (LA) and the parahippocampal place area (PPA) were localized and their responses to the Ziggerins were compared before and after training.

Apart from probing the training effects in the specific regions of interest, an increasingly common practice is to look at a wider range of regions involved in object recognition (e.g., Op de Beeck *et al.*, 2006). One specific question addressed in this study concerns the training effects in regions with different eccentricity biases. It has been suggested that face expertise imposes a higher individuation demand than letter expertise (Wong & Gauthier, in press). It has also been suggested that a higher demand for

resolution tends to recruit more lateral regions of the ventral occipitotemporal cortex that shows a bias for representation at the center of the visual field (Hasson *et al.*, 2002; Levy *et al.*, 2001; Malach *et al.*, 2002). Accordingly, face-like training should recruit more lateral regions with a larger central presentation bias than letter-like training. To understand the relationship between the two types of training effects and the regions with different eccentricity biases, separate scans were also introduced to localize regions preferring presentation at the center, middle, or periphery of the visual field.

Methods

Participants

Eighteen of the 36 participants who took part in the behavioral training also participated in two fMRI sessions, one before and one after training. Nine were from the face-like group (six females, seven right-handed, age M = 22.11, SD = 1.32) and nine from the letter-like group (five females, six right-handed, age M = 21.22, SD = 1.22). They each received $12 for each behavioral session and $25 for each fMRI session. All
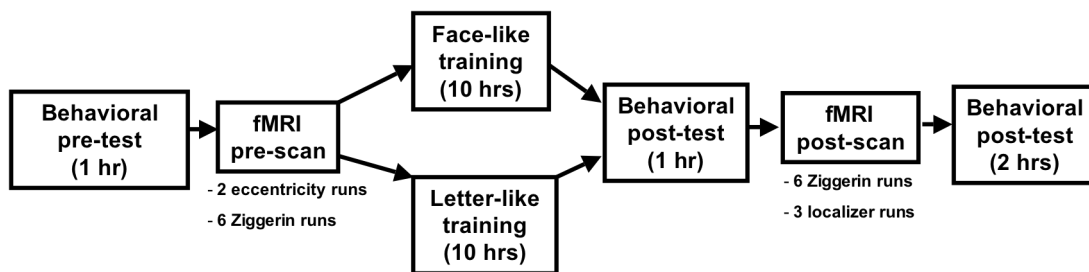
Figure 15. Sequence of training and testing sessions. During the pre-training scan, participants first went through two eccentricity runs and then six Ziggerin runs. During the post-training scan, they went through the same Ziggerin runs and then three localizer runs.

66

participants had normal or corrected to normal vision, and reported no history of neurological disorders. The sequence of training and testing sessions for each participant is shown in Figure 15.

Stimuli and Material

Anatomical 2D and 3D high-resolution T1-weighted images were acquired with conventional parameters on a 3T Philips Intera Achieva scanner at the Vanderbilt University Institute of Imaging Science. All testing was conducted on Mac computers using Matlab (MathWorks, Natick MA) with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). The stimuli were presented on an LCD panel and back-projected on a screen. Participants viewed the stimuli through a mirror mounted on top of an RF coil above their head.

The eccentricity runs used 80 grayscale pictures of familiar objects (Figure 16A). Each type of stimuli was presented either at a central circle (1.4° in diameter), in a middle ring (2.5° inner diameter and 5° outer diameter), or in a peripheral ring (10° inner diameter and 20° outer diameter). The Ziggerin runs used 72 Ziggerin (36 in trained set and 36 in transfer set) and 36 familiar object images (in six classes: beds, boats, cars, chairs, lamps, and teapots) (Figure 16B). The localizer runs used 144 grayscale images including 36 faces (half female), 36 familiar objects (those not used in the eccentricity and Ziggerin runs), 36 Roman letters (a, b, d, e, f, g, h, k, m, n, p, q, r, s, t, u, w, and y in two fonts), and 36 pseudoletters (formed by taking each Roman letters and rearranging the strokes) (Figure 16C). In both localizer and Ziggerin runs, each image spanned a visual angle of 4°.

(A)

(B)

(C)
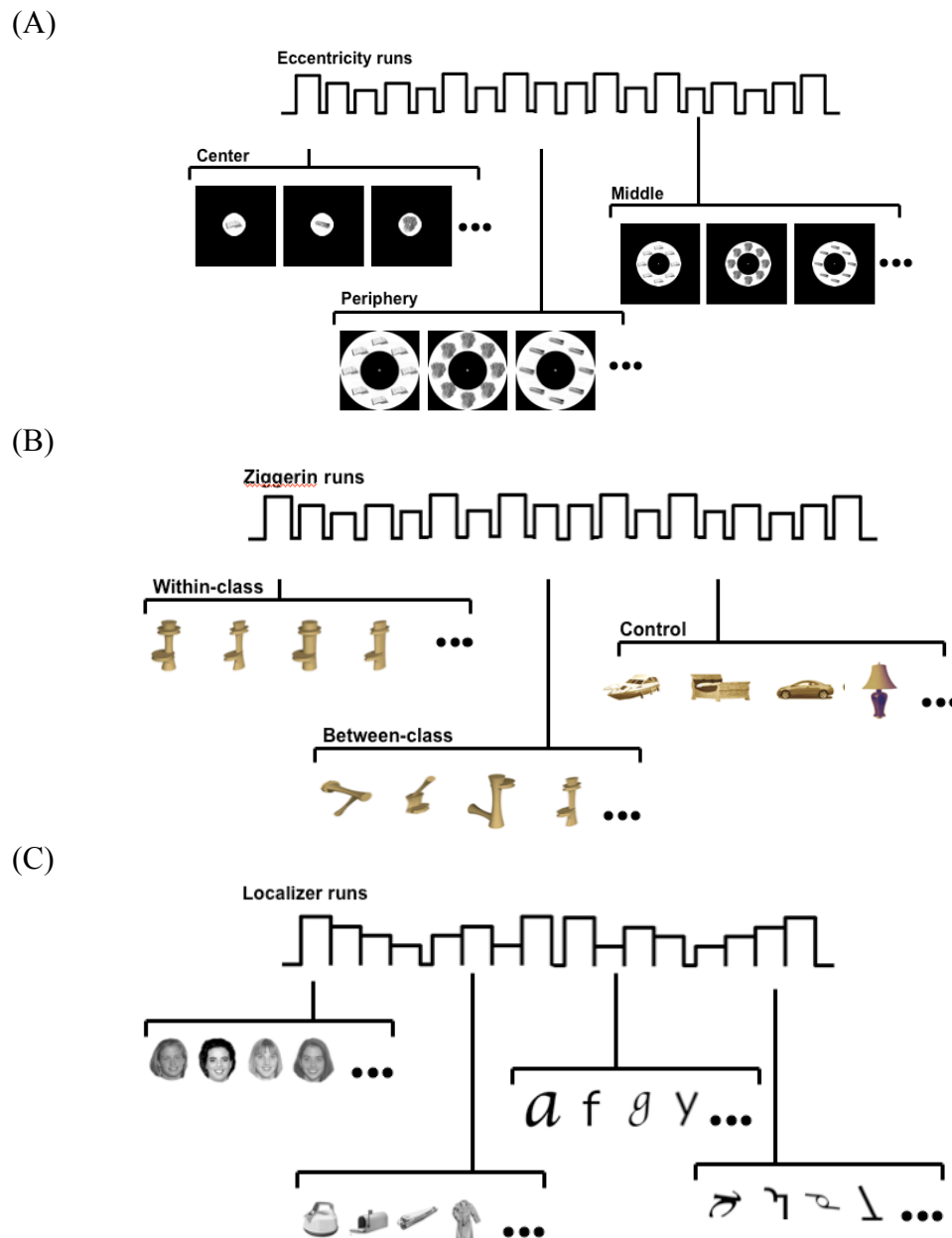
Figure 16. The conditions in the eccentricity (A), Ziggerin (B), and localizer (C) runs. Note that the stimuli in the center and middle conditions in (A) are enlarged for visualization.

Procedure

The first fMRI session was performed after the pre-training sequential matching task, and the second fMRI session was performed after the post-training sequential matching. A block design was used throughout the experiment, with different types of stimuli presented in separate blocks. In all runs, participants were required to perform a one-back image repetition task. They had to press a button with their right index finger on a response box attached to their hand when they saw two identical images presented consecutively. No response was required on a non-match trial. The ratio of match to non-match trials was 1:11. Each trial began with a blank for 275 ms followed by the stimulus for 725 ms.

In each eccentricity run, each of the three conditions (center, middle, periphery) appeared in six blocks of 10 sec, separated by 6 sec of fixation. Twelve seconds of fixation were also inserted at the beginning and the end of each run.

In each Ziggerin run, there were three conditions (Ziggerin-within, Ziggerin-between, object control). In the Ziggerin-within condition, each block contained only Ziggerins within a class, while in the Ziggerin-between condition, each block contained Ziggerins across different classes. So the Ziggerin within condition required within-class discrimination whereas the Ziggerin-between condition required between-class discrimination. Between-class discrimination was also required in the object control condition, as each block contained familiar objects from different categories. Each condition appeared in six blocks of 12 sec in each run, separated by 6 sec of fixation. Twelve seconds of fixation were inserted at the beginning and the end of each run. Three runs were devoted to trained Ziggerins and three runs were devoted to transfer Ziggerins,

and the run order was counterbalanced across participants.

In each localizer run, each of the four conditions (faces, objects, letters, pseudoletters) appeared in four blocks of 16 seconds. Eight-sec fixations were inserted at the beginning and end of each run, and after each cycle of the four conditions. In all runs, the presentation of stimuli was randomized within blocks, and the presentation of blocks was counterbalanced across runs.

Imaging parameters and analyses

Imaging was performed using a 3-Tesla, whole body GE MRI system and a birdcage head coil located at the Vanderbilt Medical Center (Nashville, USA). The field of view was 19.2 × 19.2 × 10.2 cm, with an in-plane resolution of 64 × 64 pixels and 34 contiguous axial scan planes per volume (whole brain), resulting in a voxel size of 3 × 3 × 3 mm. Images were collected using a T2*-weighted EPI acquisition (TE=25 ms, TR=2000 ms, flip angle=60) for blood oxygen-level dependent (BOLD) based imaging. High-resolution T1-weighted anatomical volumes were also acquired using a 3-D fast spoiled grass (FSPGR) acquisition (TI=400 ms, TE=4.18 ms, TR=10 ms, FA=20). The functional data were analyzed using the Brain Voyager™ (http://www.brainvoyager.com) multi-study GLM (general linear model) procedure and in-house programs written in Matlab™ (http://www.themathworks.com). Data preprocessing included slice scan time correction, 3D motion correction, temporal filtering (3 cycles/run high-pass), and spatial smoothing (6 mm FWHM Gaussian). A GLM analysis computed the correlation of predictor variables or functions with the recorded activation data (criterion variables) across scanning sessions. The predictor functions were based on the blocked stimulus

presentation paradigm of the particular run being analyzed and represented an estimate of the predicted hemodynamic response during that run. To properly model the hemodynamic response, the predictors were represented as the stimulus protocol boxcar functions convolved with the appropriate gamma function (Δ=2.5, τ=1.25) estimate of a typical hemodynamic response (Boynton *et al.*, 1996). To increase power, statistical parametric maps were computed only for gray matter within a pre-set search region (Talairach coordinates: x= -70 ~ 70, y= -20 ~ (-100), z= -2 ~ -32), in order to focus on the occipital and posterior temporal regions in both hemispheres. Regions of interest (e.g., face-selective regions) were identified at both group and individual levels. At the group level, activations for different conditions in the localizer runs were compared using random effects analyses at the group level with a threshold of p<.001 (uncorrected) and a cluster threshold of three contiguous 3-mm isometric voxels. At the individual level the same conditions were compared using fixed effects analyses with a threshold of p(FDR)<.05 and a cluster threshold of three contiguous 3-mm isometric voxels. FDR (False Discovery Rate) is a method of correcting for multiple comparisons by controlling for the expected proportion of false positive voxels among those that are above threshold (Genovese *et al.*, 2002). For analyses of the Ziggerin runs, since trained and transfer runs showed similar results in different analyses they were collapsed in the subsequent reports of results.

Predictions

Figure 17 shows the major predictions of the pre- and post-training scans. The fusiform face area (FFA) and the left letter-selective area (LLA) were expected to show

different training effects for the two groups. Whereas face-like training was expected to result in an increase in fMRI activations in the FFA but not in the LLA, letter-like training was expected to enhance activity in the LLA but not in the FFA. Since the face-like group was trained to discriminate Ziggerins within a class, one exploratory question was if the training effect for this group would be stronger for the Ziggerin-within condition, where participants had to discriminate between Ziggerins in the same class, as opposed to the Ziggerin-between condition, where participants were required to discriminate between different classes of Ziggerins. Similarly, for the letter-like group, the training effect may be clearer for the Ziggerin-between condition than the Ziggerin-within condition, since the group was trained on discrimination between classes of Ziggerins. The parahippocampal place area (PPA) was a control region and was expected to show no change in activity after training.

Exploratory analyses were also conducted to study if there was any relationship between different eccentricity biases in different regions of the ventral occipitotemporal
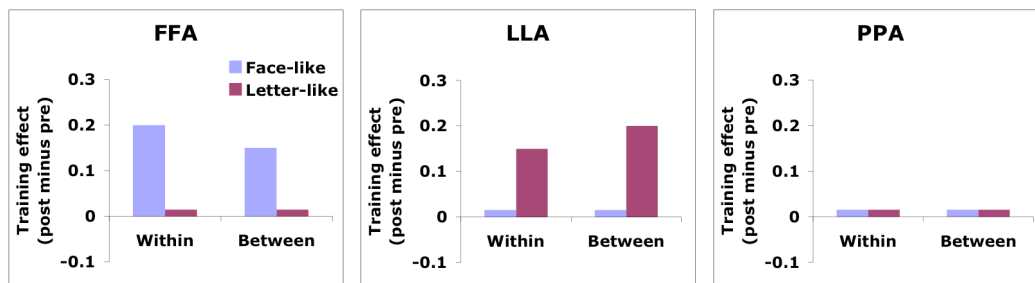
Figure 17. Predictions for the imaging results of the pre- and post-training scans. The face-like group was expected to show an increase in the activity of the fusiform face area (FFA) after training. The letter-like group, however, was expected to show an activity increase in the left letter-selective area (LLA). As a control region, the parahippocampal place area (PPA) was expected to show little training effect for either group.

cortex, the preference for basic- vs. subordinate-level categorization in these regions, and the differential training effects for the face-like and letter-like groups.

Results

Behavioral results

Eccentricity runs

Performance was similar in different conditions (Figure 18A), as indicated by the lack of main effects of Group and Eccentricity (center, middle, periphery), or their interactions (all ps>.18 for both RT and proportion correct measures).

Localizer runs

Responses were slower for faces than for other stimuli (Figure 18B). A Group × Stimulus ANOVA showed a significant Stimulus effect [$F(3,48)=5.48$, $p=.0025$]. Scheffé tests ($p<.05$) revealed longer response times for faces than for each of the other stimuli, while there was no difference among objects, letters, and pseudoletters. Accuracy measures showed a significant Group × Stimulus interaction [$F(3,48)=4.90$, $p=.0047$]. Scheffé tests showed that only the difference between faces and pseudoletters in the letter-like group was significant ($p=.051$).

Ziggerin runs

Performances improved with training (Figure 18C). This was confirmed by a Group × Pre/Post × Run Type (trained, transfer) × Block Type (within, between, control) ANOVA, which showed a main effect of Pre/Post [RT: $F(1,16)=9.57$, $p=.0070$;

Proportion correct: $F(1,16)=4.12$, $p=.0595$]. No main effect of or interaction with Group was found.
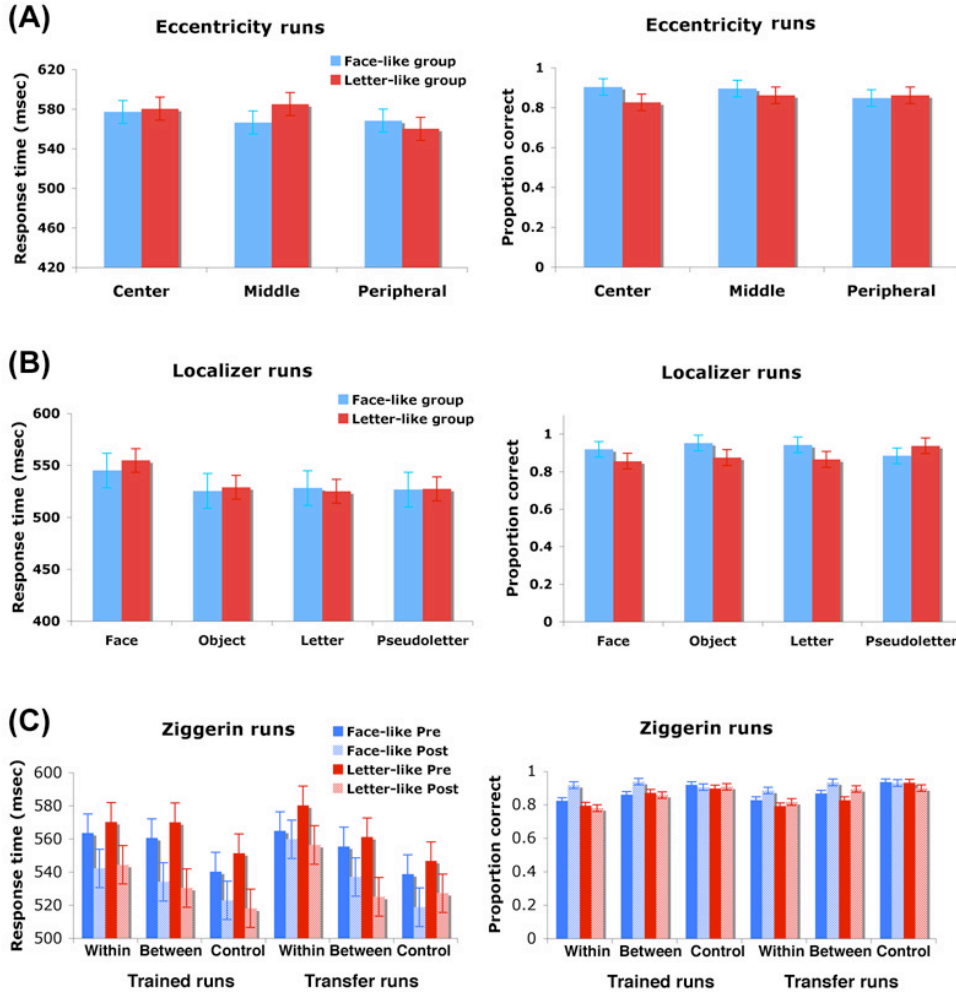

Figure 18. Behavioral performance in the (A) eccentricity, (B) localizer, and (C) Ziggerin runs. Error bars represent the 95% confidence interval for the eccentricity effect within each group in (A), the 95% confidence interval for the sitmulus effect in (B), and the 95% confidence for the pre/post effect in (C).

Other effects included a significant Block Type effect [RT: $F(2,32)=20.13$, p.0001; Proportion correct: $F(2,32)=21.96$, p.0001], but it interacted with Pre/Post [Proportion correct: $F(2,32)=3.23$, $p=.0524$] and Run Type [RT: $F(2,32)=3.62$, $p=.0383$]. Scheffé tests ($p<.05$) showed that, while accuracy followed the order of object control > Ziggerin-between > Ziggerin-within before training, performance improved such that Ziggerin-between became as accurate as the object control blocks. For the Block Type × Run Type interaction, responses were faster for trained than transfer runs only for Ziggerin-within blocks. There was also an interaction between Pre/Post and Run Type [RT: $F(1,16)=3.94$, $p=.0647$]. Responses were faster for trained than transfer runs after training but similar before training.

Imaging results

Four types of analyses were conducted. First, statistical parameter maps were generated to search for any voxels in the ventral visual stream showing significant training effects. Second, the face- and letter-selective regions were identified at the group level by comparing activity across conditions during the localizer runs. The training effects for the two groups in these regions of interest were compared. Third, the same analyses were conducted with the regions of interest identified at the individual level. Finally, the pattern of training effects over the whole ventral occipitotemporal cortex was examined by dividing the cortex into 32 regions of interest.

Statistical parameter maps

First, a search for ventral visual areas showing training effects was performed by

conducting a number of contrasts between activations for Ziggerins before and after

training. Specifically, the following two contrasts were conducted for each group: (1) the

substraction of the activity in the Ziggerin-within condition (compared with object

control) during pre-scan from that during post-scan; and (2) the substraction of the

activity in the Ziggerin-between condition (compared with object control) during pre-

scan from that during post-scan. At a threshold of $p<.001$ (uncorrected), no significant

voxel was revealed with these contrasts in either group. Analyses of the interaction

between Group and each of the above two contrasts were also performed. Again no

significant voxel was revealed.

The only significant training effect in an SPM was that for the Ziggerin-between

condition when the two groups were collapsed, i.e., when the second contrast mentioned

above was conducted averaging the two groups together. As shown in Figure 19, a region

in the left posterior fusiform gyrus (pFs) was found at $p<.001$ (uncorrected). This region

(size=88mm^3) was not part of the fusiform face area, given the higher activity for objects

than faces. The region showed a reduction in activations for both groups, to a similar

degree in the Ziggerin-within condition and more so for the letter-like group in the

Ziggerin-between condition. A Group (face-like, letter-like) × Block Type (Ziggerin-

within, Ziggerin-between) ANOVA confirmed the observation. There was a close-to-

significant Group × Block Type interaction [$F(1,16)=4.045$, $p=.0615$]. Scheffé tests

($p<.05$) showed that the training effects were similar between the groups for the Ziggerin-

within condition but larger for the letter-like group for the Ziggerin-between condition.

One-sample t-tests showed that the training effect was significantly different from zero

for the face-like group in the Ziggerin-within condition [$t(8)=-2.963$, $p=.002$] and was

close to significant in the Ziggerin-between condition [$t(8)$=-2.023, p=.076]. For the

letter-like group the training effect was significant in the Ziggerin-between condition

[$t(8)$=-4.713, p=.002] and did not reach significance for the Ziggerin-within condition

[$t(8)$=-1.702, p=.127].
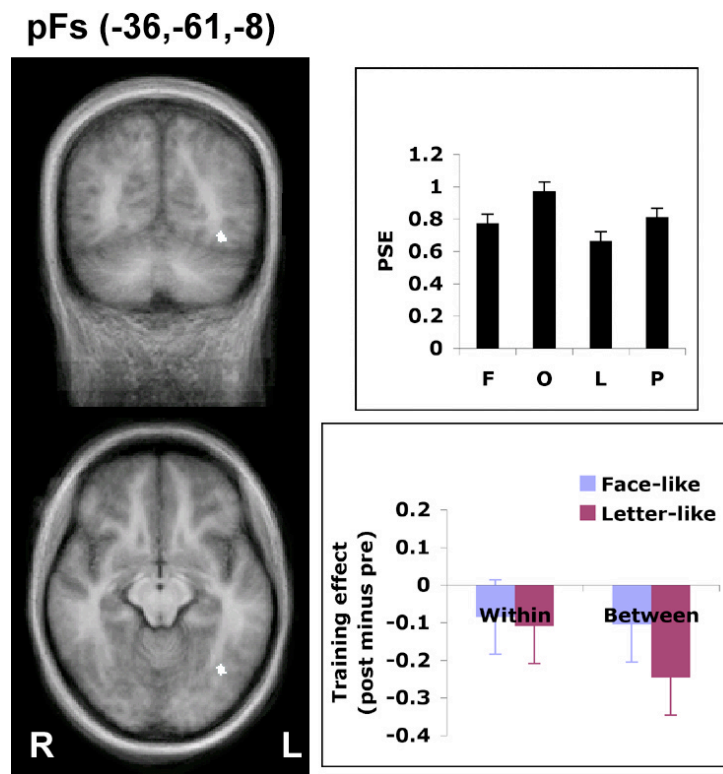
Figure 19. Training effects in the face/letter localizer runs and Ziggerin runs in a posterior fusiform (pFs) region. This region (indicated in white color) was found by comparing the activity for the Ziggerin-between condition before and after training with the two groups collapsed. The region showed a reduction in activity after training for both groups, but more so for the letter-like group in the Ziggerin-between condition. Error bars represent the 95% confidence interval for the stimulus variable in the localizer plot, and for the group variable in the Ziggerin plot.

Face- and letter-selective regions (group defined)

The main hypothesis of this study was the increased activity in face-selective and letter-selective regions after face-like and letter-like training respectively. To test this hypothesis, the training effects of these regions of interest were measured. The right fusiform face area (RFFA) was localized at the group level using the (faces minus letters & objects) contrast. The left letter area (LLA) was localized using the (letters minus faces
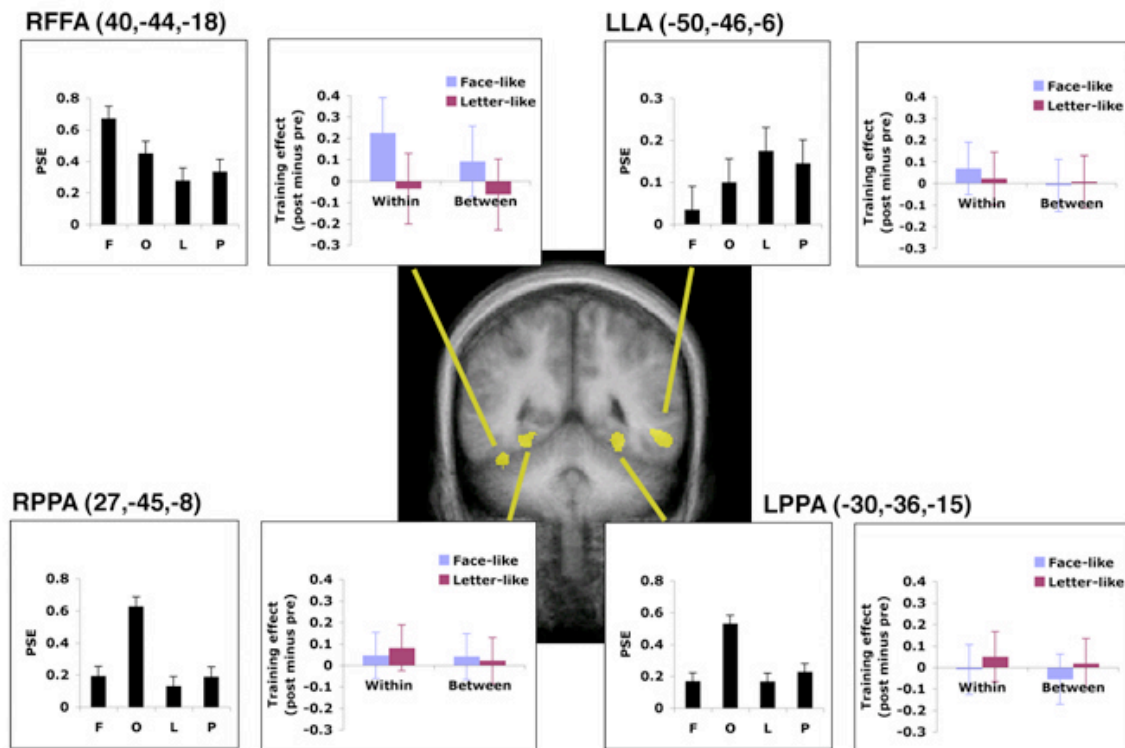
Figure 20. Training effects in the face/letter localizer runs and Ziggerin runs in different ROIs. For each ROI, the left graph shows the percent signal change during the localizer (F- faces; O – objects; L – letters; P- pseudoletters). The right graph for each ROI shows the changes (post minus pre) in BOLD signal for Ziggerins in the two groups. Error bars respresent the 95% confidence intervals for the stimulus variable in the localizer graphs and for the group variable in the Ziggerin graphs.

& objects) contrast. No left FFA or right LA was found even with a low threshold ($p<.01$ uncorrected). As control regions, the left and right parahippocampal place areas (LPPA and RPPA) were also localized by the (objects minus faces) contrast. PPA has been shown to respond more to objects than faces (e.g., Epstein *et al.*, 1999). Figure 20 shows the activity of these ROIs in the localizer runs and the training effects in the Ziggerin runs.

The training effects for each region were measured by subtracting the activity for the Ziggerins (relative to control objects) before training from that after training. There was a larger training effect for the face-like group than the letter-like group in the RFFA. A Group (face-like, letter-like) × Block Type (Ziggerin-within, Ziggerin-between) ANOVA conducted on the training effects for each ROI confirmed this observation. The effect of Group was close to significant for the RFFA [$F(1,16)=3.56$, $p=.0773$]. One sample *t*-tests showed that the RFFA training effect was close to significant for the Ziggerin-within blocks in the face-like group [$t(8)=1.866$, $p=.099$] but not for the Ziggerin-between block or for either block in the letter-like group [ps>.30]. No other effects were significant in the other ROIs.

Pearson product-moment correlations were computed between the training effects in these ROIs and behavioral signatures of learning across participants. The behavioral signatures include the difference in the basic-level advantage during the sequential matching task before and after training, and the configural effect during the part matching task. The only behavioral measure correlating with neural training effects was the configural effect obtained in the part matching task, in which one performed matching on a Ziggerin half while ignoring the other half. The configural effect was defined as the

difference in response time between the congruent and incongruent conditions when the parts were aligned compared with the difference when they were misaligned. In other words, it measured the sensitivity of the congruency effect, or holistic processing, to a change in configuration. Figure 21 shows the correlations in different regions, separately for Ziggerin-within and Ziggerin-between conditions. The amount of configural effect was highly correlated with the amount of training effect in the RFFA, for both Ziggerin-within [r=.73, p=.0006] and Ziggerin-between conditions [r=.70, p=.0013]. As shown in the scatter plots participants in the face-like group tended to have a larger configural
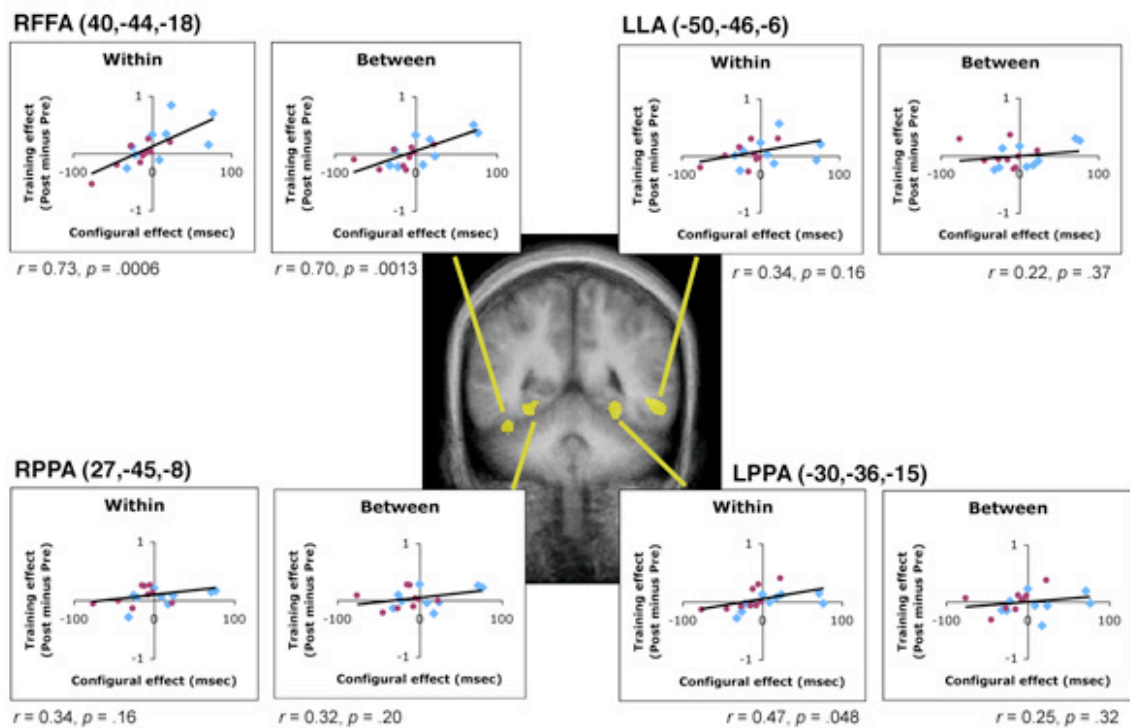
Figure 21. Correlation between the configural effect found in the part matching task and the training effects in different ROIs. The face-like (in light blue diamonds) and letter-like (in plum dots) groups were analysed together. For each ROI, the left and right graphs show the correlation for the Ziggerin-within and Ziggerin-between conditions respectively. Only the RFFA shows a significant correlation between its training effect and the behavioral configural effect in both within and between conditions.

effect as well as a larger RFFA training effect than the participants in the letter-like group. There were no significant correlations in other ROIs, except for LPPA that showed a correlation between the configural effect and the training effect in the Ziggerin-within condition [r=.47, p=.048]. As shown in the scatter plot the face-like group highly overlapped with the letter-like group in the LPPA, suggesting that this moderate correlation was not likely due to training but rather individual differences. Also, the training effect in the LPPA did not correlate with the configural effect in the Ziggerin-between condition [r=.25, p=.32].

Therefore, both the magnitude of the training effect and its correlation with configural processing indicates that face-like training resulted in more RFFA recruitment.

Face- and letter-selective regions (Individually defined)

It was expected that the training effects would be clearer with a more precise identification of the regions of interest, for instance, at an individual level. The analyses in the previous section were thus conducted on the RFFA, LPPA and RPPA identified in each participant. The LLA could not be identified for one-third of the participants even when a lower threshold was used (FDR <.10) and thus data for the region were not analyzed. Table 2 shows the Talairach coordinates of these regions for each participant.

Figure 22 shows the training effects in these individually defined regions. Contrary to the group-defined ROI analyses, there was little training effect in each of these regions. A Group (face-like, letter-like) × Block Type (Ziggerin-within, Ziggerin-between) ANOVA conducted for each region showed no significant effect in any of these regions [all ps>.16]. Planned comparisons showed that none of the training effects were

significantly different from zero [all ps>.15]. Correlation analyses showed findings

consistent with the group-defined ROI analyses, with a significant correlation found

between the configural effect (in response time) and the magnitude of the training effect

in the individually defined RFFA in the Ziggerin-between condition [r=.462, p<.053].

None of the other correlations were significant (Table 3).

Table 2. The Talairach coordinates and sizes (in mm^3) of the right fusiform face area
(RFFA), and the left and right parahippocampal place area (LPPA and RPPA) defined for
individual participants.

Face-like Subject	RFFA				LPPA				RPPA			
	x	y	z	size	x	y	z	size	x	y	z	size
1	42	-47	-14	865	-30	-50	-6	917	24	-45	-11	790
2	44	-38	-12	608	-29	-49	-11	993	29	-45	-4	855
3	36	-42	-20	523	-27	-43	-14	930	26	-48	-10	766
4	34	-44	-14	603	-26	-55	-10	899	24	-55	-8	928
5	37	-53	-16	542	-32	-56	-11	910	26	-50	-5	870
6	38	-40	-14	472	-27	-40	-12	840	28	-40	-11	960
7	46	-50	-8	823	-27	-43	-8	860	33	-47	-18	865
8	39	-42	-14	598	-28	-47	-9	931	23	-44	-9	836
9	34	-39	-17	668	-32	-39	-15	883	24	-34	-11	988
Average	38.9	-43.4	-14.6	634	-28.7	-46.9	-10.7	907	26.3	-45.3	-9.67	873
Letter-like Subject	RFFA				LPPA				RPPA			
	x	y	z	size	x	y	z	size	x	y	z	size
1	41	-50	-11	232	-26	-50	-10	777	30	-48	-10	916
2	34	-56	-14	599	-30	-41	-10	705	29	-42	-13	884
3	35	-39	-18	385	-27	-50	-11	516	23	-46	-8	527
4	32	-50	-15	407	-23	-51	-13	421	24	-49	-10	464
5	34	-47	-17	844	-33	-50	-11	959	27	-41	-9	878
6	43	-49	-17	232	-28	-47	-16	670	23	-53	-9	667
7	46	-49	-11	462	-30	-51	-13	973	34	-45	-13	931
8	40	-46	-14	902	-28	-42	-12	882	27	-41	-13	933
9	42	-48	-17	695	-28	-50	-10	964	26	-56	-8	910
Average	38.6	-48.2	-14.9	529	-28.1	-48	-11.8	763	27	-46.8	-10.3	790
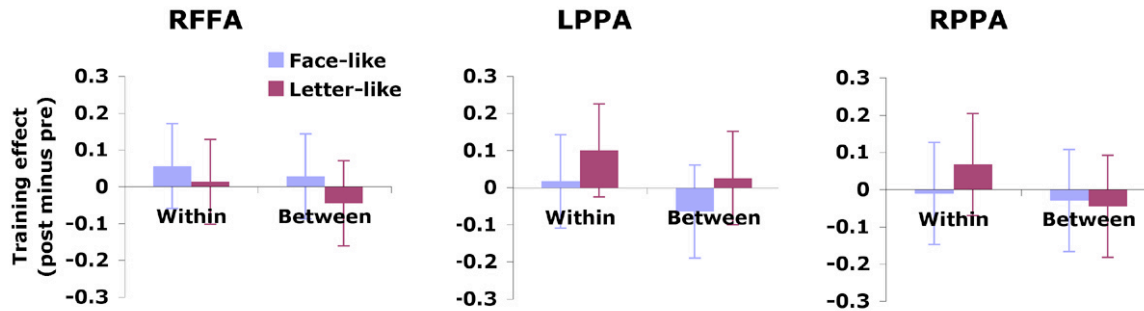
Figure 22. Training effects in the individually defined ROIs. TheThe left letter area could not be defined in the majority of participants. Error bars represent the 95% confidence intervals for the group variable.

Table 3. The coefficients (*r*) and significance (*p*) of the correlations between the configural effect in the part matching task and the training effects in the individually defined RFFA.

	Ziggerin-within		Ziggerin-between	
	r	*p*	*r*	*p*
RFFA	0.383	0.116	0.462	0.053
LPPA	0.212	0.398	0.163	0.519
RPPA	0.071	0.779	0.192	0.445

Ventral occipitotemporal cortex

To better understand the training effects over the object recognition system in general, a set of 32 ROI boxes were used to target on different parts along the ventral occipitotemporal cortex (Figure 23A). These were 10×10×15-mm ROIs situated at different points along the anterior-posterior axis (*y*-coordinate =-35, -45, -55, or -65) and the medial-lateral axis (*x*-coordinate = -20, -30, -40, or -50 on the left and 20, 30, 40, or 50 on the right). The location and extent of each ROI along the *z*-axis were adjusted for

each participant to ensure the best coverage of gray matter.

One organizing principle proposed to characterize the occipitotemporal cortex is eccentricity bias (Hasson *et al.*, 2002; Levy *et al.*, 2001; Malach *et al.*, 2002). It has been shown that the more medial parts of the ventral occipitotemporal cortex tend to prefer peripheral presentations whereas the lateral parts are more responsive to central presentations. It was thus suggested that the lateral regions are responsible for recognition tasks associated with a higher resolution demand than the medial regions that are more concerned with information integration rather than fine discrimination (Hasson *et al.*, 2002; Malach *et al.*, 2002). However, this interesting claim has never been directly tested before. The current training paradigm provides an ideal condition to test this hypothesis, as discrimination demand was varied for a set of novel objects across the Ziggerin-within and Ziggerin-between blocks.

To test the claim, the activations in the eccentricity runs and the Ziggerin runs before training were examined. Specifically, the activation pattern of the 32 ROIs for central and periphery presentations were compared with the activation pattern for within- and between-class discriminations.

As shown in Figure 23A, there was a nice correspondence between the eccentricity and resolution demand patterns. Lateral regions preferred presentation at the center as well as within-class discrimination, whereas medial regions preferred presentation at the periphery as well as between-class discrimination. A Pearson product-moment correlation coefficient was computed between the activity patterns in the eccentricity runs (Center minus Periphery) and the Ziggerin runs (Within minus Between) for each ROI set (at y=35, 45, 55 or 65) in each participant. Figure 23B showed
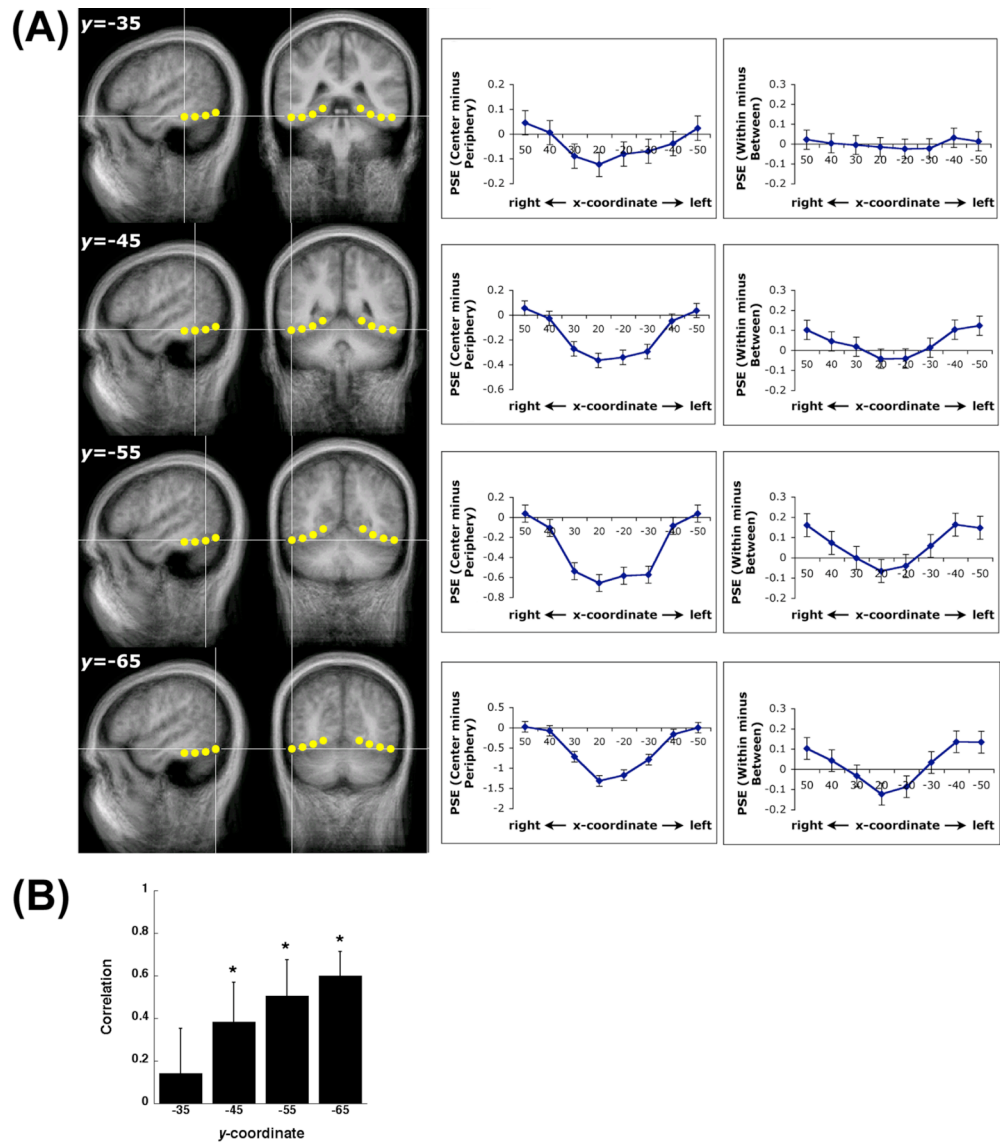
Figure 23. Relationship between eccentricity bias and preference for different discrimination demand. (A) Activations of the ventral occipitotemporal ROIs during the eccentricity and Ziggerin runs before training for all the participants (groups collapsed). At each point along the anterior-posterior axis (*y*-coordinate = -35, -45, -55, -65), there were eight ROIs covering the medial and lateral regions (*x*-coordinate = 50, 40, 30, 20, -20, -30, -40, -50). The activation patterns during the eccentricity runs and Ziggerin runs corresponded well with each other. The medial regions preferred peripheral presentation as well as between-class discrimination, whereas the lateral regions preferred central presentation and more demanding, within-class discrimination. Error bars represent the 95% confidence interval for the *x*-coordinate factor. (B) The average correlation between the eccentricity activity pattern and the pattern for within-minus-between activity in each ROI set. Asterisks indicate correlation values significantly different from zero. Error bars represent the 95% confidence intervals around the means.

the average correlations across all participants for each ROI set. The average correlation

coefficients were 0.14, 0.38, 0.50, 0.60 for y=35, 45, 55 and 65 respectively. The

coefficients for y=45, 55, and 65 were significant (all ps≤.0005) but that for y=35 was not

(p=.17). This confirms the link between the eccentricity bias and discrimination demand.

Figure 24 shows the training effects for the 32 ROIs. The patterns of training

effects were different between the two groups. The letter-like group emphasized the

medial regions more after training, as indicated by the positive training effect (activity

increase) for the medial regions and the negative training effect (activity decrease) for the

lateral regions. In contrast, the face-like group showed less clear emphasis on the medial

regions, especially in the most anterior ROIs (y=-35) and in the right hemisphere. A

Group (face-like, letter-like) × x-coordinate (50, 40, 30, 20, -20, -30, -40, -50) ANOVA

was conducted for the ROI sets at each y-coordinate (-35, -45, -55, -65), separately for

Ziggerin-within and Ziggerin-between conditions. There was a significant Group × x-

coordinate interaction for ROIs at y=-45 in both Ziggerin-within [$F(7,112)$=2.13,

p=.0458] and Ziggerin-between [$F(7,112)$=2.22, p=.0377] conditions. Further analyses

showed that the training effects were different across the medial and lateral regions only

for the letter-like group [$F(7,56)$=4.75, p=.0003] but not for the face-like group

[$F(7,56)$=1.18, p=.3238]. One-sample t-tests (p<.05) also showed that, for the letter-like

group, there were positive training effects in some medial regions [(x,y) = (30,-45) and

(20,-55) for Ziggerin-within] and negative training effects in some lateral regions [(x,y) =

(40,-45), (50,-45), (40,-55), (50,-55), (-50,-65), (-40,-65), (30,-65), (40,-65), and (50,-65),

all for Ziggerin-between]. The face-like group, however, only showed a positive training

effect in one ROI [(x,y) = (50,-35)]. These results suggested more distributed training

effects for the letter-like group, with more recruitment of the medial regions.
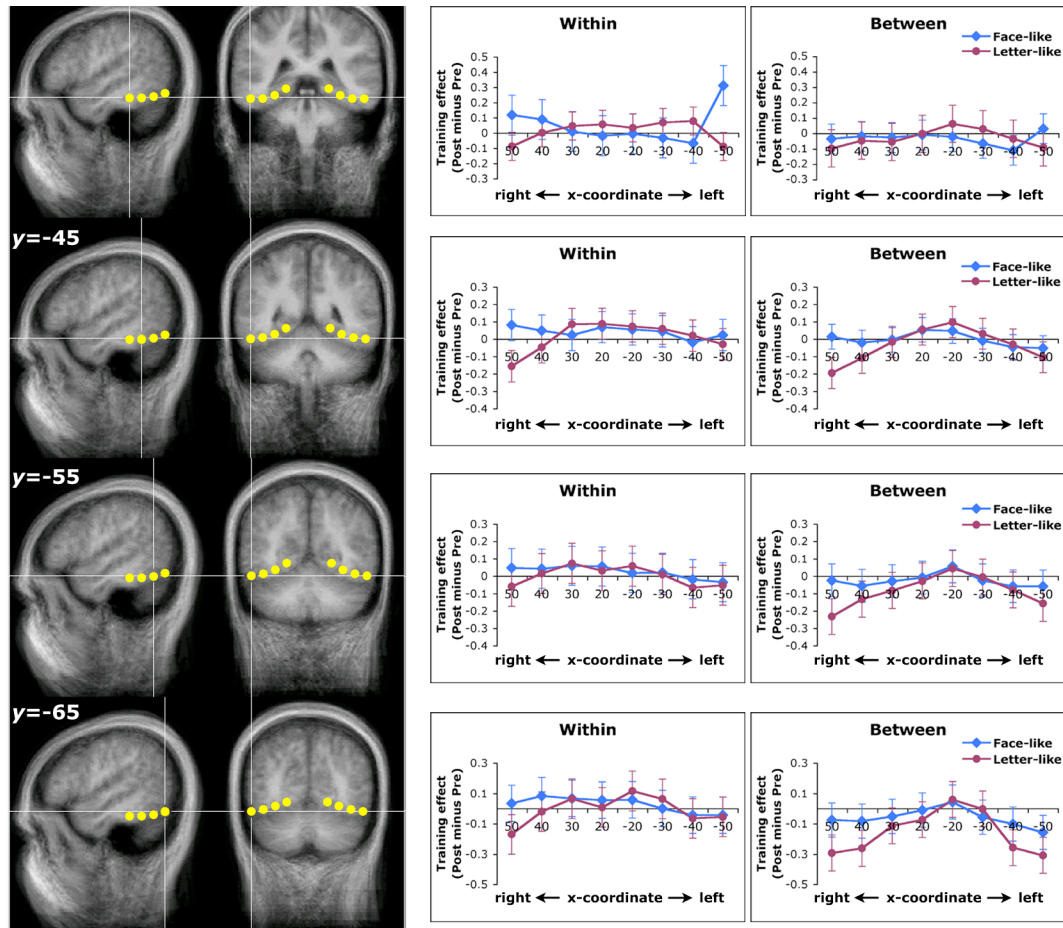
Figure 24. Training effects for the ventral occipitotemporal ROIs. The letter-like group showed more activity for the Ziggerins in the medial regions and less activity at the lateral regions after training. The face-like group didn't show the same pattern, with relatively more balanced involvement of medial and lateral regions. Error bars represent the 95% confidence intervals for the *x*-coordinate factor.

To better probe the differential training effects for the medial and lateral regions, correlations were computed between the training effects and the distance from the midline of the brain. For example, one can look at the set of four ROIs in the right hemisphere (i.e., at x=20, 30, 40, 50) at y=-65. If letter-like training indeed caused a shift
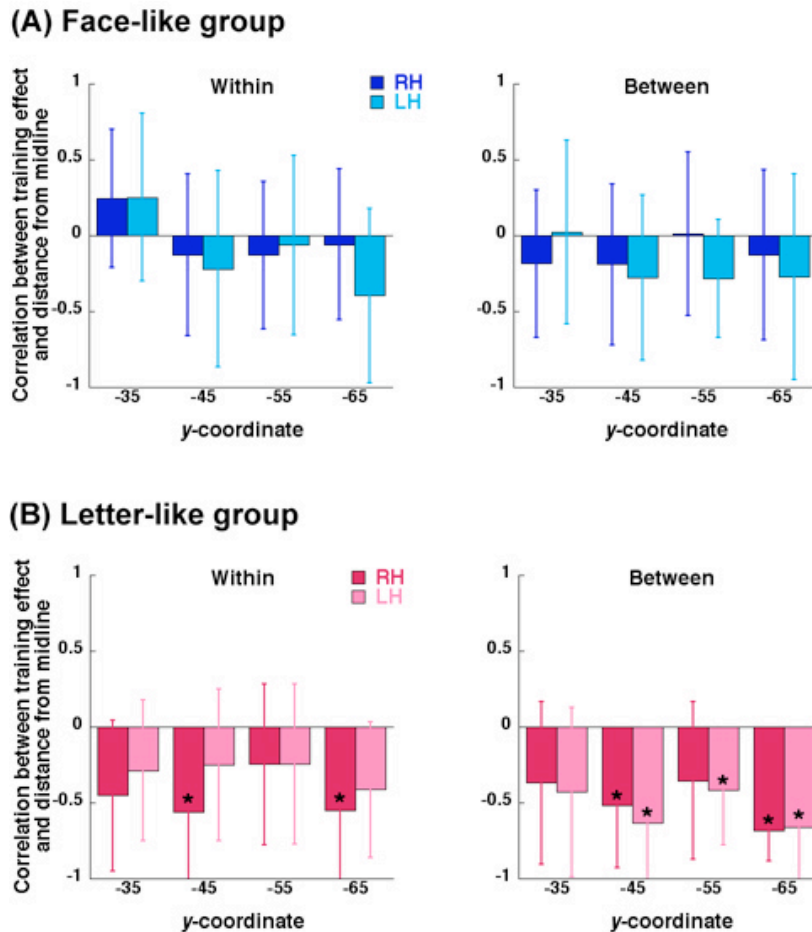
Figure 25. Correlation between the training effect (Post minus Pre) of an ROI and the distance of an ROI from the midline (absolute value of the x-cooridnate) for each of the four ROI sets (at y=35, 45, 55, 65) in each hemisphere (RH, LH). While the face-like group (A) showed no significant correlations, the letter-like group (B) showed negative correlations for multiple ROI sets in both hemispheres (marked by asterisks). A negative correlation indicates that the training effect led to increases especially in ROIs closer to the midline. In other words, the letter-like group put a larger emphasis on the medial, periphery-biased regions after training than before. Error bars represent the 95% confidence interval around the mean.

of emphasis to the medial regions, then one should expect the training effect to be negatively correlated with the distance from the midline. In other words, the further away an ROI is from the midline, the more negative the training effect should be. The correlation analyses confirmed the observation of the shift of emphasis to the medial regions for the letter-like group (Figure 25). A number of significant, negative correlations were found for the letter-like group in both hemispheres, especially for the Ziggerin-between condition. For the face-like group, however, no significant correlation was obtained.

General Discussion

Different neural markers of training were shown for the two groups. Face-like training resulted in a more focal effect, with a larger involvement of neural substrates near to the right fusiform face area (RFFA). The activity in the group-defined RFFA was heightened after training for the Ziggerins. In addition, the RFFA was found to have its training effects correlated with the magnitude of the configural effects with Ziggerins. In other words, participants who showed a higher sensitivity to the configurations of Ziggerin parts after training also showed a larger training effect in the RFFA. This is consistent with findings of the Greeble training studies (Gauthier & Tarr, 2002). Thus the face-like training with Ziggerins has succeeded in creating neural changes similar to expertise shown for Greebles and faces.

One aspect of the face-like training effects is worth our attention. The training effects were stronger in the RFFA defined at the group level than in the individually defined RFFA. The group-defined RFFA showed an increase in activity with training in

the Ziggerin-within condition and significant correlations with the configural effects in both Ziggerin-within and Ziggerin-between conditions. The individually defined RFFA, however, did not show any significant changes in activity, and showed a significant correlation with the configural effect only for the Ziggerin-between condition. This indicates that the training effect was not the strongest in the most face-selective voxels for each participant. As shown in Figure 26, the individually defined RFFAs for the participants in the face-like group clustered around the group-defined RFFA, though there was no perfect overlap (on average 20% of the group-defined RFFA overlapped
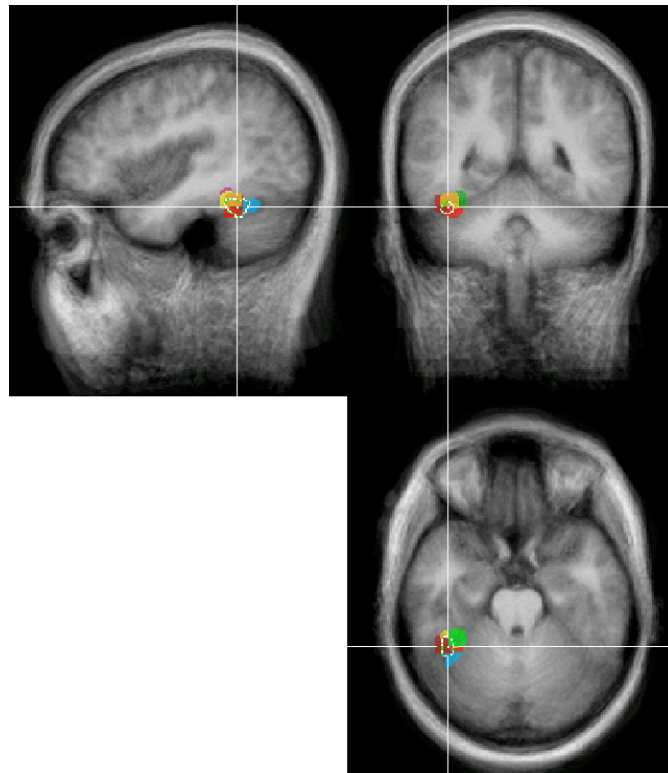

Figure 26. The RFFA defined for all the participants as a group (enclosed by the white line) and the RFFAs defined for each individual in the face-like group (in different colors). The individually defined RFFA clustered around the group-defined RFFA.

with the individually defined RFFA). The average Talairach coordinates of individual RFFAs (38.9, -43.4, -14.6) were also very similar to those of the group RFFA (40, -44, -18), with an average distance of 7.6mm. Therefore, it can be said that the face-like training effect overlapped the individually defined RFFA and was maximal at locations very close to the individually defined RFFA. This is different from the findings of previous Greeble studies (Gauthier *et al.*, 1999) in which the individually defined FFA showed a clear increase in activity after training.

There are at least two possible reasons of this pattern of results. It could be that this particular training with these particular objects could only engage a region near but not exactly overlapping with each individual's RFFA. This could be because of the geometry of the inanimate Ziggerins. An alternative possibility is that participants in the face-like group have acquired an intermediate level of face-like expertise, translating into only a partial involvement of the FFA proper. As discussed before the participants had to learn subordinate-level discrimination for six Ziggerin classes instead of one homogeneous set of Greebles in a similar training period (~10 hours). This means that participants received less training on each Ziggerin class compared with the Greebles (about 1/6th of the training experienced with Greebles). It should be noted that, unlike in Greeble studies, the basic-level advantage still existed for the face-like group after training, i.e., performance was still better at the class than individual level during the sequential matching task. More training could lead to an even smaller basic-level advantage and activation changes closer to the center of the individually defined RFFA. Note that discrimination training with novel objects in Op de Beeck *et al.*'s (2006) study also led to training effects in regions close but not right in the RFFA.

Letter-like training resulted in a more distributed pattern of training effects. It did not lead to more recruitment of the letter-selective area. Instead, a redistribution of activity occurred after letter-like training, such that the medial parts of the ventral occipitotemporal cortex became more important than the lateral parts. It has been shown in past work (Hasson *et al.*, 2002; Levy *et al.*, 2001) and here that the medial regions of the ventral occipitotemporal cortex are more responsive to objects presented in the periphery than at the center of the visual field. The current study also showed that the medial extrastriate regions are more responsive to basic-level categorization (discriminating among different classes of the novel Ziggerins) while the lateral regions are more recruited during subordinate-level discrimination (discriminating among Ziggerins within the same class). It is thus intuitive to find that letter-like training, relative to face-like training, led to more recruitment of neural substrates for basic-level categorization than those for subordinate-level discrimination.

There are several reasons for the lack of training effects in the letter-selective area. One plausible reason is that the current letter-like training was not demanding enough. There were only six classes of Ziggerins for participants to learn, much fewer than the number of basic characters or alphabets in most writing systems. The relatively easy task may be insufficient to push the participants to recruit the letter-selective neural mechanisms. It could be the same reason for the lack of style tuning effects during behavioral testing. If the training had been harder, the regularity in style could have been more useful for facilitating Ziggerin recognition and thus could have been extracted by the participants. The second plausible reason for the lack of training effects in the letter-selective area is that the letter-like training did not provide sufficient conditions for the

acquisition of expertise in visual letter perception. The training captured certain aspects of letter recognition: participants were trained to recognize, at the basic-level level, a large number of instances presented in a spatially organized clusters with a coherent style. However, one feature of letter perception – its close association with other processes involved in reading and writing – was not captured. The Ziggerin training involved little linguistic processing, with no elements of orthography, phonology, and semantics introduced in the training. Neither was there any type of motor learning that normally comes with letter writing or typing. The sensorimotor interaction with letters was demonstrated in imaging studies showing activation of the motor network by letter perception, and activation of the letter-selective visual regions during letter writing (James & Gauthier, 2006). It could take the involvement of the linguistic and/or motor areas to induce a training effect in the letter-selective area. Another reason for the lack of training effects in the letter-selective area concerns the properties of the Ziggerins. They are three-dimensional objects with texture and shading and a drastically different geometry compared with two-dimensional letters and characters. Perhaps only objects similar enough to letters can activate the letter-selective area, which would mean that there are strong constraints on selectivity that have to do with object geometry. Recent studies have shown that characters of unfamiliar writing systems can activate the letter-selective and word form areas to a similar degree as familiar letters (Reinke *et al.*, in press).

Overall, the face-like training increased recruitment of the face-selective area whereas the letter-like training did not affect the letter-selective region. Despite this, the current study succeeded in causing different patterns of neural activity changes as a result

of two different types of training experience. While this study reveals limitations in our current understanding of the factors that determine the specific recruitment of different neural substrates in visual expertise, it clearly establishes the role of experience in determining the recruitment of selective neural substrates in object recognition.

CHAPTER IV

CONCLUDING REMARKS: IMPLICATIONS AND FUTURE DIRECTIONS FOR STUDYING THE EFFECT OF EXPERIENCE ON VISUAL OBJECT RECOGNITION

Summary and overview

This dissertation was aimed at augmenting our understanding of the role of experience on selectivity for different categories in the visual object recognition system. As discussed in CHAPTER I, while existing theories postulate different mechanisms to account for the category selectivity found in the object recognition system, relatively little empirical work has addressed the effects of specific factors. In the current study, the role of experience was examined more directly using a training paradigm in which two groups of participants were exposed to different types of training experiences on the same set of novel objects. This allowed a relatively pure test of the effect of experience with object geometry controlled. Also, the use of novel objects presumably minimized higher-level processing (e.g., language, social and emotional processing) typically associated with real-world objects (e.g., letters, faces). This was a novel approach as previous studies have mostly focused on one type of training (Gauthier *et al.*, 1999; Moore *et al.*, 2006; Yue *et al.*, 2006).

The two types of training compared here were designed to produce two kinds of perceptual expertise, namely face and letter expertise. Face and letter perception represent two very different cases of object perception, with different behavioral and neural signatures of expertise. Face and letter perception display drastically different behavioral phenomena and neural selectivity patterns. With such differences, the contrast

between face and letter perception present the ideal model for studying the role of experience.

After the review of literature and motivation for this dissertation laid out in the first chapter, CHAPTERS II and III report the procedures for the two types of training and the behavioral and neural training effects. The face-like training was similar to the Greeble training studies which led to face-like behavioral phenomena and neural activations (Gauthier & Tarr, 1997, 2002; Gauthier *et al.*, 1999; Gauthier *et al.*, 1998; Rossion *et al.*, 2002; Rossion *et al.*, 2004). Face-like training was defined as learning to individuate a set of novel objects that share a configuration of parts, just like one has to discriminate very similar faces during person identification. Letter-like training, which has not been implemented before, was designed according to the hypothesis of the task demands associated with letter perception during reading (Wong & Gauthier, in press). It was defined as learning to rapidly recognize at the basic level a large number of items presented in a spatially organized cluster with a coherent style, just as letters appear in texts.

Face-like training resulted in behavioral and neural effects consistent with findings from prior Greeble training studies. Training caused a smaller basic-level advantage as well as increases in configural processing effects. Besides, fMRI activity in the right fusiform face area (RFFA) was enhanced after training, and the magnitude of this enhancement was positively correlated with the magnitude of the configural processing effect. This is consistent with the correlation of RFFA activity with the amount of holistic processing found for Greebles after training (Gauthier & Tarr, 2002). While these effects of face-like training replicate prior findings, the importance of the

current study lies on the fact that it is the first time that such effects are obtained in contrast to a training group that received an equivalent exposure to the same objects.

Unlike Greebles, the Ziggerins used in the current study were inanimate and definitely not face-like in geometry. Also, the Ziggerin set was much more heterogeneous, with six classes each with a unique set of parts. This means that participants were actually required to acquire face-like expertise in six basic-level categories. Given this, it is impressive that face-like training effects were obtained with only ten hours of training, similar to the time spent in Greeble training studies.

Letter-like training resulted in some but not all of the effects associated with letter expertise. Behaviorally, letter-like training caused a trend of enhanced basic-level advantage, i.e., recognition performance improved at the basic but not at a subordinate level. This is consistent with our recent study showing a larger basic-level advantage for characters in a familiar writing system (Wong & Gauthier, in press). Besides, the letter-like training group was faster than the other group in recognizing, at the basic level, triplets of Ziggerins. However, the letter-like group did not reveal any style tuning effect, and showed no changes in the activity of the letter-selective area. It is likely that the letter-like training in this study only introduced some but not all of the necessary factors to drive selectivity in the letter-selective area. Despite this, letter-like training resulted in a distributed pattern of neural changes, with more emphasis on the medial than the lateral regions in the ventral occipitotemporal cortex. These medial regions were also shown to be more responsive to object presentation in the periphery, and more active when one makes a basic-level judgment than a subordinate-level judgment. Therefore, the letter-like training in this study may better be named "basic-level categorization training",

which caused a larger basic-level advantage as well as more emphasis on neural substrates responsible for basic-level categorization. In this aspect, the current study is related to a recent bird training study (Scott *et al.*, 2006), in which participants learned subordinate-level categorization with a kind of birds (e.g., owls) and basic-level categorization with another kind (e.g., wading birds). In that study, the basic-level training was easy (with only two categories) and thus regarded as the control condition to show that the subordinate-level training effects were not merely a result of exposure. The current study was different in that the basic-level training was much more demanding (involving six basic-level categories with novel objects and requiring fast recognition of a large number of objects). Therefore, while Scott *et al.*'s (2006) study did not show any effects unique to basic-level training, in the current study the letter-like/basic-level training led to behavioral and neural changes not found for the face-like training.

Implications and future directions

Process-map theory

The results of this dissertation have important implications for our understanding of the origins of category selectivity in the visual object recognition system. Specifically, the study demonstrated how two different types of trainings led to qualitatively different patterns of selectivity for objects after training, when participants were tested in the same task with the same objects. The process-map theory provides a natural explanation of the mechanisms underlying the effect of experience (Bukach *et al.*, 2006; Gauthier, 2000). According to this hypothesis, the ventral occipitotemporal cortex contains regions with different pre-existing biases, or preference for specific types of recognition demand.

When recognition of a certain object category imposes a particular recognition demand, the substrates that best fulfill the recognition demand would be used more. This could be accomplished by feedback from frontal areas to different parts of extrastriate cortex that process the same object in parallel in different manners. Upon prolonged experience with an object category under the same recognition demands, the visual areas that best meet the recognition demands may become associated with activity in areas that represent the physical properties of a category. Through this process, visual areas can come to demonstrate category-selectivity independently of the current task demands.

While the process-map hypothesis was developed to explain how face-like expertise with non-face categories may recruit face-like processes and neural substrates, its applicability to the whole object recognition system has not been put to a real test. This dissertation confirmed the findings of previous studies, and went a step further by showing the interplay between recognition demands and experience over a wide range of regions in the ventral occipitotemporal cortex. For instance, the current study demonstrated how, regardless of training, different ventral occipitotemporal regions bias for perception at different levels of categorization, and that these regions were in turn recruited to different extent for the two types of training, depending on the level of categorization required.

Certainly it should be noted that while the current study succeeded in producing different effects with two types of training, it was not able to recreate all the behavioral and neural markers of letter-like expertise as expected. This does not mean that the process-map theoretical framework is necessarily inadequate – the process-map theory does not dictate the specific correspondences between training conditions and training

outcomes, although it does postulate that they exist. As we begin to understand additional organizing principles of the object recognition system, we may be in a position to more specifically predict the sufficient training conditions that can produce certain behavioral effects and recruit a given visual area. For letter selectivity to occur, it remains to be tested if stimulus geometry, task demand, association with linguistic and motor processing, or some or all of these factors, is necessary. The same question can be asked when one tries to explain selectivity for other object categories. The current paradigm demonstrates a fruitful approach to address such questions, much more directly than studies with common objects that confound geometry and past experience, or studies that involve only one type of training, in which training effects could result from exposure only.

Organizing principles of the ventral occipitotemporal cortex

Certainly gathering the sufficient conditions for creating selectivity for different categories is important. It may be even more valuable, however, to isolate specific principles with which the object recognition system is organized and dimensions along which processing of different categories vary. Two organizing principles of the visual object recognition system were revealed – level of categorization, and configural processing.

Level of categorization has been implied as one dimension along which substrates in the object recognition system vary. An eccentricity-bias gradient has been shown, with the lateral parts of the ventral occipitotemporal cortex preferring object presentation at the center of the visual field and the medial part preferring presentation in the periphery

(Hasson *et al.*, 2002; Levy *et al.*, 2001; Malach *et al.*, 2002). It was suggested that the lateral regions preferring central presentations receive more visual input from the fovea and thus can handle recognition requiring a higher resolution. The current study provides the first empirical support for the link between eccentricity bias and preferred level of categorization: regions with a center bias are also more active for more subordinate-level discrimination, while regions with a periphery bias are more active for coarser, basic-level discrimination.

Another principle that distinguishes between the properties of different neural substrates in the ventral occipitotemporal cortex is configural processing. A link between configural processing and the fusiform face area (FFA) can be implied from the finding that face perception is associated with configural processing (Young *et al.*, 1987) as well as higher FFA activity than other object categories (Downing *et al.*, 2006; Kanwisher *et al.*, 1997; Puce *et al.*, 1996). More direct evidence for this link comes from studies showing a correlation between configural processing and FFA activity for Greebles (Gauthier & Tarr, 2002). The current study also revealed a positive correlation between the magnitude of configural processing and the magnitude of training effects in the RFFA, thus confirming the link between the FFA and configural processing.

The discovery of the organizing principles of the ventral occipitotemporal cortex has very important implications for the study of functional specialization in the object recognition system. The nature of neuroimaging techniques may have encouraged researchers to focus on identifying localized regions for different categories and dividing the brain into multiple systems. However, postulating multiple systems to account for differences between categories is like re-description of data and does not allow

predictions to be made for new categories not studied before. A deeper understanding of category selectivity requires the understanding of its underlying mechanisms. The current findings suggest that it would be more accurate to regard the object recognition system as one single system within which many discrete and continuous functional dimensions interact, rather than as a collection of segregated, discrete compartments, or modules. A good future direction would be to discover more potential organizing principles of the object recognition system. This can be done with behavioral and modeling methods that contrast the computational demands of perception of different categories.

Explaining perceptual learning in general

This dissertation focuses on the experience-dependent changes in the object representations in the higher-level visual areas. In fact perceptual learning has been shown to occur at different points along the ventral visual pathway, from the early retinotopic areas to the higher visual areas that support object recognition directly (see reviews in Fahle, 2004; Fine & Jacobs, 2002; Kourtzi & DiCarlo, 2006).

Perceptual learning studies involving "simple features" are often found to involve early visual areas. For instance, discrimination training with oblique orientation patterns led to increased human fMRI responses in V1 (Furmanski *et al.*, 2004). A recent single cell recording study with monkeys showed that orientation discrimination training actually caused larger changes in V4 than in V1 (Raiguel *et al.*, 2006). The changes involved sharpening of the tuning curves in neurons optimal for signaling small differences around the trained orientation. Sigman *et al.* (2005) illustrated the role of top-down processes in the reorganization of visual pathway in a visual search task. After

learning to search for a target (e.g., T in a certain orientation) among distractors (Ts in other orientations), the early visual cortex was more active (and more correlated with performance) during the search for trained than untrained targets, whereas the lateral occipital, frontal and parietal areas were less active (and less correlated with performance) for the trained target search. Interestingly, the changes in the early visual cortex did not depend on whether the target was present or not, suggesting the modulation of top-down task demand.

As described in CHAPTER I, training studies involving more complex objects tend to reveal changes in higher visual areas (see Figure 1 for stimuli in some of the studies). Several studies revealed that training to perform extremely fine discrimination among morphed or highly similar objects caused the lateral occipital complex (LOC) to increase its activity (Jiang *et al.*, 2007; Op de Beeck *et al.*, 2006) and to become more sensitive to shape changes (Jiang *et al.*, 2007; Yue *et al.*, 2006). Subordinate-level categorization training with Greebles resulted in enhanced activity in the fusiform face area (FFA), as well as in areas overlapping with the LOC (Gauthier & Tarr, 2002; Gauthier *et al.*, 1999). Working memory training with novel objects made with polygons led to heightened activity not only in the LOC and FFA, but also in prefrontal and parietal areas (Moore *et al.*, 2006).

It is clear that there is no single locus of neural changes underlying perceptual learning. The most obvious difference among the findings is that some training changed the early visual areas while others led to changes in later stages of object recognition. There are also other differences. For example, orientation discrimination training can cause in some studies effects specific to the trained eye, indicating early visual area

103

changes (Karni & Sagi, 1991), but similar procedures can also lead to training effects

generalizable across eyes, suggesting changes in later stages (Schoups *et al.*, 1995).

Another difference concerns the recruitment of the FFA. Although the FFA has been

suggested to be associated with subordinate-level categorization (Gauthier, 2000;

Gauthier *et al.*, 1997), studies that emphasized fine discrimination among objects (Jiang

et al., 2007; Op de Beeck *et al.*, 2006; Yue *et al.*, 2006) did not always find an increase in

FFA activity. It is not surprising to see these differences among studies, given the variety

of object geometries and training experiences.

One attempt to offer a unifying account for the behavioral and neural effects of

perceptual learning comes from the reverse hierarchy theory, or the RHT (Ahissar &

Hochstein, 1998, 2004). The RHT asserts that perceptual learning results from the

increased accessibility of task-relevant information in different neural substrates. Also,

"perceptual improvement largely stems from a gradual top-down-guided increase in

usability of first high- then lower-level task-relevant information" (Ahissar & Hochstein,

2004, p.457). More specifically, easier training tasks with greater stimulus variability

(e.g., in terms of position) tend to increase recruitment of the later stages in the visual

pathways, while harder training tasks with smaller stimulus variability are learned at

earlier visual areas. Accordingly, both stimulus characteristics and task demands play

important roles in determining which part of the visual pathway would change upon

learning.

Here one can see the parallel between the challenge in explaining category

selectivity in the object recognition system and that in accounting for different loci

involved in perceptual learning. Multiple factors could contribute to the training effects

found in different neural regions in different studies, and empirical work is needed to tease apart the contributions of individual factors. In this sense, the type of training paradigm used in this dissertation would be a useful tool to bridge the fields of "simple feature" perceptual learning and of learning in complex object recognition.

Final conclusions

This dissertation reveals the unique contribution of recognition experience to the recruitment of different areas in the object recognition system. Manipulating training experiences with the same set of novel object resulted in face-like training effects consistent with previous findings in one case, and in the other case a distributed pattern of changes reflecting more emphasis on basic-level categorization. The comparison of two training procedures in one study demonstrated its power in probing individual principles that govern the organization of the ventral occipitotemporal cortex. It would be a fruitful approach in explaining not only category selectivity in the object recognition system but also mechanisms underlying perceptual learning in general.

Individual response times and group average proportion correct in naming for face-like
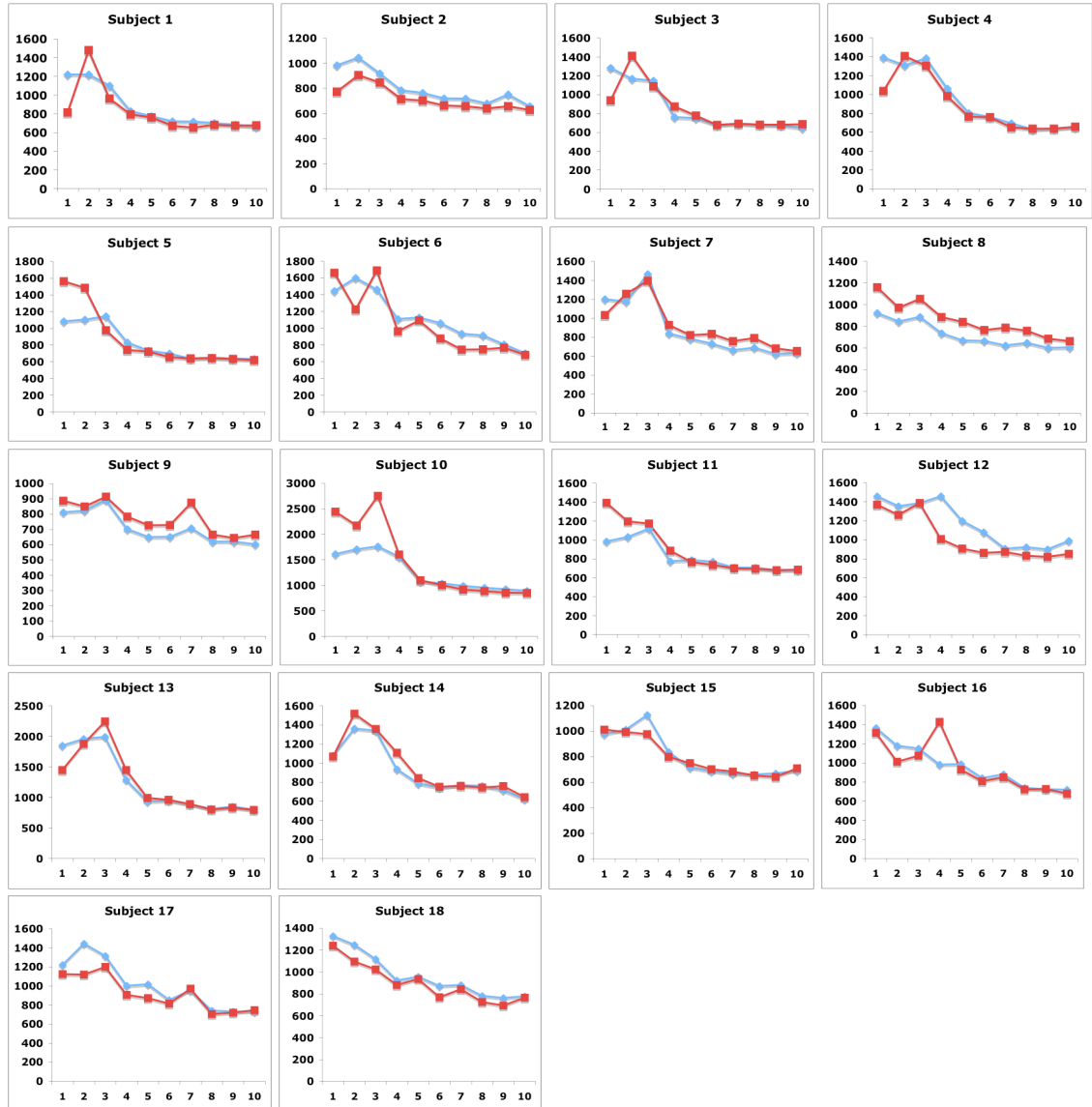
training.

Figure A. Changes in naming response times (msec) across sessions in face-like training (Blue – Ziggerins with names; Red – Ziggerins without names).

Table A. Group average naming accuracy (proportion correct) across sessions in face-like training.

Session	With names		Without names	
	M	SD	M	SD
1	0.96	0.03	0.93	0.12
2	0.94	0.04	0.91	0.13
3	0.94	0.04	0.94	0.07
4	0.97	0.03	0.98	0.04
5	0.96	0.03	0.97	0.04
6	0.97	0.02	0.97	0.03
7	0.96	0.03	0.96	0.04
8	0.97	0.02	0.97	0.03
9	0.97	0.03	0.96	0.04
10	0.97	0.02	0.98	0.02

Individual response times and group average proportion correct in verification for face-

like training.

Figure B. Changes in verification response times (msec) across sessions in face-like training.

Table B. Group average verification accuracy (*d'*) and bias across sessions in face-like training.

Session	Sensitivity (*d'*)		Bias	
	M	SD	M	SD
1	2.52	0.46	-0.41	0.17
2	2.62	0.53	-0.44	0.10
3	2.71	0.51	-0.47	0.21
4	2.87	0.40	-0.44	0.19
5	3.04	0.55	-0.40	0.20
6	2.99	0.57	-0.36	0.19
7	2.99	0.67	-0.33	0.22
8	3.05	0.56	-0.33	0.20
9	3.05	0.81	-0.26	0.21
10	3.26	0.68	-0.27	0.14

Individual response times and group average proportion correct in matching for face-like
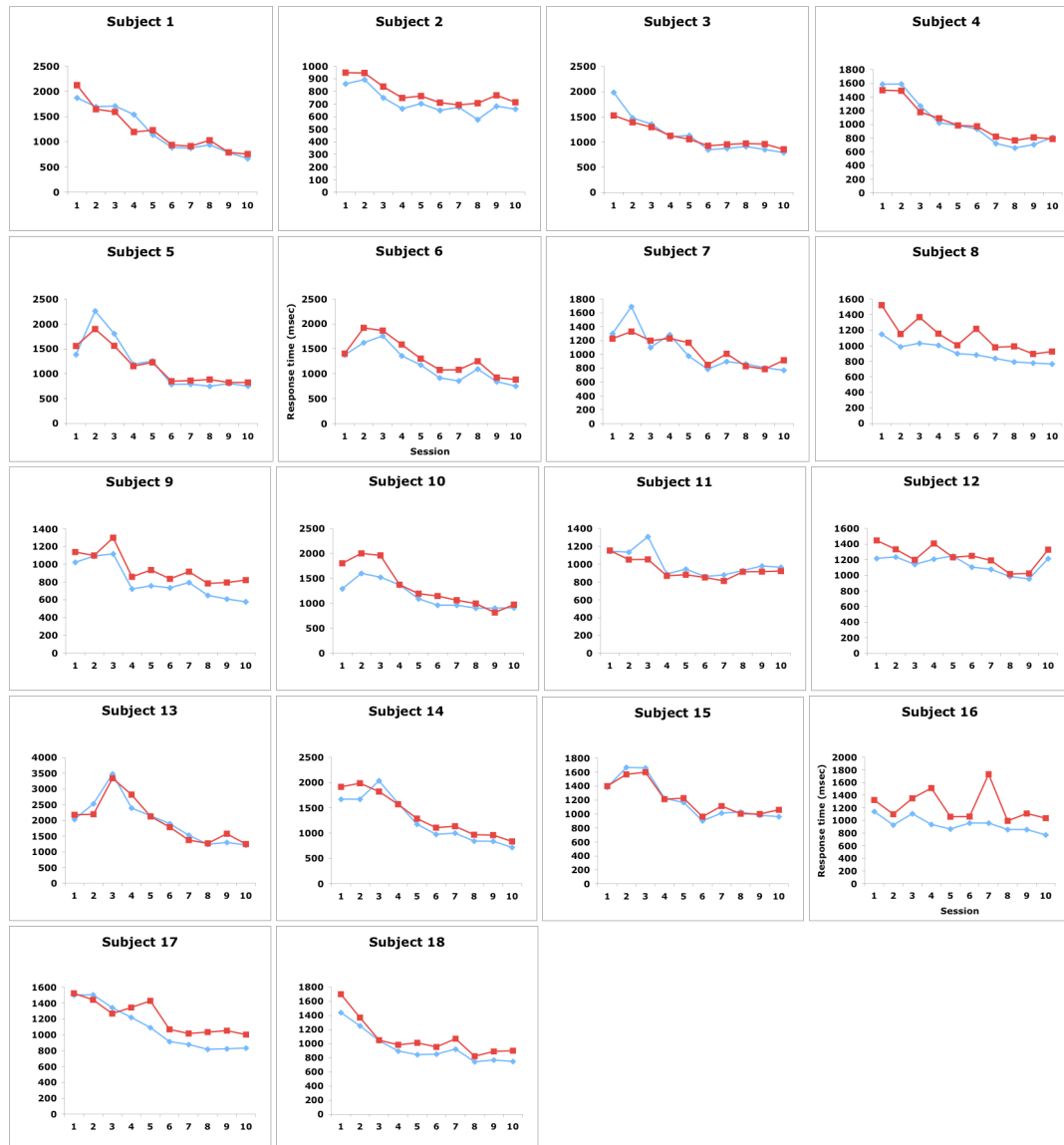
training.

Figure C. Changes in matching response times (msec) across sessions in face-like training (Blue – target present trials; red – target absent trials).

Table C. Group average matching accuracy (proportion correct) across sessions in face-like training.

Session	Target present		Target absent	
	M	SD	M	SD
1	0.91	0.03	0.69	0.11
2	0.92	0.03	0.65	0.15
3	0.92	0.04	0.64	0.14
4	0.93	0.03	0.68	0.14
5	0.95	0.02	0.72	0.11
6	0.94	0.03	0.72	0.11
7	0.95	0.02	0.73	0.12
8	0.95	0.02	0.74	0.12
9	0.95	0.04	0.74	0.11
10	0.96	0.03	0.74	0.12

Individual response times and group average proportion correct in naming for letter-like training.
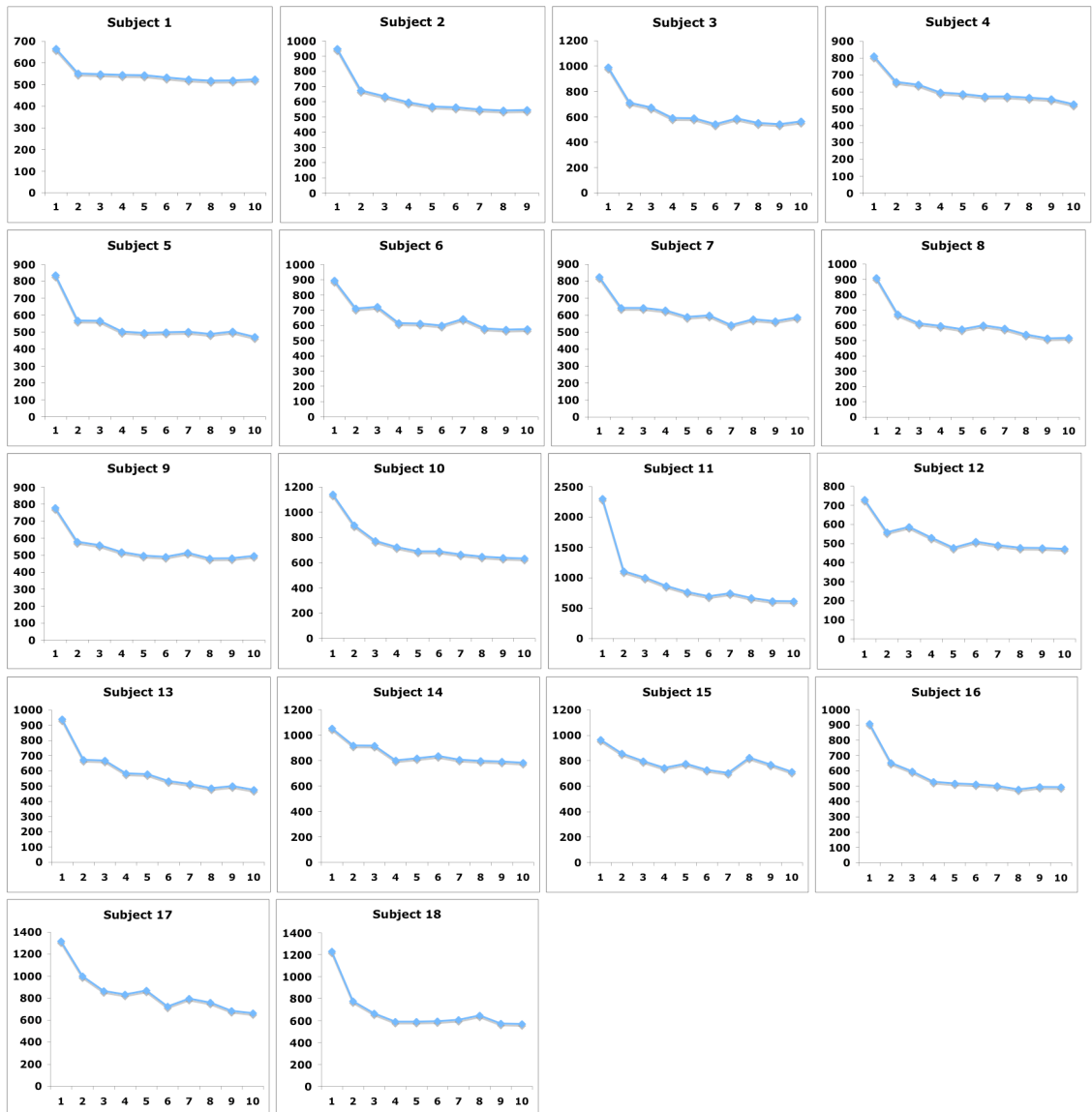
Figure D. Changes in naming response times (msec) across sessions in letter-like training.

Table D. Group average naming accuracy (proportion correct) across sessions in letter-like training.

Session	M	SD
1	0.97	0.03
2	0.98	0.02
3	0.98	0.02
4	0.97	0.03
5	0.97	0.03
6	0.97	0.03
7	0.97	0.03
8	0.97	0.03
9	0.96	0.04
10	0.97	0.03

Individual response times and group average proportion correct in verification for letter-like training.
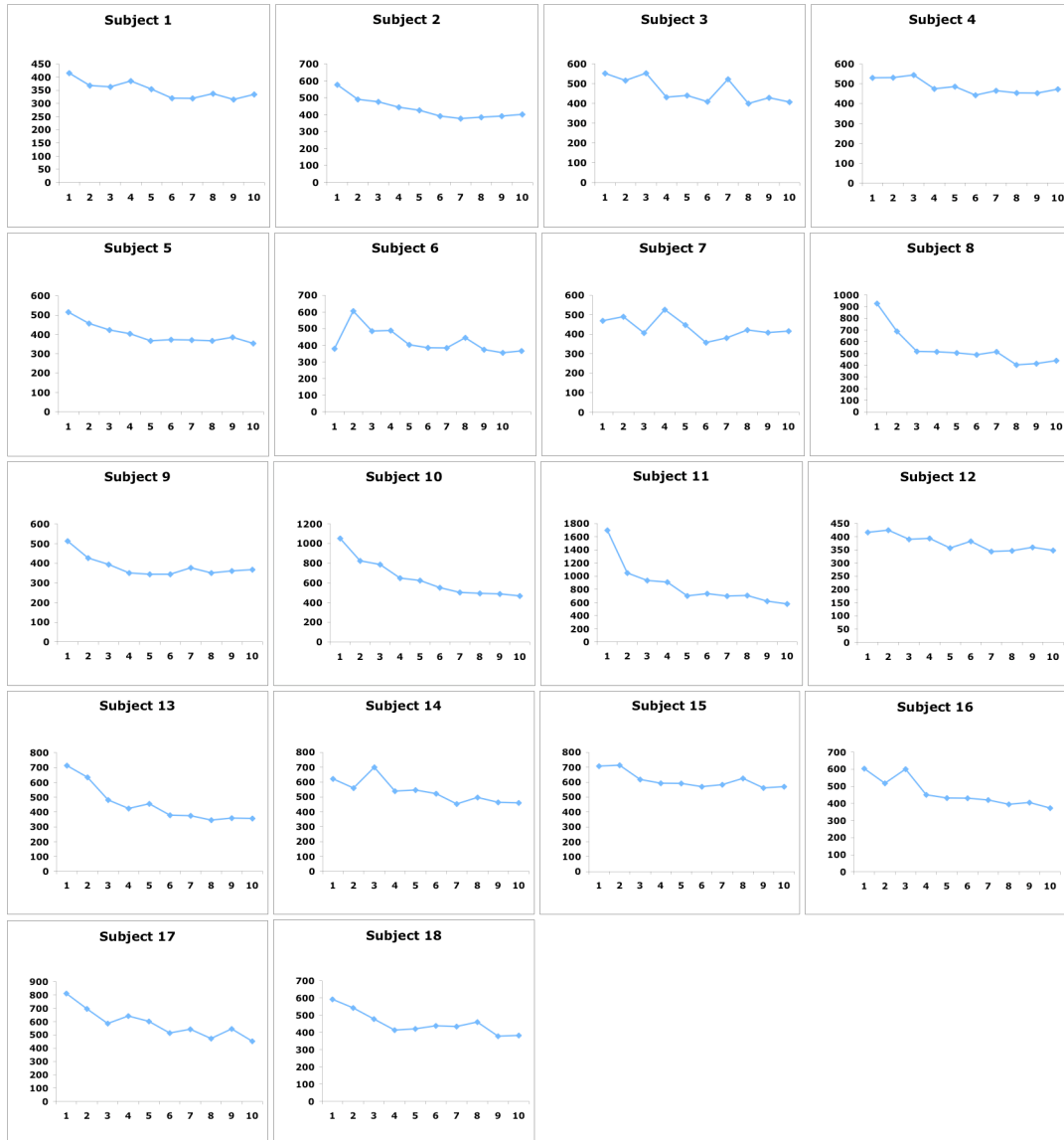
Figure E. Changes in verification response times (msec) across sessions in letter-like training.

Table E. Group average verification accuracy (*d'*) and bias across sessions in letter-like training.

Session	Sensitivity (*d'*)		Bias	
	M	SD	M	SD
1	4.10	0.71	0.04	0.25
2	4.11	0.60	-0.08	0.23
3	4.00	0.69	-0.04	0.16
4	3.88	0.66	-0.07	0.21
5	3.82	0.74	-0.06	0.20
6	3.76	0.72	-0.01	0.19
7	3.61	0.83	-0.11	0.18
8	3.85	0.91	-0.12	0.21
9	3.76	0.89	-0.05	0.22
10	3.62	0.75	-0.02	0.26

Individual response times and group average proportion correct in matrix scanning for letter-like training.
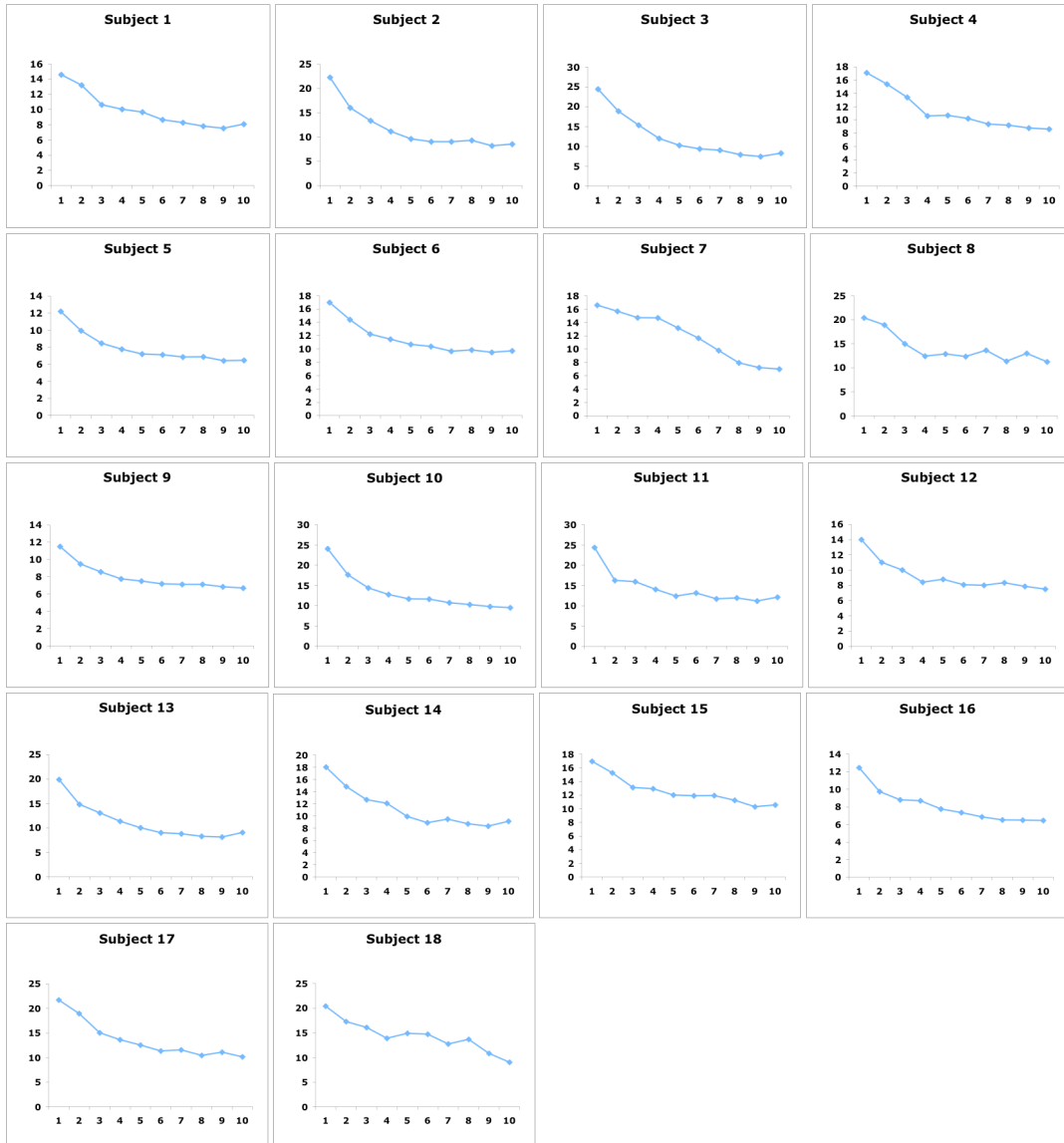

Figure F. Changes in matrix scanning response times (sec) across sessions in letter-like training.

Table F. Group average matrix scanning accuracy (proportion correct) across sessions in letter-like training.

Session	M	SD
1	0.88	0.06
2	0.92	0.06
3	0.94	0.04
4	0.94	0.03
5	0.93	0.05
6	0.94	0.05
7	0.94	0.04
8	0.93	0.04
9	0.93	0.06
10	0.92	0.07

APPENDIX G

Performance for the different target-distractor matching conditions in the part matching task

The congruent and incongruent conditions were split for further analyses in the part matching task where distractor changes can occur at the individual level. A Three-way ANOVA was conducted for response times with the factors of Group (face-like, letter-like), Alignment (aligned, misaligned), and Matching Conditions (Target same Distractor same, Target same Distractor different, Target different Distractor different, Target different Distractor same). The Group × Alignment × Matching Condition interaction was close to significant for response time [$F(6,129)=1.996$, $p=.0708$]. Scheffé tests ($p<.05$) on the aligned trials showed that, for the face-like group, responses were faster in target-same trials when the distractor parts were also the same compared with when the distractor parts were different. In target-different trials, responses were faster when the distractor parts were also different than when the distractor parts were the same. In other words, performance was better when the target and distractor parts led to the same correct response. Such a difference between congruent and incongruent trials was not found for the face-like group when the target and distractor parts were misaligned. Neither did it occur for the letter-like group or the control group, whether the target and distractor parts were aligned or not. The three-way ANOVA on proportion correct did not show a significant three-way interaction [$F<1$].
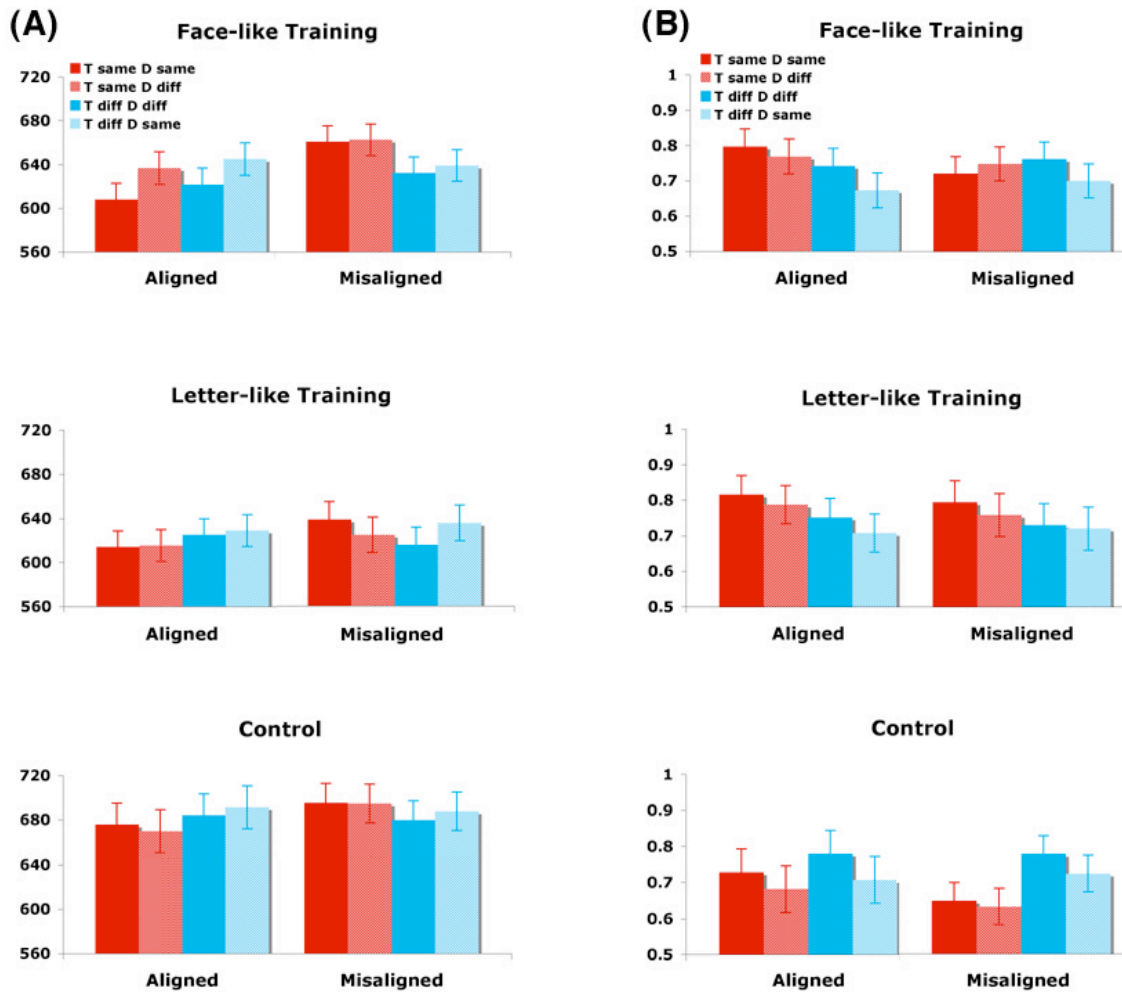
Figure G. Response times (A) and proportion correct (B) for the different target-distractor matching conditions in the part matching task when distractor changes occurred at the individual level. Error bars represent 95% confidence interval for the matching condition variable (matching conditions: T same D same = Target part same Distractor part same; T same D diff = Target part same Distractor part different; T diff D same = Target part different Distractor part same; T diff D diff = Target part different Distractor part different).

REFERENCES

Ahissar, M., & Hochstein, S. (1998). Perceptual learning. In V. Walsh & J. J. Kulikowski (Eds.), *Perceptual constancies: Why things look as they do* (pp. 455-498): New York: Cambridge University Press.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *TRENDS in Cognitive Sciences, 8*(10), 457-464.

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M., *et al.* (1994a). Face recognition in human extrastriate cortex. *J Neurophysiol, 71*(2), 821-825.

Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994b). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex, 4*(5), 544-554.

Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *PNAS, 104*(21), 9087-9092.

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.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *The Journal of Neuroscience, 16*(13), 4207-4221.

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

Bub, D. N., Arguin, M., & Lecours, A. R. (1993). Jules dejerine and his interpretation of pure alexia. *Brain and Language, 45*(4), 531-559.

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *TRENDS in Cognitive Sciences, 10*(4), 159-166.

Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience, 15*, 704-717.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current theories. In D. Besner, T. G. Waller & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice*. Orlando: Academic Press.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2*(10), 913-919.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *NeuroImage, 22*, 466-476.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., *et al.* (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*(Pt 2), 291-307.

Cohen, L., Lehericy, S., Cohochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain, 125*, 1054-1069.

Cox, D., & Savoy, R. (2003). Functional magnetic resonance imaging (fmri) "brain reading": Detecting and classifying distributed patterns of fmri activity in human visual cortex. *Neuroimage, 19*, 261-270.

Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J.-B., Le Bihan, D., *et al.* (2004). Letter binding and invariant recognition of masked words. *Psychological Science, 15*(5), 307-313.

Dehaene, S., Le Clec'H, G., Poline, J. B., Le Bihan, D., & Cohen, L. (2002). The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *Neuroreport, 13*, 321-325.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16*, 1453-1461.

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

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 914-927.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron, 23*(1), 115-125.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601.

Fahle, M. (2004). Perceptual learning: A case for early selection. *Journal of Vision, 4*, 879-890.

Farah, M. J. (2004). *Visual agnosia* (2nd ed.). Cambridge, MA: The MIT Press.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning tasks: A review. *Journal of Vision, 2*, 190-203.

Flowers, D. L., Jones, K., Noble, K., VanMeter, J., Zeffiro, T. A., Wood, F. B., *et al.* (2004). Attention to single letters activates left extrastriate cortex. *Neuroimage, 21*(3), 829-839.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*(5502), 312-316.

Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Current Biology, 14*(573-578).

Garrett, A. S., Flowers, D. L., Absher, J. R., Fahey, F. H., Gage, H. D., Keyes, J. W., *et al.* (2000). Cortical activity related to accuracy of letter recognition. *Neuroimage, 11*(2), 111-123.

Gauthier, I. (2000). What constrains the organization of the ventral temporal cortex? *Trends in Cognitive Science, 4*(1), 1-2.

Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P., & Gore, J. C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Current Biology, 7*(9), 645-651.

Gauthier, I., Curby, K. M., Skudlarski, P., & Epstein, R. A. (2005). Individual differences in ffa activity suggest independent processing at different spatial scales. *Cognitive, Affective, & Behavioral Neuroscience, 5*(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. (2000a). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191-197.

Gauthier, I., Tanaka, J. W., & Brown, D. D. (submitted). Failures of selective attention in the complete face composite paradigm.

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.

Gauthier, I., Tarr, M. J., Moylan, J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2000b). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12*(3), 495-504.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38*(15/16), 2401-2428.

Gauthier, I., Wong, A. C.-N., Hayward, W. G., & Cheung, O. S.-C. (2006). Font-tuning associated with expertise in letter perception. *Perception, 35*(4), 541-559.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15*, 870-878.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature, 349*(6305), 154-156.

Grelotti, D. J., Klin, A. J., Gauthier, I., Skudlarski, P., Cohen, D. J., Gore, J. C., *et al.* (2005). Fmri activation of the fusiform gyrus and amygdala to cartoon characters but not to faces in a boy with autism. *Neuropsychologia, 43*, 373-385.

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

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Reviews Neuroscience, 27*, 649-677.

Grill-Spector, K., Sayres, R., & Ress, D. (2007). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nat Neurosci, 10*(1), 133.

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron, 37*, 1027-1041.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, M. (2002). Eccentricity bias as an organizing principle for human high order object areas. *Neuron, 34*, 479-490.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., *et al.* (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc. Natl. Acad. Sci. USA, 88*, 1621-1625.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc. Natl. Acad. Sci. USA, 96*, 9379-9384.

James, K. H., & Gauthier, I. (2006). Letter processing automatically recruits a sensory-motor brain network. *Neuropsychologia, 44*, 2937-2949.

James, K. H., James, T. W., Jobard, G., Wong, A. C.-N., & Gauthier, I. (2005). Letter processing in the visual system: Different activation patterns for single letters and strings. *Cognitive, Affective, and Behavioral Neuroscience, 5*(4), 452-466.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron, 53*(6), 891.

Jolicoeur, P. (1990). Orientation congruency effects on the identification of disoriented shapes. *Journal of Experimental Psychology: Human Perception and Performance, 16*(26), 351-364.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*, 759-763.

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

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of National Academy of Sciences of the United States of America, 88*, 4966-4970.

Kourtzi, Z., & DiCarlo, J. J. (2006). Learning and neural plasticity in visual object recognition. *Current Opinion in Neurobiology, 16*(2), 152.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience, 4*(533-539).

Levy, I., Hasson, U., Harel, M., & Malach, R. (2004). Functional analysis of the periphery effect in human building related areas. *Human Brain Mapping, 22*, 15-26.

Longcamp, M., Anton, J.-L., Roth, M., & Velay, J.-L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage, 19*(4), 1492-1500.

Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences, 6*(176-184).

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *TRENDS in Cognitive Sciences, 7*(7), 293-299.

McNeil, J. E., & Warrington, E. K. (1991). Prosopagnosia: A reclassification. *Quarterly Journal of Experimental Psychology, 43A*(2), 267-287.

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

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-377.

Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature, 372*(6503), 260-263.

Noppeney, U., Price, C. J., Penny, W. D., & Friston, K. J. (2006). Two distinct neural mechanisms for category-selective responses. *Cerebral Cortex, 16*(3), 437-445.

O'Toole, A. J. O., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience, 17*(4), 580-590.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience, 26*(50), 13025-13036.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neuroscience, 193*, 603-608.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Polk, T. A., Stallcup, M., Aguirre, G. K., Alsop, D. C., D'Esposito, M., Detre, J. A., *et al.* (2002). Neural specialization for letter recognition. *Journal of Cognitive Neuroscience, 14*(2), 145-159.

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *Neuroimage, 19*, 473-481.

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

Raiguel, S., Vogels, R., Mysore, S. G., & Orban, G. A. (2006). Learning to see the difference specifically alters the most informative v4 neurons. *Journal of Neuroscience, 26*(24), 6589-6602.

Reinke, K., Fernandes, M., Schwindt, G., O'Craven, K., & Grady, C. L. (in press). Functional specificity of the visual word form area: General activation for words and symbols but specific network activation for words. *Brain and Language*.

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

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.

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., *et al.* (2000). The n170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *Neuroreport, 11*(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 Sciences of the United States of America, 101*, 14521-14526.

Sanocki, T. (1987). Visual knowledge underlying letter perception: Font-specific, schematic tuning. *Journal of Experimental Psychology: Human Perception and Performance, 13*(2), 267-278.

Sanocki, T. (1988). Font regularity constraints on the process of letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 472-480.

Sanocki, T. (1991). Looking for a structural network: Effects of changing size and style on letter recognition. *Perception, 20*(4), 529-541.

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularity. *Journal of Physiology, 483*, 797-810.

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience, 25*(47), 11055-11059.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., & Curran, T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. *Journal of Cognitive Neuroscience, 18*, 1453-1465.

Sigman, M., Pan, H., Yang, Y., Stern, E., Sillbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron, 46*, 823-835.

Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex. *Human Brain Mapping, 27*(1), 77-89.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fmri study. *Neuron, 35*, 1157-1165.

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., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science, 16*(2), 145-151.

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

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.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109-139.

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex, 13*, 90-99.

Tarkiainen, A., Cornelissen, P. L., & Salmelin, R. (2002). Dynamics of visual feature analysis and object-level processing in face versus letter-string perception. *Brain, 125*(5), 1125-1136.

Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain, 122*(Pt 11), 2119-2132.

Tarr, M. J., & Gauthier, I. (2000). Ffa: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience, 3*(8), 764-769.

Vigneau, M., Jobard, G., Mazoyer, B., & Tzourio-Mazoyer, N. (2005). Word and non-word reading: What role for the visual word form area? *Neuroimage, 27*, 694-705.

Vinckier, F., Dehaene, S., & Jobert, A. (2007). Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron, 55*(1), 143-156.

Warrington, E. K., & Shallice, T. (1980). Word form dyslexia. *Brain, 103*, 99-112.

Weisberg, J., van Turennout, M., & Martin, A. (2006). A neural system for learning about object function. *Cereb. Cortex*, bhj176.

Wong, A. C.-N., & Gauthier, I. (in press). An analysis of letter expertise in a levels-of-categorization framework. *Visual Cognition*.

Wong, A. C.-N., Gauthier, I., Woroch, B., Debuse, C., & Curran, T. (2005). An early electrophysiological response associated with expertise in letter perception. *Cognitive, Affective, and Behavioral Neuroscience, 5*(3), 306-318.

Wong, A. C.-N., Jobard, G., James, K. H., James, T. W., & Gauthier, I. (submitted). Expertise with characters in alphabetic and non-alphabetic writing systems engage overlapping occipito-temporal areas.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex, 15*(8), 1234-1242.

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

Young, A. W. (1992). Face recognition impairments. *Phil Trans R Soc Lond, 335*, 47-54.

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

Yovel, G., & Kanwisher, N. (2004). Face perception: Doman specific, not process specific. *Neuron, 44*, 889-898.

Yovel, G., Paller, K. A., & Levy, J. (2005). A whole face is more than the sum of its halves: Interactive processing in face perception. *Visual Cognition, 12*(2), 337-352.

Yue, X. M., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision Research, 46*(22), 3802-3811.