Neural correlates of visual object learning

by

Jesse S. Husk

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Psychology, Neuroscience and Behaviour
McMaster University

Descriptive Note

Doctor of Philosophy (2009)
Psychology, Neuroscience and Behaviour
McMaster University
Hamilton, Ontario

TITLE: Neural correlates of visual object learning

AUTHOR: Jesse S. Husk

SUPERVISORS: Professor Patrick J. Bennett & Professor Allison B. Sekuler

NUMBER OF PAGES: xiv, 145

Abstract

Neural correlates of visual object learning

Jesse S. Husk
Doctor of Philosophy
Department of Psychology, Neuroscience and Behaviour
McMaster University
2009

Faces are often deemed special objects because they are associated with behavioural and physiological characteristics that differ from those of other objects. These characteristics may indicate that faces are processed with separate mechanisms than other objects. On the other hand, these characteristics may be the result of our extensive experience with faces. If so, other objects should exhibit these same characteristics with sufficient exposure. This prediction has begun to be addressed both from studies of real-world experts and from studies that explicitly manipulate experience with non-face objects in the lab.

Contributing to this larger framework, here we demonstrate that : (1) large inversion effects can be obtained through training alone, therefore large face inversion effects are insufficient evidence of specialized face-processing mechanisms; (2) house-identification training substantially improves behavioural performance but has minimal impact on fMRI activity recorded in areas that preferentially respond to houses or faces, nor in retinotopically-defined early visual areas. (3) house-identification training systematically reduces the amplitude of late ERP components in the range of 200-300 ms, and (4) the relative patterns of ERP responses to faces and houses remain quite stable after house-identification training, with faces continuing to exhibit larger, earlier N1 responses than houses.

Together, these results suggest that, although some behavioural characteristics attributed to specialized face processing can be adequately explained through experience alone, training of non-face objects does not readily reduce existing differences in the fMRI and EEG signatures of face and object processing.

Preface

This thesis comprises 6 chapters. Chapters 2-5, presenting the main research, are written in journal article format.

Chapter 2 was written collaboratively with my supervisors, Allison Sekuler and Patrick Bennett, and has been previously published in the journal Vision Research. The article is reproduced here (re-formatted to meet McMaster thesis guidelines) with full permission from Vision Research. As stated under "Author's Rights", authors retain "the right to include the journal article, in full or in part, in a thesis or dissertation". Full author's rights can be accessed at the site:

http://www.elsevier.com/wps/find/authorsview.authors/copyright#whatrights

Chapter 3 was written collaboratively with Allison Sekuler, Patrick Bennett, and Lisa Betts. Chapters 4 and 5 were written collaboratively with Allison Sekuler and Patrick Bennett. These three chapters have not yet been submitted for publication.

For all chapters, as primary author, I oversaw all aspects of the research. The generation of ideas and the development of these ideas into experimental designs was a collaborative effort between myself and the other co-authors. I was solely responsible for the experimental programming for Chapters 2, 4, and 5, and was assisted with programming by Lisa Betts for Chapter 3. Data collection was performed largely by myself, with some assistance from research assistant Donna Waxman (Chapter 2), co-author Lisa Betts (Chapter 3), and several undergraduate research assistants (Chapters 4 & 5). I was primarily responsible for the data analysis for all chapters. My supervisors and other co-authors worked with me to determine appropriate analyses, and to interpret the subsequent results. I was responsible for initial drafts of all chapters. Subsequent edits were shared collaboratively amongst all co-authors.

Acknowledgements

Pat and Allison: Thank you for the guidance you've provided over the years. Particularly, for helping to mould my thoughts into workable experiments, for teaching me to think more critically about what I read, for encouraging me to speak and write with conviction, and for fine-tuning my grammar. Zahra: As much as I dreaded the idea of communal grad-student workspace, it's been a relief and a joy to have you at my side from the beginning to the end. I will miss our long, distracting conversations tremendously. Thanks to Carl for being my first lab mentor, to Guillaume for being the most fun and energetic model of academic excellence I can imagine, and to Lisa for all the long hours spent patiently guiding me through the rigors of fMRI analysis. Thanks, especially, to Donna: I may not have given you as many subject-hours as some of the others, but I still don't know what I would have done without your astonishing efficiency, and much more importantly, your friendship, over the years. Thanks also to Erin, for your sage, grounding advice, to Sam, for your humour and gentle concern, and to Jana for all the long phone-hours that have so enriched this past year. To Mom for always being there (and knowing exactly what all this is about!), and to Liam for teaching me I don't (always) have to have the last word.

Contents

Chapter 1

General Introduction

1.1 Overview

Face perception research has become a dominant domain of research in the field of visual pattern perception, occupying large volumes of yearly publications and conference presentations. At the heart of much of this research are the questions: Are faces special objects? And, does the brain process faces differently than it processes other object classes?

This thesis is not directly about face processing, and will employ face stimuli only occasionally for purpose of control, but the projects that make up this thesis exist within the context of recent increased interest in face processing, and many of the questions addressed by these projects arise in direct response to the face processing literature. In particular, these projects examine the development of expertise with sets of novel objects. The purpose of this introduction, therefore, is to ground this body of research within the framework of the face perception literature.

The aim of this thesis is to explore the properties associated with the development of visual object expertise across short-term in-lab training of young adult subjects, both behaviourally and physiologically. The experiments reported here not only contribute to an understanding of visual pattern learning, but also, by comparison with the existing face literature, address questions about how faces are processed differently from other objects. Specifically, this thesis: considers whether the inversion effect is a valid indicator of face-specific processing (Chapter 2), tests existing models of object-organization within the ventral temporal cortex (Chapter 3), and re-examines electrophysiological markers

of face processing and expertise (Chapters 4 & 5).

1.2 Faces as special objects

Humans, as a social species, spend much of their time in social groupings, from family settings to school or work settings. The ability to differentiate among people, and to quickly recognize those individuals that we regularly encounter, are critical parts of daily life. Faces do not only provide information about the identity of a person, but also provide information about a person's emotional and cognitive state, age, gender, and attractiveness. Other facial cues (such as the direction of eye-gaze) provide information about salient events in the near environment. For these reasons, efficient face processing presumably is important for successfully maneuvering through social encounters. Due to their familiarity and behavioral relevance, it is intuitively appealing to assume that faces may have acquired some special status during our evolutionary and/or developmental history, and that humans may be predisposed to attend to faces, to process faces with special efficiency, and might be equipped with neural circuitry designed to facilitate the processing of faces.

The face perception literature has been mainly dominated by three major questions. First: Is face processing qualitatively different from the processing of other objects? Second: If faces are processed differently, in what way does the processing differ from that of other objects: how does the processing differ mechanistically, and how is that implemented cortically? Third: If face and object processing differ, how did this difference arise? Is there an innate predisposition to treat faces as special objects, or did the differences arise due to a high level of experience with face processing (perhaps due to the prevalence and social importance of faces)? The following sections briefly review the dominant points that have arisen within the field.

1.3 Mechanisms of face perception

One of the earliest suggestions that face perception might differ from the perception of other classes of objects arose from the inversion paradigm. It has been long known that object processing is impaired by picture-plane rotation (e.g., Jolicoeur, 1985). Recognition and identification of common objects is more difficult (i.e., accuracy decreases and response latency increases) when the objects are rotated away from their standard, or canonical, orientation. Yin (1969) noticed that this effect was larger for faces. Yin com-

pared the reaction times for recognition upright and inverted presentations of a variety of stimuli, including faces, person stick-figures, and costumes. Face recognition was particularly impaired by inversion, more so than for other tested classes. This result was not related to task difficulty, as the face task was not the most difficult of discriminations subjects were asked to perform. Importantly, Yin (1969) noted that the subjects reported processing the faces with a different strategy than the other objects: they seemed to see the faces as a single whole, rather than through the piecemeal extraction of features:

> [Subjects] seemed to use two alternative strategies, either searching for some distinguishing feature or attempting to get a general impression of the whole picture. The first tended to be used mostly for materials; the second was used mostly for faces....None of the subjects, however, reported being able to use the second strategy when looking at the inverted face. - (Yin, 1969, page 145)

1.3.1 Featural and configural processing

Yin's work sparked a massive line of research on the mechanisms underlying face inversion that continues to this day. Consistent with Yin's original speculation, it is now widely held that face and object processing differ qualitatively. Specifically, that object recognition primarily relies on featural processing, but that face recognition relies on both featural and configural processing. It is worth taking care to define featural and configural processing because both the terms configural and featural have been applied loosely, and sometimes inconsistently, within the field (Maurer et al., 2002).

Featural processing generally refers to the separate processing of individual features, a definition that is complicated by the difficulty of defining which aspects of a face should be labeled as features. In most cases, features refer to the nameable parts of a face (e.g. eyes, nose, mouth, eyebrows)[1]. A challenge for this dichotomy between features and configurations is to demonstrate that the visual system recognizes the nameable parts as features, and gives them special preference as features over other parts of the face that could equally be considered so. For example, there is no inherent reason that the space between the eyes or between the nose and upper lip could not, themselves, be treated as features; However, they are rarely (if ever) defined as such.

[1]Note that this definition also has problems because many nameable parts of faces (e.g., cheek, forehead) generally are not considered to be "features".

Configural processing is a term that sometimes refers to holistic processing (the quality of perceiving the face as a united whole, with a corresponding deficit in the ability to process subsections of the face independently), and other times refers to the processing of first- and second-order spatial relations among the nameable parts (Maurer et al., 2002). In the case of faces, the first-order relations refer to the consistent vertical structure of faces: eyes are above the nose which is above the mouth. Second-order relations, however, refer to the metric distances among nameable parts. Because first-order relations are similar in all human faces, such information may be useful for detecting faces but would not be useful for discriminating among faces. Therefore, second-order relations are thought to be the configural information that contributes most significantly to face discrimination and identification.

Generally, it is assumed that faces engage a combination of featural and configural processing, whereas objects engage only featural processing. Furthermore, inverting a face by 180 deg typically is thought to disrupt configural processing more than featural processing, rendering the processing of inverted faces more similar to that of upright or inverted objects. Objects, in this view, are less impacted by inversion because object processing does not engage configural processing, even in the upright orientation. When reduced to featural processing due to inversion, there is little deficit because the objects were already processed featurally, when upright. This explanation of the inversion effect implies that when inversion effects *are* observed for non-face objects, these objects must also be processed configurally (perhaps to a lesser degree than for faces), or that featural processing is also somewhat impaired by inversion.

1.3.2 Evidence for configural processing

Where, then, is the evidence that faces are processed configurally? The evidence comes predominantly from three experiment paradigms: the part/whole effect and the composite face effect (both of which investigate holistic configural processing), and second-order relation manipulations.

The part/whole effect refers to the observation that subjects are poorer at recognizing facial features in isolation than features presented within the context of the face (Tanaka and Farah, 1993). For example, Tanaka and Farah found that subjects were more accurate in a task that required them to discriminate two schematic faces that differed only in terms of the nose, than in a task that required them to discriminate two noses presented in isolation. Tanaka and Farah interpreted this result as showing that faces are processed

and learned holistically, so that the features, in isolation, are not recalled as readily as when presented in the context of the face. Further, when subjects are required to learn either features in isolation or whole faces, only those required to learn whole faces later exhibit inversion effects (for whole faces, at test); those trained (and tested) on isolated features did not exhibit inversion effects for these features (Farah et al., 1995). Farah et al. interpreted these results as indicating that inversion effects are not obtained for featural processing.

The composite face effect (Young et al., 1987) is tested by presenting subjects with two faces, each of which is the composite of two face-halves: the top of one face and the bottom of a different face. In one common version of this task, the bottom half always differs across the pair. Subjects are asked to judge whether the top halves are the same or different, when the faces are presented aligned (fused to appear to be a single face) or misaligned (so that the division between the face halves is clearly visible). Typically, in the "same" condition, where the face tops are identical, subjects are more likely to judge the face tops as different (presumably, influenced by the presence of the different bottoms) when the faces are aligned than when misaligned. This result is interpreted as implying that when faces are presented as a unit (with the top and bottom halves aligned), subjects are unable to avoid perceiving the face holistically, and therefore, incorrectly perceive the top halves as differing, even though only the bottom half differs. When misaligned, subjects are not forced to process the stimuli holistically, and can adequately perform the task. The strength of the composite face effect is interpreted as a measure of holistic processing.

In second-order relation spacing tasks, subjects are provided with a set of faces that share a common base-face and differ only in terms of either features or the spacing of those features. For example, in the Jane face task, subjects are presented with two versions of Jane and asked to identify whether the faces are the same or different (LeGrand et al., 2001; Maurer et al., 2002). In one condition, the examples of Jane differ only by their internal features (eyes, nose, mouth); In a separate condition, the examples of Jane differ only by the spacing between the eyes and/or the spacing between the nose and upper lip. Subjects are better at recognizing spacing changes when the faces are presented upright than when presented inverted, but do not differ in their accuracy at recognizing featural set changes across orientations (Freire et al., 2000; LeGrand et al., 2001; Maurer et al., 2002). This is interpreted as evidence that configural processing (recognition of spacing changes) is superior with upright faces, relative to inverted.

The part/whole effect, the composite face effect, and second-order spacing tasks each represent artificial contexts in which faces are easier to discriminate upright than inverted, and are assumed to identify the source of the better performance with upright faces across contexts. The logic, simply put, is as follows: (1) Face recognition suffers when inverted; (2) Detection of configural changes suffers when inverted; (3) Therefore, inverted face recognition suffers *because* it is dependent on configural processing, which is disrupted by inversion. Unfortunately, the face perception literature rarely recognizes that there is a logical fallacy in this reasoning: it is equally possible that inversion separately disrupts both face recognition and these configural processing tasks. In fact, there is little known about how performance in these configural tasks is related to other measures of face processing. What little work that has been done on this issue suggests that there is only a weak association between configural processing and face recognition. Konar et al. (VSS 2008), for example, found the magnitude of the composite face effect accounts for less than 10% of the variation in face identification accuracy among a large sample of young adults. If upright face processing is reliant on configural processing, a much stronger relation between the composite face effect and face identification would have been expected.

Thus, despite their extensive use in the face perception field, it is worth emphasizing that these configural processing tasks do not provide direct evidence that faces are processed configurally, let alone provide direct evidence that a deficit in configural processing is at the root of the face inversion effect. Nonetheless, these and other similar manipulations have had great weight in the field. This is troubling, in part because some counter evidence is beginning to emerge, questioning the assumption that upright and inverted faces are processed using qualitatively different mechanisms. For example, Sekuler et al. (2004) used classification images to map the portions of faces used in upright and inverted face discrimination tasks. Subjects made use of a small portion of the face around the eye/eyebrow region (contrary to holistic models of face perception), and most importantly, these regions did not differ significantly from upright to inverted face discrimination. Because the classification image analysis assumes that subjects used a linear strategy, Sekuler et al. (2004) also examined the extent of unexplained subject performance from their linear template model. The linear model accounted for nearly the entirety of subject performance. This result suggested that if other more complex modes of processing are engaged (such as configural processing), the contribution of these other forms of processing must be quite minimal, and thus, are unable to account for the substantial loss of performance caused by inversion. These results, amongst others

(Gaspar et al., 2008b; Martelli et al., 2005; Riesenhuber et al., 2004; Troje and Westhoff, 2006), caution against employing inversion as a method to tap into configural or featural processing. These results also suggest a more general caution: it is not yet clear that faces and objects are indeed processed by separate mechanisms.

The face literature has been largely focused on the featural/configural distinction. However, a few other important characteristics of face processing have been noted. For example, faces are processed using a spatial frequency filter that responds to 1-2 octave range of frequencies centered approximately on 10 cycles per face, tuned, surprisingly, to the object size rather than to the retinal size: faces are processed using a narrow band of information around 10 cycles per face (e.g., Nasanen, 1999; Gold et al., 1999a; Gaspar et al., 2008a). Object processing may also be spatial frequency specific. For example, letter identification relies on spatially tuned filters as well, though the tuning differs from that of faces (Gold et al., 1999b). Face recognition is also strongly reliant on information around the eye region: recognition of famous faces is disproportionately impaired by the removal of eyes and eyebrows, relative to other features (Sadr et al., 2003), and classification studies reveal that when distinguishing between a pair of faces across many trials, subjects rely almost entirely on this same small region around the eyes and eyebrows (Gold et al., 1999b; Sekuler et al., 2004; Gaspar et al., 2008b; Gosselin and Schyns, 2001; Schyns et al., 2002). These findings do not necessarily imply that faces are processed in a manner that is fundamentally different from other object classes, but they do, at minimum, suggest that observers make use of fairly consistent and reliable strategies when processing faces.

1.4 Neural mechanisms of face processing

Single-unit neural recordings in monkeys have revealed cells, in inferotemporal (e.g., Gross et al., 1972) and superior temporal cortex (e.g., Rolls and Baylis, 1986; Bruce et al., 1981), that respond preferentially to faces over other object classes. The responses of these face cells often remain invariant despite changes in low-level stimulus properties such as size and contrast (e.g., Rolls and Baylis, 1986). Although face cells often respond selectively to particular head viewpoints (Perrett et al., 1992), some cells respond to particular face identities across a wide range of viewpoints and photographic contexts (Quiroga et al., 2005).

1.4.1 Functional neuroimaging (fMRI)

Functional neuroimaging (fMRI) studies with humans have revealed several regions that respond more to faces than to other object classes or to scrambled versions of faces (Kanwisher et al., 1997; Puce et al., 1996). One region in particular, in the ventral temporal cortex has received more attention than most because of the consistency with which the area has been identified across experiments: the fusiform face area (FFA) (e.g. Kanwisher et al., 1997, 1999; Pourtois et al., 2005; Tong et al., 2000; O'Craven and Kanwisher, 2000; Grill-Spector et al., 2004). Although the FFA responds quite strongly to faces, it also responds to objects of other categories (e.g. Tong et al., 2000; Joseph and Gathers, 2002), leading to uncertainty about whether the FFA is truly a face selective region. Further, it is unclear whether the FFA is best thought of as an isolated region, or as part of a larger network because voxels that define regions of interest (ROIs), like the FFA, often have considerable overlap. In fact, the areas responsive to faces, chairs, and houses show a great deal of overlap, and are best separated by the relative patterns of activations across the voxels, rather than by differences in the voxels themselves (Ishai et al., 1999, 2000). The overlap between areas responsive to different object classes may imply that the object responsive regions of the brain are better understood as a collection of neurons that are responsive to a wide range of stimuli and object classes, but whose relative activations define the boundaries of object categories. Together, these studies imply that regions in the fusiform gyrus are involved in the processing of faces, but whether the FFA is an isolated face processing area or an integrated part of a larger network of visual object processing areas remains heavily debated.

1.4.2 Event-Related Potentials (ERPs)

Cortical markers of face processing in humans have also been explored using techniques that are more sensitive to the timing of neural events. Electroencephalography (EEG) records the activity at the cortex that results from the synchronized firing of large batches of neurons. EEG activity is often analysed by way of event related potentials (ERPs): By collecting EEG recordings across many trials and temporally aligning the trials to the stimulus onset, the activity across the trials can then be averaged together to reduce noise and reveal the signal related to the stimulus of interest. The resulting ERP should approximate an isolated signal produced by the stimulus alone.

Jeffreys and Tukmachi (1992) and Bentin et al. (1996) identified ERP signals associated with the presentation of faces, identifying patterns of ERP components that have

since been replicated and studied across many experimental contexts (e.g., Botzel et al., 1995; Itier and Taylor, 2004; Rousselet et al., 2005, 2007, 2008). Faces reliably produce a complex of ERP components. When recording from electrodes above the right and left occipital cortex, a large positivity (P1) arises about 100 ms after stimulus onset. This positivity is followed by a large negativity (N1) approximately 170 ms after stimulus onset, and a second positivity (P2) around 250 ms post stimulus. These components are observable after the presentation of a range of objects and scenes, but the N1 tends to be larger for faces than for other object classes (Jeffreys and Tukmachi, 1992; Bentin et al., 1996; Eimer, 2000a). The dynamics of the N1 response remain poorly understood. Various manipulations affect the size and timing of the resulting N1 response. For example, the N1 is increased by stimulus inversion (e.g., Itier et al., 2006; Rossion et al., 2002; Rousselet et al., 2008), and by the presentation of eyes in isolation (Bentin et al., 1996; Taylor et al., 2001), but is insensitive to face familiarity (Eimer, 2000b; Bentin and Deouell, 2000), implying that the N1 is more responsive to face structure than to individual face identity (although see Caharel et al., 2009). The N1 may be more generally involved in the processing of object structure where clear contours and boundary information is available, because the strength of these components is systematically reduced when object phase is scrambled (Guillaume et al., 2009).

The existence of face cells, the reliability of activity in the fusiform gyrus, and the unusual strength of the N1 to faces are routinely interpreted as evidence for special cortical mechanisms for face processing. However, these findings need not indicate that faces are processed differently from other objects. The fusiform face area may be one part of a larger object complex, and the strength of the N1 may be a matter of scale rather than indicative of a qualitative difference in processing. Similarly, the behavioural findings that faces have large inversion effects point to an unusual susceptibility of faces to the effect of distortion or modification from typical viewing conditions. Yet, there is again debate about whether these differences reflect differences in degree or differences in kind. Is face processing genuinely qualitatively different from the processing of objects, or only quantitatively so? These interesting questions continue to be debated within the face literature. Yet regardless of one's position on these matters, there is agreement that face processing possesses these unusual behavioural and neural signatures. Explaining why faces seem to be processed differently from other objects (be that by degree or by kind) remains an interesting question. To this end, studies in both infants and adults have been conducted to determine the extent to which the characteristics of face processing can be explained by either an innate face processing mechanism or by extensive experience with

faces across the lifespan.

1.5 Innateness and Expertise

Humans seem to be innately predisposed toward face processing. Even within hours of birth, infants preferentially orient toward faces, spending more time viewing faces than other control stimuli (e.g., Goren et al., 1975; Johnson et al., 1991). Neonatal infants also can discriminate faces: for example, newborns prefer to look at their mother's face over other faces (e.g., Pascalis et al., 1995), and more attractive faces over less attractive faces (Slater et al., 1998). The immediacy of these face preferences seems indicative of innately acquired face processing mechanisms, though some have argued that these results might be accounted for by cross-modal face learning, via tactile and proprioceptive inputs within the womb (Quinn and Slater, 2003). Regardless, learning begins to play a role in face processing early in life: young infants have the ability to discriminate both human and monkey faces, and faces across different human races, but gradually develop adult-like specificity for face discrimination within their own species (Pascalis et al., 2002) and race (Kelly et al., 2007).

Rather than trying to establish whether face processing itself is innate, expertise studies investigate whether non-face objects exhibit common behavioural and neural signatures of face processing after a sufficient degree of expertise has been acquired. These studies have made use of two main approaches: either studying object perception in real world experts (e.g., expert chess players, bird watchers, etc.), or studying how object perception is affected by expertise acquired over the course of several days of practice in laboratory experiments. For example, several studies have focused on whether large inversion effects are found only with faces or if they can be found with very familiar non-face stimuli. Diamond and Carey (1986), for example, measured recognition accuracy in a task that used upright and inverted pictures of dogs. For subjects who had expertise at identifying dogs, large inversion effects were obtained with pictures of the particular breeds of dogs with which the experts were familiar. Importantly, the size of this expertise-related inversion effect did not differ significantly from the inversion effect obtained with faces. Diamond and Carey interpreted these findings as evidence that faces are not unique in their sensitivity to inversion, and speculated that inversion effects might emerge any time expertise is obtained for a class of objects that share a common configuration. Consistent with this interpretation, large inversion effects have also been found for bird and car experts, both of whom show larger inversion effects for their own

class of expertise, than for the other non-expertly processed category (Gauthier et al., 2000), though Chapter 2 of this thesis will demonstrate that large inversion effects can also be obtained after training with textures that do not share a common configuration. Recently, Robbins and McKone (2007) reported an experiment that failed to replicate Diamond and Carey's results, leading the authors to speculate that Diamond and Carey's large inversion effect for dogs was caused by the experts being familiar with the particular stimuli used in that experiment, rather than their expertise *per se*. Even if Robbins and McKone's interpretation of Diamond and Carey's results is correct, it remains interesting that pictographic learning can result in inversion effects of the same magnitude as observed for faces.

The effects of expertise on physiological markers of face processing also have been explored. As described earlier, fMRI studies have identified a region of cortex that is reliably activated more strongly by faces than by other stimulus classes (Kanwisher et al., 1997). Subsequently, researchers proposed the possibility that this region might more generally code for within-category discriminations of expertly-processed object classes (Gauthier and Tarr, 1997). Support for this hypothesis has been mixed: An initial study examining real-world experts (bird experts, and car experts) found somewhat stronger FFA activity for the class of expertise (birds, for bird experts) (Gauthier et al., 2000). Attempts to replicate this study have been mixed (Xu, 2005; Rhodes et al., 2004; Grill-Spector et al., 2004), but those studies that failed to replicate (Rhodes et al., 2004; Grill-Spector et al., 2004) had serious design flaws: the experimental tasks were poorly matched to the expertise of the real-world experts, such that the experts did not outperform the novices in the experimental context. Similarly, training on a set of novel objects (Greebles) increased activity in the right FFA (Gauthier and Tarr, 1997), but training on a separate set of novel objects was unable to replicate these findings (Op de Beeck et al., 2006). Similar studies have been conducted to examine the role of expertise in EEG markers of face and object processing. For example, faces exhibit larger N1 responses for inverted than upright faces, an EEG inversion effect that is less prominent with other object classes (Rossion et al., 2000). Real-world fingerprint experts exhibit this N1 inversion effect for fingerprints (Busey and Vanderkolk, 2005), as do Greeble experts after in-lab training (Rossion et al., 2002). Overall N1 amplitude is larger for faces than for other objects (Bentin et al., 1996; Eimer, 2000a; Rossion et al., 2000), and both in lab training with cars and with owls has resulted in an amplitude increase for these object classes (Scott et al., 2006, 2008). Together, these findings tentatively suggest that some of the cortical signatures of face processing may be affected by expertise.

1.6 Summary and Conclusions

This chapter has provided a brief overview of some of the principal questions within the face perception literature. Specifically, the chapter identified some common behavioural measures (e.g., the inversion effect) that have been used to highlight differences between face perception and the perception of non-face objects; reviewed evidence for and against the existence of face-specific perceptual processing that has been offered as an explanation for these differences in perception; and outlined the ongoing attempts to determine whether face-specific perceptual processing can be explained by innate face mechanisms or experience-driven modification to more general object processing mechanisms.

The following chapters focus largely on the extent to which the characteristics of face processing can best be understood as the result of a lifetime of experience perceiving faces. Following in the footsteps of earlier research, I have approached this question indirectly, specifically by considering how controlled in-lab experience with non-face objects modifies the processing of these objects. By examining how experience shapes the processing of non-face objects, I can determine whether characteristics that are currently deemed face-specific can arise in other object classes in response to training. Specifically, the following chapters will consider: (1) whether the larger size of the face inversion effect is an indicator of specialized face processing mechanisms; (2) whether the cortical regions involved in the processing of faces and houses modify their relative activity after house discrimination training; and (3 & 4) how the early EEG components of face and object processing alter with systematic house discrimination training.

References

Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience 8 (6), 551–565.

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

Botzel, K., Schulze, S., Stodieck, S. R., 1995. Scalp topography and analysis of intracranial sources of face-evoked potentials. Exp Brain Res 104 (1), 135–143.

Bruce, C., Desimone, R., Gross, C. G., 1981. Visual properties of neurons in a polysensory

area in superior temporal sulcus of the macaque. J Neurophysiol 46 (2), 369–384.

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

Caharel, S., d'Arripe, O., Ramon, M., Jacques, C., Rossion, B., 2009. Early adaptation to repeated unfamiliar faces across viewpoint changes in the right hemisphere: evidence from the N170 ERP component. Neuropsychologia 47 (3), 639–643.

Diamond, R., Carey, S., 1986. Why faces are and are not special: an effect of expertise. J Exp Psychol Gen 115 (2), 107–117.

Eimer, M., 2000a. Attentional modulations of event-related brain potentials sensitive to faces. Cognitive Neuropsychology 17 (1-3), 103–116.

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

Farah, M. J., Tanaka, J. W., Drain, H. M., 1995. What causes the face inversion effect? J Exp Psychol Hum Percept Perform 21 (3), 628–634.

Freire, A., Lee, K., Symons, L. A., 2000. The face-inversion effect as a deficit in the encoding of configural information: direct evidence. Perception 29 (2), 159–170.

Gaspar, C., Sekuler, A. B., Bennett, P. J., 2008a. Spatial frequency tuning of upright and inverted face identification. Vision Res 48 (28), 2817–2826.

Gaspar, C. M., Bennett, P. J., Sekuler, A. B., 2008b. The effects of face inversion and contrast-reversal on efficiency and internal noise. Vision Res 48 (8), 1084–1095.

Gauthier, I., Skudlarski, P., Gore, J. C., Anderson, A. W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci 3 (2), 191–197.

Gauthier, I., Tarr, M. J., 1997. Becoming a "greeble" expert: exploring mechanisms for face recognition. Vision Res 37 (12), 1673–1682.

Gold, J., Bennett, P. J., Sekuler, A. B., 1999a. Identification of band-pass filtered letters and faces by human and ideal observers. Vision Res 39 (21), 3537–3560.

Gold, J., Bennett, P. J., Sekuler, A. B., 1999b. Signal but not noise changes with perceptual learning. Nature 402 (6758), 176–178.

Goren, C. C., Sarty, M., Wu, P. Y., 1975. Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics 56 (4), 544–549.

Gosselin, F., Schyns, P. G., 2001. Bubbles: a technique to reveal the use of information in recognition tasks. Vision Res 41 (17), 2261–2271.

Grill-Spector, K., Knouf, N., Kanwisher, N., 2004. The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci 7 (5), 555–562.

Gross, C. G., Rocha-Miranda, C. E., Bender, D. B., 1972. Visual properties of neurons in inferotemporal cortex of the macaque. J Neurophysiol 35 (1), 96–111.

Guillaume, C., Guillery-Girard, B., Chaby, L., Lebreton, K., Hugueville, L., Eustache, F., Fiori, N., 2009. The time course of repetition effects for familiar faces and objects: an ERP study. Brain Res 1248, 149–161.

Ishai, A., Ungerleider, L. G., Haxby, J. V., 2000. Distributed neural systems for the generation of visual images. Neuron 28 (3), 979–990.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., Haxby, J. V., 1999. Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci U S A 96 (16), 9379–9384.

Itier, R. J., Latinus, M., Taylor, M. J., 2006. Face, eye and object early processing: what is the face specificity? Neuroimage 29 (2), 667–676.

Itier, R. J., Taylor, M. J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cereb Cortex 14 (2), 132–142.

Jeffreys, D. A., Tukmachi, E. S., 1992. The vertex-positive scalp potential evoked by faces and by objects. Exp Brain Res 91 (2), 340–350.

Johnson, M., Dziurawiec, S., Ellis, H., Morton, J., 1991. Newborns preferential tracking of face-like stimuli and its subsequent decline. Cognition 40 (1-2), 1–19.

Jolicoeur, P., 1985. The time to name disoriented natural objects. Mem Cognit 13 (4), 289–303.

Joseph, J. E., Gathers, A. D., 2002. Natural and manufactured objects activate the fusiform face area. Neuroreport 13 (7), 935–938.

Kanwisher, N., McDermott, J., Chun, M. M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17 (11), 4302–4311.

Kanwisher, N., Stanley, D., Harris, A., 1999. The fusiform face area is selective for faces not animals. Neuroreport 10 (1), 183–187.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., Pascalis, O., 2007. The other-race effect develops during infancy - Evidence of perceptual narrowing. Psychological Science 18 (12), 1084–1089.

Konar, Y., Bennett, P. J., Sekuler, A. B., VSS 2008. The composite face effect is still not correlated with face identification accuracy. Journal of Vision 8 (6), 891a (Abstract).

LeGrand, R., Mondloch, C. J., Maurer, D., Brent, H. P., 2001. Neuroperception. early visual experience and face processing. Nature 410 (6831), 890.

Martelli, M., Majaj, N. J., Pelli, D. G., 2005. Are faces processed like words? A diagnostic test for recognition by parts. J Vis 5 (1), 58–70.

Maurer, D., Grand, R., Mondloch, C., 2002. The many faces of configural processing. Trends Cogn Sci 6 (6), 255–260.

Nasanen, R., 1999. Spatial frequency bandwidth used in the recognition of facial images. Vision Res 39 (23), 3824–3833.

O'Craven, K. M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. J Cogn Neurosci 12 (6), 1013–1023.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., Kanwisher, N. G., 2006. Discrimination training alters object representations in human extrastriate cortex. J Neurosci 26 (50), 13025–13036.

Pascalis, O., de Haan, M., Nelson, C., 2002. Is face processing species-specific during the first year of life? Science 296 (5571), 1321–1323.

Pascalis, O., Deschonen, S., Morton, J., Deruelle, C., Fabregrenet, M., 1995. Mothers face recognition by neonates - A replication and an extension. Infant Behavior & Development 18 (1), 79–85.

Perrett, D. I., Hietanen, J. K., Oram, M. W., Benson, P. J., 1992. Organization and functions of cells responsive to faces in the temporal cortex. Philos Trans R Soc Lond B Biol Sci 335 (1273), 23–30.

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., Vuilleumier, P., 2005. Portraits or people? Distinct representations of face identity in the human visual cortex. J Cogn Neurosci 17 (7), 1043–1057.

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. J Neurosci 16 (16), 5205–5215.

Quinn, P. C., Slater, A., 2003. Face perception at birth and beyond. The development of face processing in infancy and early childhood: Current perspectives, 3–12.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., Fried, I., 2005. Invariant visual representation by single neurons in the human brain. Nature 435 (7045), 1102–1107.

Rhodes, G., Byatt, G., Michie, P. T., Puce, A., 2004. Is the fusiform face area specialized for faces, individuation, or expert individuation? J Cogn Neurosci 16 (2), 189–203.

Riesenhuber, M., Jarudi, I., Gilad, S., Sinha, P., 2004. Face processing in humans is compatible with a simple shape-based model of vision. Proc Biol Sci 271 Suppl 6, S448–50.

Robbins, R., McKone, E., 2007. No face-like processing for objects-of-expertise in three behavioural tasks. Cognition 103 (1), 34–79.

Rolls, E. T., Baylis, G. C., 1986. Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. Exp Brain Res 65 (1), 38–48.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., Crommelinck, M., 2002. Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. Psychol Sci 13 (3), 250–257.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., Crommelinck, M., 2000. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. Neuroreport 11 (1), 69–74.

Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B., 2005. Spatial scaling factors explain eccentricity effects on face ERPs. J Vis 5 (10), 755–763.

Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B., 2007. Single-trial EEG dynamics of object and face visual processing. Neuroimage 36 (3), 843–862.

Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B., 2008. Time course and robustness of ERP object and face differences. J Vis 8 (12), 3.1–18.

Sadr, J., Jarudi, I., Sinha, P., 2003. The role of eyebrows in face recognition. Perception 32 (3), 285–293.

Schyns, P. G., Bonnar, L., Gosselin, F., 2002. Show me the features! Understanding recognition from the use of visual information. Psychol Sci 13 (5), 402–409.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., Curran, T., 2006. A reevaluation of the electrophysiological correlates of expert object processing. J Cogn Neurosci 18 (9), 1453–1465.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., Curran, T., 2008. The role of category learning in the acquisition and retention of perceptual expertise: a behavioral and neurophysiological study. Brain Res 1210, 204–215.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., Bennett, P. J., 2004. Inversion leads to quantitative, not qualitative, changes in face processing. Curr Biol 14 (5), 391–396.

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., Samuels, C., 1998. Newborn infants prefer attractive faces. Infant Behavior & Development 21 (2), 345–354.

Tanaka, J. W., Farah, M. J., 1993. Parts and wholes in face recognition. Q J Exp Psychol A 46 (2), 225–245.

Taylor, M. J., Edmonds, G. E., McCarthy, G., Allison, T., 2001. Eyes first! Eye processing develops before face processing in children. Neuroreport 12 (8), 1671–1676.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., Kanwisher, N., 2000. Response properties of the human fusiform face area. Cognitive Neuropsychology 17 (1-3), 257–279.

Troje, N. F., Westhoff, C., 2006. The inversion effect in biological motion perception: evidence for a "life detector"? Curr Biol 16 (8), 821–824.

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

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

Young, A. W., Hellawell, D., Hay, D. C., 1987. Configurational information in face perception. Perception 16 (6), 747–759.

Chapter 2

Inverting houses and textures: Investigating the characteristics of learned inversion effects

2.1 Abstract

Faces, more than other objects, are identified more accurately when upright than inverted. This inversion effect may be linked to differences in expertise. Here, we explore how stimulus characteristics and expertise interact to determine the magnitude of inversion effects. Observers were trained to identify houses or textures. Inversion effects were not found with either stimulus before training, but were found following 5 days of practice. Additionally, the learning-induced inversion effects showed partial transfer to novel exemplars. Although similar amounts of learning were observed with both types of stimuli, inversion effects were significantly larger for textures. Our results suggest that the size of the inversion effect is not a reliable index of face-specific processing.

Reference: Husk, J.S., Bennett, P.J., & Sekuler, A.B. (2007). Inverting houses and textures: investigating the characteristics of learned inversion effects. Vision Research, 47, 3350-3359.

2.2 Introduction

Objects are recognized more rapidly at their canonical orientations than when rotated within the picture plane (e.g., Jolicoeur, 1985), or in depth (Lawson et al., 2000). However, rotation (inversion) seems to impair face processing with particular severity, both in terms of accuracy and reaction time (e.g., Yin, 1969; Diamond and Carey, 1986, for a more extensive review, see Valentine (1988)). The impairment is so much more pronounced for faces than for other objects, that the inversion effect has become a hallmark of hypothesized face-specialized processing, particularly configural processing mechanisms (e.g., Tanaka and Farah, 1993; Farah et al., 1995; Moscovitch and Moscovitch, 2000; Leder and Bruce, 2000; Rhodes et al., 2004).

However, recent research suggests that the difference between upright and inverted face processing may be quantitative in nature, rather than qualitative. For example, studies using the classification image and bubbles techniques have shown that observers rely heavily on the eye and eyebrow region when identifying both upright (Gosselin and Schyns, 2001; Schyns et al., 2002; Gold et al., 2004; Sekuler et al., 2004) and inverted (Sekuler et al., 2004) faces, but the efficiency with which observers use available information in this region is reduced when faces are inverted (Sekuler et al., 2004; Gaspar et al., 2008). This difference in processing efficiency between upright and inverted faces mirrors the change in processing efficiency for objects seen as a result of practice (Gold et al., 1999b, 2004), suggesting that the superiority of upright processing may reflect greater practice identifying upright faces than inverted ones. Consistent with this idea, sizeable inversion effects have been observed for other non-face objects, when observers have developed expertise with that object class. For example, Diamond and Carey (1986) found that dog experts exhibited inversion effects when discriminating amongst breeds for which they had developed expertise (an effect that did not generalize to dogs in general), whereas novices did not perform significantly differently across orientations (but see Robbins and McKone, 2007). Similarly, inversion effects have been reported for body position discrimination (Reed et al., 2003), and inversion effects were larger when discriminating amongst bodies in biologically possible positions than in biologically impossible positions. Because observers likely have far more exposure to biologically possible positions than impossible ones, these results are consistent with the notion that inversion effects emerge for expertly-processed stimuli.

Furthermore, there is some evidence that practice can induce inversion effects. The most compelling evidence comes from (McLaren, 1997), who trained observers to discrim-

inate amongst checkerboard patterns, and demonstrated a strong inversion effect after practice: not only was upright performance greater for familiar than unfamiliar checkerboards, but inverted performance was actually impaired for familiar checkerboards relative to unfamiliar exemplars. Practice-induced inversion effects also have been reported for Greebles (a specially designed class of novel stimuli; Gauthier and Tarr, 1997). Observers who had been trained previously to recognize upright Greebles discriminated configural changes faster (though not more accurately) for upright stimuli than for inverted stimuli; observers who received no previous experience with Greebles did not differ in their performance across orientations. Similarly, in a separate task involving Greeble recognition (Gauthier et al., 1998), both novices and experts initially showed a small RT advantage for upright Greebles (relative to Greebles misoriented by 60, 120 & 180 deg). Recognition became faster with practice on the task for both experts and novices, but upright performance benefited disproportionately for experts, such that the inversion effect was enhanced for experts but lost for novices. Moreover, practice-induced inversion effects do not seem to be limited to the visual modality, because face perception and training of pattern discrimination in the tactile domain also can induce inversion effects (Newell et al., 2001; Behrmann and Ewell, 2003; Kilgour and Lederman, 2006). Taken together, these results show that inversion effects are present for expertly-processed stimuli, and can be induced through laboratory training tasks with novel stimuli.

However, many of the characteristics of practice-induced inversion effects remain largely unexplored. For example, it is not clear whether the size of trained inversion effects depends on prior knowledge brought to the task, such as knowledge about the canonical orientation of the object class. Further, the limited number of studies that have induced inversion effects through practice have not examined whether these inversion effects transfer to novel members of that class (a characteristic of face inversion effects). Finally, there is a suggestion within this body of research that the size of the inversion effect is a direct indicator of expertise (for faces or other highly trained object classes). Yin (1969), for example, emphasized the greater size of inversion effects for faces relative to other object classes, and studies of expertise generally have demonstrated larger inversion effects for experts than for novices (Diamond and Carey, 1986; McLaren, 1997; Gauthier et al., 1998; Behrmann and Ewell, 2003; Reed et al., 2003). The extent to which inversion effects differ across object sets with an equal extent of practice remains unknown.

The following experiments explore some of the characteristics of learned inversion effects by comparing face inversion effects to inversion effects generated before and after

practice on house and texture discrimination tasks. If knowledge of the canonical orientation is sufficient to induce an inversion effect, then houses, but not textures, should exhibit inversion effects prior to practice. Further, practice with upright houses might be expected to induce larger inversion effects than practice with inverted houses. If inversion effects that are induced by training are qualitatively similar to face inversion effects, then these inversion effects should, like faces, transfer to novel houses. Finally, if the size of the inversion effect is a direct indicator of expertise, then equivalent amounts of training on house and texture discrimination tasks should result in similarly sized inversion effects.

2.3 Experiment 1

Experiment 1 examined whether practice on a house discrimination task would be specific to the trained orientation. Different sets of observers were trained across eleven days to differentiate either amongst 10 upright houses or amongst 10 inverted houses, and both sets of observers subsequently were tested at both orientations.

2.3.1 Methods

2.3.1.1 Subjects

Twelve observers (mean age = 25.6 years; range: 19-45) were recruited from McMaster University's Vision and Cognitive Neuroscience Lab participant pool. Observers were undergraduate and graduate students at McMaster University, and received $10/hour for their participation. All observers had normal or corrected-to-normal visual acuity, and all were naive with respect to the purpose of the study.

2.3.1.2 Stimuli

Object classes differ in their degree of structural homogeneity. For example, faces are a highly homogenous stimulus category: the relative locations of eyes, nose and mouth are consistent across all exemplars. By contrast, houses are far more heterogeneous: the numbers and locations of doors and windows usually vary significantly across exemplars. Different strategies may well be needed to differentiate members of homogenous and heterogeneous classes, because the demands are likely to differ. For example, the most distinctive differences between the stimuli are more likely to be in a spatially predictable location in a homogenous class, than in a heterogeneous class. For these reasons, despite the typical heterogeneity of houses in the real world, the ten houses used in this exper-

iment (Figure 2.1) were constructed to tap into strategies used to learn to discriminate homogenous stimuli. This was accomplished by constructing all houses with a single shared template: Each house possessed a door to the lower left, a large window to the lower right, and two small windows to the upper left and right. Individual houses could be differentiated on the basis of any one of these components (as with the features in faces), but the house background was identical for all stimuli. The house background and the individual windows and doors were cropped (in Adobe Photoshop 7.0) from photographs of local houses, taken with a digital camera at a front-view angle. Finally, the spatial frequency content of the images was equated by applying the average amplitude spectrum to each house, ensuring that the stimuli differed only in terms of their global phase spectra. Because the structure of images is carried largely by the global phase spectrum (Oppenheim and Lim, 1981; Sekuler and Bennett, 1996), this manipulation preserves the discriminability of the stimuli while ensuring that observers could not rely upon differences in overall contrast, or relative contrast differences across spatial frequencies and orientations, to perform the task. Despite the structural homogeneity of the house stimuli, these stimuli display typical EEG and fMRI markers of house processing: the N170 component of the event-related potential has reduced amplitude for these houses relative to faces (Rousselet et al., 2005, 2007), and maximal BOLD responses are located in the parahippocampal place area (Husk et al., VSS 2006).

All stimuli were displayed on a Sony Trinitron GDM-F520 monitor (800 x 600 pixels, 21 pixels/cm, refresh rate 85 Hz) and viewed from a distance of 1 meter. The target houses (subtending 5 x 7 degrees) were presented centrally on an otherwise uniform screen. On the selection screen, the 10 houses (the target plus nine distractors) were presented in two rows above and below fixation, each house subtending 3.4 x 4.8 degrees. The position of each individual house in the selection screen was held constant across all trials, and was the same for all observers. Throughout the experiment, all stimuli were presented at a fixed contrast variance of 0.0015. Background luminance equaled 15.85 cd/m2 and was held constant throughout the experiment.

2.3.1.3 Procedure

Observers completed eleven sessions on consecutive weekdays (with most sessions separated by 24 hrs, but some as much as 72 hrs when crossing a weekend). The first session consisted of a pre-test followed by the first practice session. Sessions 2-10 each contained a single practice session, and session 11 consisted solely of a post-test. The general procedure employed during all phases of the experiment was a 10-alternative

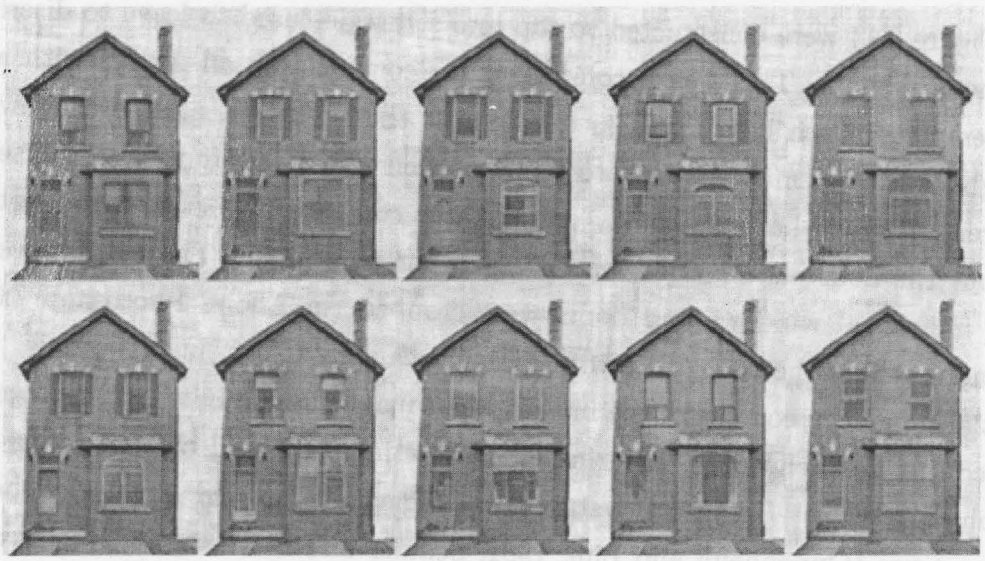

Figure 2.1: House set employed in Experiment 1. These images are displayed in the same order as presented on the selection screen. However, during actual presentation the images were of lower contrast, and the size and spacing of the images were greater than displayed here.

forced choice discrimination task. Each trial began with a fixation point, presented for 1 second, after which a house was presented for 500 ms. The selection screen, consisting of an array of 10 houses, was presented immediately afterward, and remained on the screen until the observer made a response. The observer responded by using the computer mouse to select the target house. Auditory feedback was provided during all phases of the experiment (pre-test, training, and post-test): a high-pitch tone indicated a correct selection, and a low-pitch tone, an incorrect selection. The next trial was automatically started immediately following the auditory feedback.

Pre-test To obtain an initial comparison of upright and inverted discrimination performance, observers partook in a pre-test composed of 200 trials (100 trials each of upright and inverted stimuli) separated into 10 blocks of 20 trials. The upright and inverted blocks were alternated across the session, with the order counterbalanced across observers. All houses were presented equally often in each block, and at each orientation across blocks.

Practice Observers then were trained across 10 sessions to discriminate amongst the 10 houses (half were trained on upright houses, and half on inverted houses). For those trained in the inverted condition, the target and distractors on the selection screen were also inverted. Each practice session consisted of 400 trials (40 presentations with each 10 houses serving as a target, randomly intermixed).

Post-test Observers were re-tested on upright and inverted performance. The post-test procedure was the same as that of the pre-test, and observers maintained the same order of exposure to upright and inverted blocks as in the pre-test.

2.3.2 Results

In the pre-test, observers assigned to the inverted practice condition outperformed those assigned to the upright practice condition (Figure 2.2a). Despite this difference, neither group showed an initial inversion effect: performance with upright and inverted stimuli did not differ either in the upright practice group ($t(5) = 0.15, p = .89$), or in the inverted practice group ($t(5) = 0.48, p = .65$). As seen in Figure 2.3, although performance in both groups was significantly greater after training than before (main effect of Session: $F(1, 10) = 118.95, p < .001$), the improvement in the upright group was somewhat greater, so that the difference in performance between the groups was smaller by the tenth day of practice than it was on the first day (Group x Session interaction: $F(1, 10) = 6.01, p < .05$). Following practice with upright houses, a significant inversion effect was observed (Figure 2.2b): upright houses were identified with 14% greater accuracy than inverted houses ($t(5) = 8.37, p < .001$). Practice with inverted houses induced a reversed inversion effect of approximately the same magnitude: inverted houses were identified with 13% greater accuracy than upright houses ($t(5) = -3.62, p < .02$; Figure 2.2b).

Because the upright and inverted groups performed differently at pre-test, Figure 2.2c represents the performance on upright and inverted houses before and after learning in terms of improvement (i.e., the change from baseline performance), illustrating the striking cross-over Group x Orientation interaction ($F(1, 10) = 14.73, p < .01$): Performance was better with the trained orientation, regardless of whether observers were trained with upright or inverted stimuli. It is important to note, though, that performance improved substantially from baseline for both the trained and untrained orientations. Thus, there seems to be partial, but incomplete, transfer of learning across orientations. Practice im-

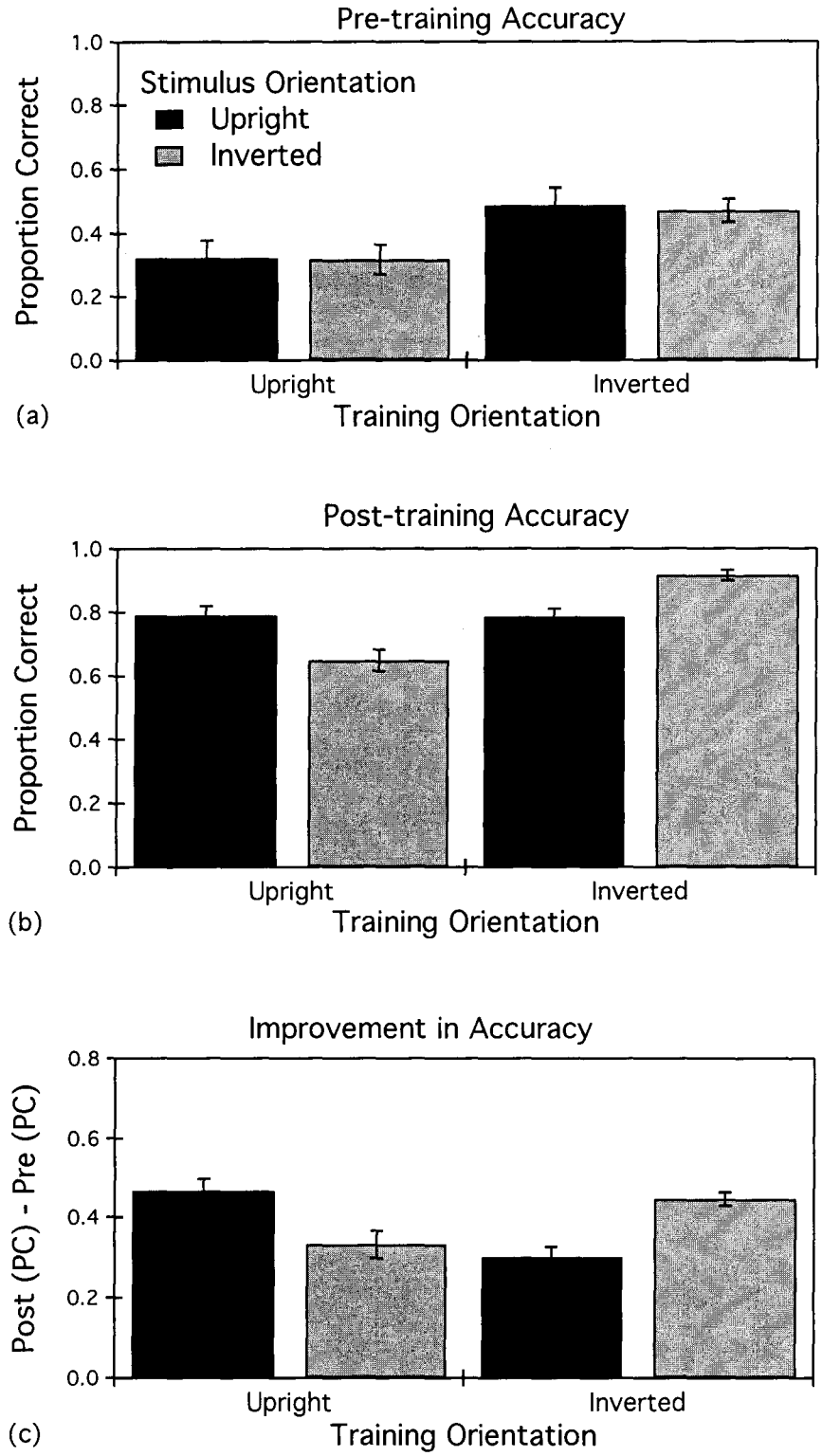

Figure 2.2: Results of Experiment 1. Error bars represent +/- 1 standard error of the mean. (a) Pre-test (b) post-test, and (c) learning-related improvement in accuracy on upright and inverted house discrimination for observers trained on upright or inverted house discrimination.

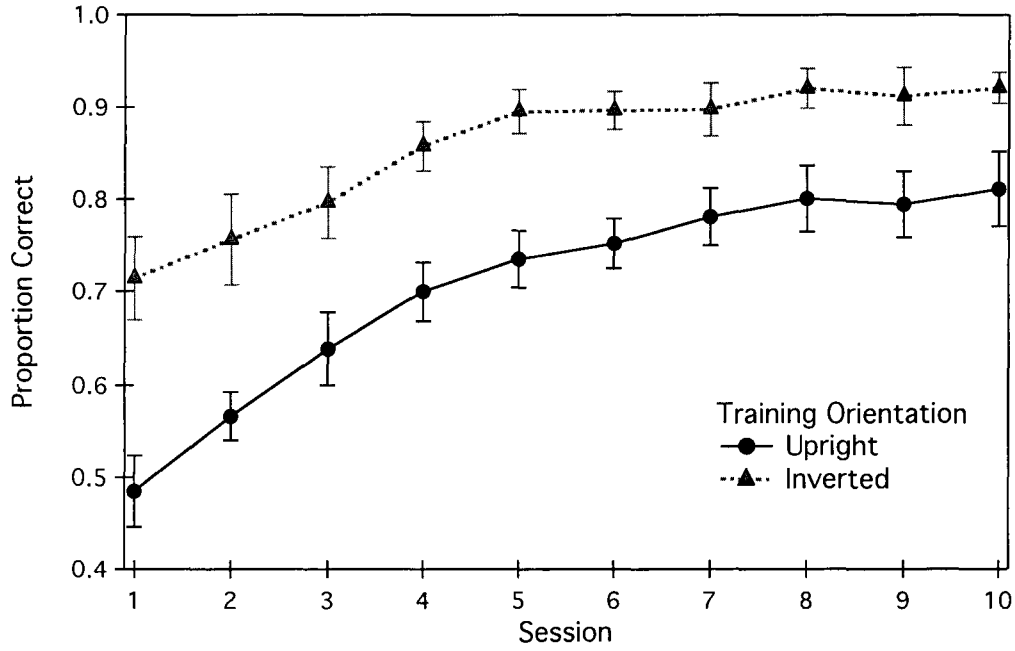

Figure 2.3: Learning curves for observers trained on upright or inverted houses in Experiment 1. Error bars represent +/- 1 standard error of the mean.

proved performance in both groups. Hence, it is possible that the inversion effects that were found after practice were associated with a change in task difficulty (i.e., response accuracy), rather than practice per se. We evaluated this possibility by testing eight novice observers with upright and inverted houses at five different levels of contrast variance (0.001, 0.0015, 0.003, 0.01, 0.02). By varying stimulus contrast, it was possible to manipulate task difficulty independently of the amount of practice. Observers completed a single session consisting of 5 blocks of 80 trials. Each block presented houses at a different contrast level, with order counterbalanced across observers. The results are shown in Figure 2.4: Response accuracy increased monotonically with contrast and, averaged across contrasts, was 3.4% higher for upright than inverted houses. A repeated-measures ANOVA found significant main effects of contrast ($F(4, 28) = 34.39, p < .001$) and stimulus orientation ($F(1, 7) = 8.81, p < .05$). Importantly, however, the interaction between contrast and orientation was not significant ($F(4, 28) = 0.65, p = .56$), so the effect of stimulus orientation did not depend on contrast or on overall accuracy. These results suggest that the change in the inversion effect produced by practice was not simply due to a change in task difficulty, but rather reflects an effect of practice per se.

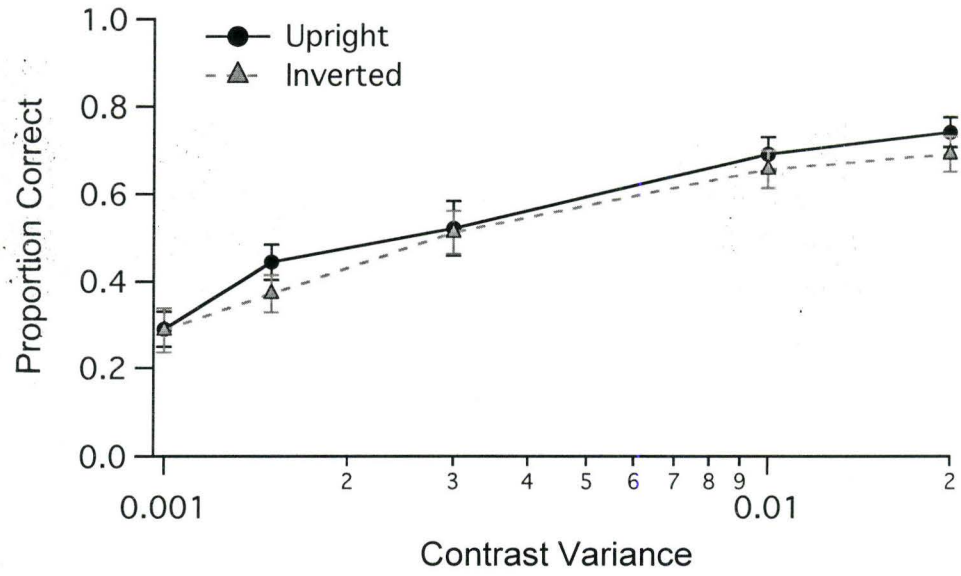

Figure 2.4: Accuracy on a single session of upright and inverted house discrimination as a function of contrast variance (plotted on a logarithmic scale). Error bars represent +/- 1 standard error of the mean. In the absence of training, the inversion effect size does not vary with accuracy.

2.4 Experiment 2

Face inversion effects are notable for their generalization to novel stimuli. Indeed, most studies examining the face inversion effect use face stimuli that are unfamiliar to the participants, yet face inversion effects are pronounced despite the unfamiliarity of the individual faces. Experiment 1 demonstrated that practice with houses at a given orientation can produce performance that, like faces, is impaired at novel orientations, but it is not clear whether such inversion effects generalize to novel exemplars. If so, this would suggest that observers have not only learned to recognize the individual exemplars, but have also learned characteristics about the set of stimuli that can then be applied to novel exemplars. Further, comparing the degree of impairment due to changes in orientation versus changes in exemplars can provide insight into the relative extents of stimulus-specific versus category-specific learning.

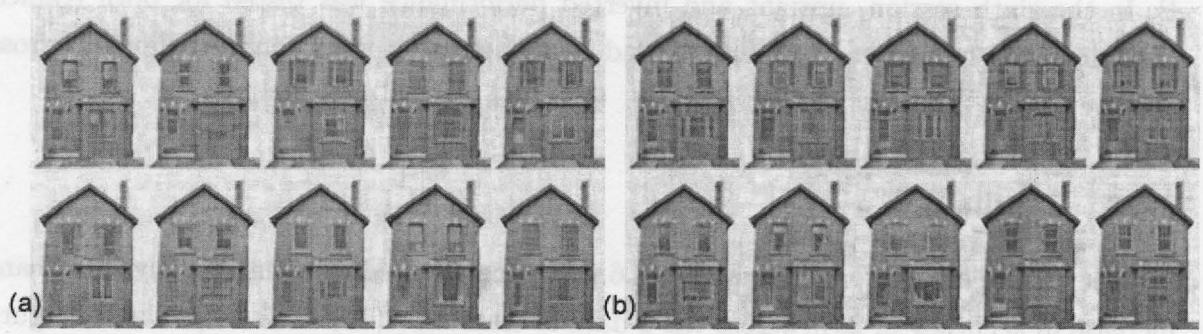

Figure 2.5: House-sets A and B used in Experiment 2. Within each house set, the items are displayed in the same order as presented on the selection screen. However, during actual presentation the images were of lower contrast, and the size and spacing of the images were greater than displayed here.

2.4.1 Methods

2.4.1.1 Subjects

Sixteen new observers (mean age = 19.4 years; range: 18-22) were recruited from the McMaster Undergraduate Psychology participant pool (and received partial course credit for their participation) or from the Vision and Cognitive Neuroscience Lab participant pool (and received $10/hour for their participation). All observers had normal or corrected-to-normal vision, and all were naive with respect to the purpose of the study.

2.4.1.2 Stimuli

Two sets of 10 houses were employed in this experiment. Sets A and B were each composed of five houses from Experiment 1 plus an additional five novel houses (Figure 2.5b). As with Experiment 1, all houses had the same amplitude spectrum.

2.4.1.3 Procedure

Observers were trained to discriminate amongst upright presentations of ten houses using the same training procedure described in Experiment 1. To ensure that observers had no prior experience with inverted stimuli prior to testing, the pre-test was eliminated (although preliminary data on 20 additional observers confirmed that no initial inversion effect was present at the training contrast for either house set: see "pre-test" group in Figure 2.6a. Half the observers were trained for five sessions on Set A, and half on Set

B. In the sixth session, upright and inverted performance was tested for both trained and untrained house sets (orientation and house-set order were counterbalanced across observers).

2.4.2 Results

The results are presented in Figure 2.6. Practice resulted in a statistically significant increase in accuracy of 29% across the 5 days of training ($t(15) = 15.35, p < .001$, Figure 2.6b). As seen in Figure 2.6a, after practice, accuracy was greater for trained than untrained houses, ($F(1, 15) = 43.38, p < .001$). Consistent with the earlier experiments, a significant inversion effect was generated after practice, ($F(1, 15) = 59.80, p < .001$). Moreover, the size of the inversion effect did not differ from the inversion effect obtained in Experiment 1 after 10 practice sessions with upright houses ($t(14.33) = 1.81, p = 0.092$). Inversion effects were observed for both trained ($t(15) = 6.113, p < .001$) and untrained ($t(15) = 2.66, p < .02$) houses, but the size of the inversion effect was marginally greater for trained (10% greater accuracy for upright than inverted) than untrained (4% greater accuracy for upright than inverted) houses, ($F(1, 15) = 3.96, p = .065$). The superior performance for upright novel houses, compared to inverted novel houses, suggests that the advantage for upright houses is, at least in part, due to the learning of factors that are common to both sets of houses, and not specific to the individual learned houses. However, the somewhat greater magnitude of the inversion effect for the trained houses suggests some orientation-specific learning at the level of the individual houses, as well as at the level of the category. Note that the inverted familiar houses were also novel in the sense that observers had never before viewed those specific stimuli in that orientation. Nevertheless, performance on inverted familiar houses was superior to that of upright novel houses ($t(15) = 3.536, p < .01$), suggesting that learning transfers more to the familiar exemplars in novel orientations than it does to novel exemplars in familiar orientations.

2.5 Experiment 3

If expertise is the determining factor of inversion effect size, we should expect similarly sized inversion effects across stimulus sets, as long as observers are trained to the same extent. On the other hand, if the effect of inversion depends on stimulus structure, we should find that inversion effect size can vary across stimulus sets, even with equal training. To examine this issue, Experiment 3 replicates the house-training procedure

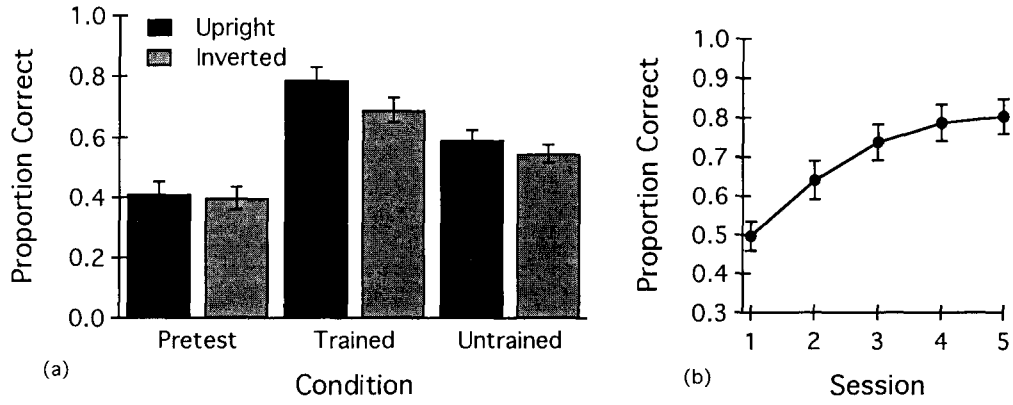

Figure 2.6: Results of Experiment 2. (a) The Pre-test condition shows the average proportion correct for upright and inverted houses for a group of 20 observers. There was no significant effect of orientation. The Trained condition shows the average proportion correct for a second group of 16 observers using the same stimuli that they had viewed during five days of practice in an identification task using upright houses. The Untrained condition shows post-practice performance in the same 16 subjects using stimuli that they had not viewed during practice. (b) Learning curve for the 16 observers who received practice with one of two sets of upright houses. Error bars represent +/- 1 standard error of the mean.

with a different non-face stimulus: texture patches. To compare both texture- and house-training results to inversion effects typically observed for face stimuli, Experiment 3 also tested a separate set of observers (without prior practice) on upright and inverted face discrimination.

2.5.1 Methods

2.5.1.1 Subjects

Twenty-six new observers (mean age = 21.1 years; range: 17-26) were recruited from the Vision and Cognitive Neuroscience Lab participant pool and received $10/hour for their participation. Eight observers were tested in a single session of upright and inverted texture discrimination to obtain a pre-test measure of the texture inversion effect. An additional eight observers were trained on upright texture discrimination for 5 days then tested on upright and inverted textures. The final ten observers participated in the face discrimination task. All observers had normal or corrected-to-normal vision, and all were naive with respect to the purpose of the study.

2.5.1.2 Stimuli and Procedure

Observers in the texture training task were trained for five days on a set of 10 textures (Figure 2.7a). Because the textures have no canonical orientation, the training orientation was assigned arbitrarily. At post-test, textures were presented both upright and inverted (relative to the training orientation). To ensure that there was no initial advantage for either orientation, a separate group of 8 observers performed the"pre-test" on a single session of upright and inverted textures. The pre-test, training, and post-test followed the same procedures used in Experiment 1, with the exception that the texture pre-testing was conducted on separate observers. The textures were constructed by generating 10 Gaussian white noise patterns, then band-pass filtering each stimulus to pass only 2-4 cycles per image (the construction of these stimuli is described in more detail elsewhere: Gold et al. (1999b,a, 2004)). At the viewing distance of 100 cm, target texture stimuli subtended 5 deg x 5 deg, and stimuli in the selection array each subtended 3.5 deg x 3.5 deg. Despite the apparent difficulty of identifying individual stimuli when one first views the stimulus set, pilot testing indicated that this task was actually slightly easier than the house identification task. Therefore, the contrast variance of the textures was reduced to 0.0005 to approximately equate the pre-practice performance on the textures to that previously obtained for houses. A separate set of observers performed a face discrimination task, in which performance was tested on upright and inverted faces in a single session without prior training. The procedure was identical to that used in the pre- and post-tests of Experiment 1. The faces (Figure 2.7b; see Gold et al. (1999b,a), for more details about the construction of the face stimuli) subtended 5.2 deg x 5.2 deg in the target screen, and 3.5 deg x 3.5 deg in the selection screen. The contrast variance of the faces was equated to that used in Experiments 1 and 2.

2.5.2 Results

Practice on the texture discrimination task resulted in a statistically significant improvement in accuracy of 35% for upright textures on the final day of testing, compared to the initial day ($t(7) = 8.67, p < .001$, Figure 2.7e), which is similar to the amount of improvement (29%) observed for the equivalent amount of training on houses in Experiment 2. Pre-testing (with separate observers) indicated no initial significant difference in discriminating amongst upright and inverted texture patches, ($t(7) = .56, p = .59$). However, after practice, a large inversion effect emerged, ($t(7) = 7.18, p < .001$). Upright performance was 28% greater than inverted performance (Figure 2.7d). In fact,

inverted performance after practice did not differ significantly from upright performance on the initial day of practice, ($t(7) = -1.44, p = .19$), indicating that, unlike what was found in Experiment 2 with houses, there was no significant transfer of learning across orientations. The inversion effects obtained with textures, faces, and houses were submitted to a one-way, between-subjects ANOVA. The effect of stimulus type was significant ($F(2, 31) = 28.58, p < 0.001, w2 = 0.618$). Differences among the three sets of inversion effects were evaluated by using Tukey's HSD to construct 95% confidence intervals: The difference between faces and houses (C.95 = 0.35,0.17) and textures and houses (C.95 = 0.28,0.09) were both significant, but the difference between faces and textures (C.95 = 0.18, -0.03) was not.

The results of practicing with textures suggest that the size of the inversion effect is neither a good indicator of the extent to which stimulus processing is "face-like," nor of the familiarity with the stimuli. Subjects received equal amounts of practice with textures and houses, and the amount of learning, as measured by changes in response accuracy, did not differ significantly for observers trained on houses ($M = 29.2, SEM = 1.9$) and textures ($M = 33.9, SEM = 3.2$, $t(22) = 1.36, p = .19$). Nevertheless, the size of the inversion effect was larger for textures than for houses (Figure 2.7c). Although overall accuracy was greater on the texture discrimination task than on the house task, this difference is unlikely to account for the difference in the inversion effects because the contrast control study in Experiment 1 established that there was no relation between overall accuracy and the size of the inversion effect for houses. Moreover, the magnitude of the inversion effect did not differ between those trained on upright and inverted houses in Experiment 1, despite the overall accuracy difference between these groups. Thus, the current findings suggest that similar practice effects can induce different inversion effects with different classes of stimuli. After just five days of practice, the inversion effect obtained with textures did not differ from the inversion effect obtained with faces, and both of these inversion effects were significantly larger than the one obtained after practice with houses. Note that there is no a priori reason to expect that learned textures are processed more configurally than learned houses, and response classification results suggest that observers rely primarily on local features for texture discrimination (Gold et al., 2004; Nagai et al., 2007), thus the size of the inversion effect may not be a reliable index of configural processing. Rather, we suggest that the magnitude of an inversion effect indicates the efficiency with which a stimulus is processed in a preferred orientation (Riesenhuber et al., 2004; Sekuler et al., 2004; Martelli et al., 2005; Troje and Westhoff, 2006; Gaspar et al., 2008).

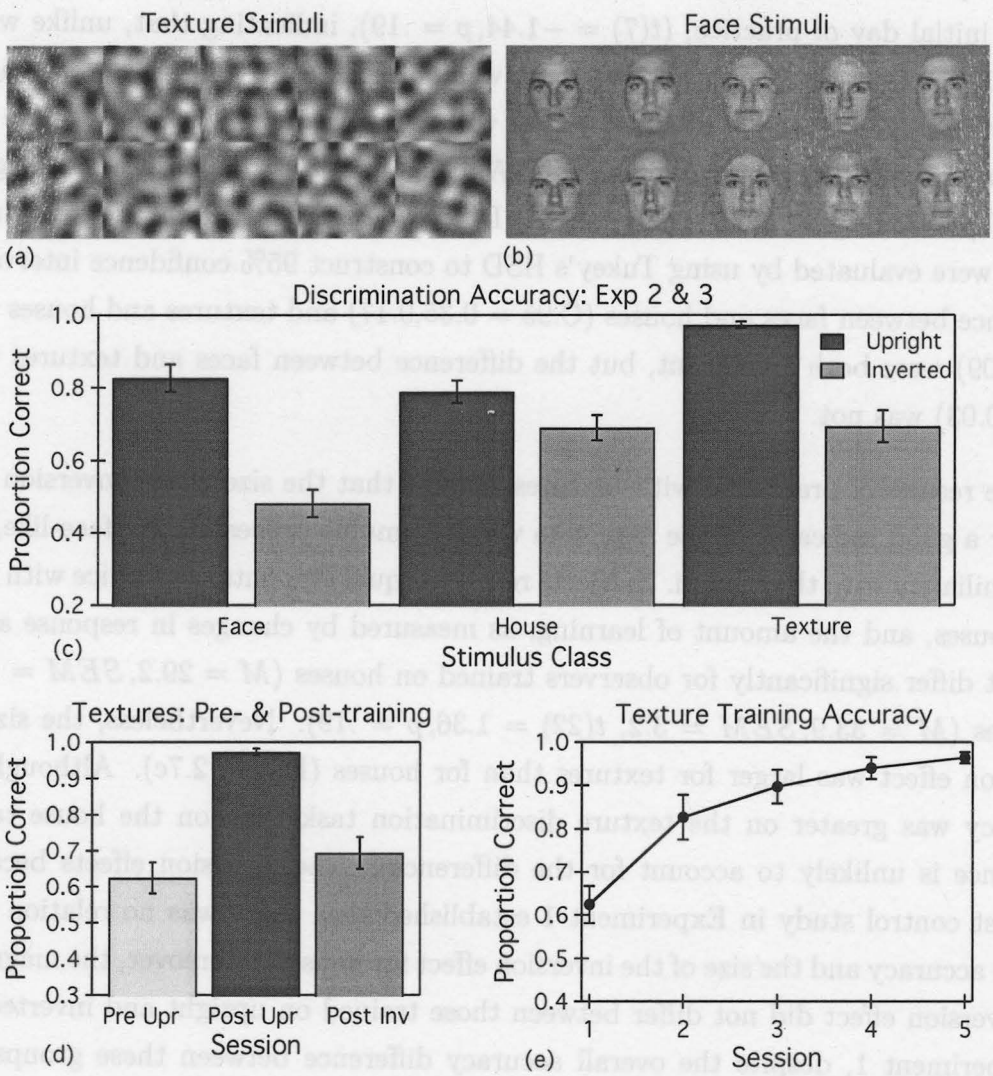

Figure 2.7: Results of Experiment 3. Texture (a) and face (b) sets employed in Experiment 3. Both sets are displayed in the same order as presented on the selection screen. However, during actual presentation the images were of lower contrast, and the size and spacing of the images were greater than displayed here. (c) Comparison of inversion effects for untrained faces (Exp3), trained houses (Exp2), and trained textures (Exp3). The data shown for houses and textures was obtained after observers had received five days of practice with upright stimuli. (d) Accuracy on texture discrimination before and after practice on upright texture discrimination (Exp3). Pre-training results are presented for upright textures only, because trained observers were not exposed to inverted textures until the post-test. Post-training results were obtained for both upright and inverted textures. (e) Learning curve for observers trained on upright texture discrimination (Exp3). Error bars represent +/- 1 standard error of the mean.

2.6 General Discussion

In accord with other studies that have generated inversion effects through training (e.g., Diamond and Carey, 1986; Gauthier et al., 1998; McLaren, 1997; Behrmann and Ewell, 2003; Reed et al., 2003), we have demonstrated that inversion effects can be generated with relatively small amounts of practice. Training on house and texture discrimination tasks for as little as five days was sufficient to induce substantial differences in identification accuracy for upright and inverted stimuli. Furthermore, the inversion effect induced after five days of training with upright houses in Experiment 2 (10%) did not differ significantly from the inversion effect obtained after 10 practice sessions in Experiment 1 (14%). Thus, although training induced an inversion effect, the cumulative effect of training appeared to diminish after approximately five days. This result does not imply, however, that an existing inversion effect cannot be altered by practice: Hussain et al. (VSS 2006) showed that the face inversion effect obtained with a particular set of faces does increase significantly with practice, and that the effect of practice did not generalize to a novel set of faces.

2.6.1 Inversion effects and configuration

The face inversion effect often is attributed to differences in the kind of information that observers use to discriminate amongst upright and inverted stimuli. Although individual faces vary along many potential dimensions, much of the face literature has come to dichotomize these dimensions into two overarching sets of changes: featural changes (based on the shape or appearance of the nameable face parts), and configural changes (based on variations in the spatial arrangement of nameable parts within the face). Several studies report that observers are more accurate at identifying upright faces that differ in terms of the spatial arrangement of features, but that accuracy does not differ across stimulus orientation when faces differ in terms of the individual parts (e.g., Freire et al., 2000; LeGrand et al., 2001). These and similar findings are frequently regarded as a demonstration that configural information strongly influences the perception of upright, but not inverted, faces (e.g., Tanaka and Farah, 1993; Farah et al., 1995; Moscovitch and Moscovitch, 2000; Leder and Bruce, 2000; Rhodes et al., 2004). However, the evidence supporting a qualitative, featural vs. configural, distinction between processing of upright and inverted faces is decidedly mixed (for reviews, see Rakover, 2002; Valentine, 1988). Furthermore, not all studies have found larger inversion effects for configural than parts-based manipulations (e.g., Yovel and Kanwisher, 2004), nor is the size of the in-

version effect associated with configural manipulation consistent across the face (Barton et al., 2001; Malcolm et al., 2004; Rutherford et al., 2007). Classification image studies (Sekuler et al., 2004) have not found qualitative differences in the information used to discriminate upright and inverted faces. Finally, the difference in inversion effect size for featural versus configural sets can be eliminated by intermixing the two trial types (Riesenhuber et al., 2004). For these reasons, we suggest that the role of configural information in generating inversion effects in faces and other objects remains an open question.

2.6.2 Inversion effects and canonical orientation

Familiar objects often have a canonical orientation. That is to say, we often have prior knowledge about an object class that permits us to identify whether a new exemplar of that class is upright or inverted. It is reasonable to suppose that having knowledge of an object's canonical orientation may lead to a processing advantage when encountering objects in their upright (or most familiar) orientation (Palmer et al., 1981). However, the results of this study indicate this is not always the case. Faces and houses both have canonical orientations; textures do not. Yet, whereas faces resulted in an inversion effect with novel exemplars (i.e., without any specific training for these particular faces), neither houses nor textures displayed an initial upright advantage for novel exemplars. The absence of an initial inversion effect for houses is not specific to the house stimuli used in the current experiments, as a similar finding has recently been reported for other house stimuli (Leder and Carbon, 2006). Having a life-time of prior exposure to upright houses apparently does not give observers an advantage when processing upright, as opposed to inverted, houses, so knowledge of canonical orientation alone does not seem to result in inversion effects. Clearly, something more specific to the individual exemplars or to the object set must be learnt before inversion effects arise.

2.6.3 Face specificity of inversion effects

In the current study, the inversion effect varied substantially between houses and textures, but very little between textures and faces. Differences between the size of inversion effects obtained with faces and other objects have often been attributed to the effects of face-specific processing. Where substantial inversion effects have been demonstrated for non-face stimuli (e.g., Greebles or dogs), critics have suggested that these results may indicate that these stimuli are too "face-like" (e.g., Kanwisher, 2000),

and therefore, might come to be processed by face mechanisms after training. The current results do not support the claim that inversion effects are the result of a switch from face-specific to object-general processing. House and texture training not only induced an inversion effect, but the size of this effect for textures was large, and did not differ significantly from the effect measured with faces. Because textures bear no resemblance to faces, it is clear that large inversion effects are not restricted to stimuli that look like faces. Moreover, quantitative changes within a linear template can account for the learning of these textures (Gold et al., 2004), suggesting that practice increases the efficiency of a single processing strategy, rather than leading to a qualitative switch from featural to configural processing. Accordingly, the large inversion effect typically obtained with faces may not reflect a qualitative change in the types of mechanisms used to identify upright and inverted stimuli (Sekuler et al., 2004; Ikeda et al., 2005; Troje and Westhoff, 2006; Gaspar et al., 2008).

2.6.4 Conclusions

Large inversion effects are neither specific to faces nor are they likely the result of face-specific mechanisms, as they can be induced with non-face stimuli, even when no prior canonical orientation is known. Some minimal level of expertise is necessary for inversion effects to emerge, yet there does not appear to be a simple relationship between degree of training and the size of the inversion effect.

2.7 Acknowledgements

This work was supported by NSERC Discovery Grants 42133 and 105494, Canada Research Chairs to PJB and ABS, and an NSERC PGS-D award to JSH. We thank Donna Waxman for assistance with the data collection, and Carl Gaspar and Daphne Maurer for their helpful comments on the manuscript.

References

Barton, J., Keenan, J., Bass, T., 2001. Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. Br J Psychol 92 Part 3, 527–549.

Behrmann, M., Ewell, C., 2003. Expertise in tactile pattern recognition. Psychol Sci 14 (5), 480–486.

Diamond, R., Carey, S., 1986. Why faces are and are not special: an effect of expertise. J Exp Psychol Gen 115 (2), 107–117.

Farah, M. J., Tanaka, J. W., Drain, H. M., 1995. What causes the face inversion effect? J Exp Psychol Hum Percept Perform 21 (3), 628–634.

Freire, A., Lee, K., Symons, L. A., 2000. The face-inversion effect as a deficit in the encoding of configural information: direct evidence. Perception 29 (2), 159–170.

Gaspar, C. M., Bennett, P. J., Sekuler, A. B., 2008. The effects of face inversion and contrast-reversal on efficiency and internal noise. Vision Res 48 (8), 1084–1095.

Gauthier, I., Tarr, M. J., 1997. Becoming a "greeble" expert: exploring mechanisms for face recognition. Vision Res 37 (12), 1673–1682.

Gauthier, I., Williams, P., Tarr, M. J., Tanaka, J., 1998. Training 'greeble' experts: a framework for studying expert object recognition processes. Vision Res 38 (15-16), 2401–2428.

Gold, J., Bennett, P. J., Sekuler, A. B., 1999a. Identification of band-pass filtered letters and faces by human and ideal observers. Vision Res 39 (21), 3537–3560.

Gold, J., Bennett, P. J., Sekuler, A. B., 1999b. Signal but not noise changes with perceptual learning. Nature 402 (6758), 176–178.

Gold, J., Sekuler, A., Bennett, P., 2004. Characterizing perceptual learning with external noise. Cognitive Science 28 (2), 167–207.

Gosselin, F., Schyns, P. G., 2001. Bubbles: a technique to reveal the use of information in recognition tasks. Vision Res 41 (17), 2261–2271.

Husk, J. S., Betts, L. R., O'Craven, K. M., Bennett, P. J., Sekuler, A. B., VSS 2006. House training: Neural correlates of object learning. Journal of Vision 6 (6), 664a (Abstract).

Hussain, Z., Bennett, P. J., Sekuler, A. B., VSS 2006. Face-inversion effects flex with perceptual learning. Journal of Vision 6 (6), 153a (Abstract).

Ikeda, H., Blake, R., Watanabe, K., 2005. Eccentric perception of biological motion is unscalably poor. Vision Res 45 (15), 1935–1943.

Jolicoeur, P., 1985. The time to name disoriented natural objects. Mem Cognit 13 (4), 289–303.

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

Kilgour, A. R., Lederman, S. J., 2006. A haptic face-inversion effect. Perception 35 (7), 921–931.

Lawson, R., Humphreys, G. W., Jolicoeur, P., 2000. The combined effects of plane disorientation and foreshortening on picture naming: one manipulation or two? J Exp Psychol Hum Percept Perform 26 (2), 568–581.

Leder, H., Bruce, V., 2000. When inverted faces are recognized: the role of configural information in face recognition. Q J Exp Psychol A 53 (2), 513–536.

Leder, H., Carbon, C.-C., 2006. Face-specific configural processing of relational information. Br J Psychol 97 (Pt 1), 19–29.

LeGrand, R., Mondloch, C. J., Maurer, D., Brent, H. P., 2001. Neuroperception. Early visual experience and face processing. Nature 410 (6831), 890.

Malcolm, G. L., Leung, C., Barton, J. J. S., 2004. Regional variation in the inversion effect for faces: differential effects for feature shape, feature configuration, and external contour. Perception 33 (10), 1221–1231.

Martelli, M., Majaj, N. J., Pelli, D. G., 2005. Are faces processed like words? A diagnostic test for recognition by parts. J Vis 5 (1), 58–70.

McLaren, I. P., 1997. Categorization and perceptual learning: an analogue of the face inversion effect. Q J Exp Psychol A 50 (2), 257–273.

Moscovitch, M., Moscovitch, D., 2000. Super face-inversion effects for isolated internal or external features, and for fractured faces. Cognitive Neuropsychology 17 (1-3), 201–219.

Nagai, M., Bennett, P. J., Sekuler, A. B., 2007. Spatiotemporal templates for detecting orientation-defined targets. J Vis 7 (8), 11.

Newell, F. N., Ernst, M. O., Tjan, B. S., Bulthoff, H. H., 2001. Viewpoint dependence in visual and haptic object recognition. Psychol Sci 12 (1), 37–42.

Oppenheim, A., Lim, J., 1981. The importance of phase in signals. Proceedings of the IEEE 69 (5), 529–541.

Palmer, S. E., Rosch, E., Chase, P., 1981. Canonical perspective and the perception of objects. Attention and Performance 9, 135–151.

Rakover, S. S., 2002. Featural vs. configurational information in faces: a conceptual and empirical analysis. Br J Psychol 93 (Pt 1), 1–30.

Reed, C. L., Stone, V. E., Bozova, S., Tanaka, J., 2003. The body-inversion effect. Psychol Sci 14 (4), 302–308.

Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., Clifford, C. W. G., 2004. Orientation-contingent face aftereffects and implications for face-coding mechanisms. Curr Biol 14 (23), 2119–2123.

Riesenhuber, M., Jarudi, I., Gilad, S., Sinha, P., 2004. Face processing in humans is compatible with a simple shape-based model of vision. Proc Biol Sci 271 Suppl 6, S448–50.

Robbins, R., McKone, E., 2007. No face-like processing for objects-of-expertise in three behavioural tasks. Cognition 103 (1), 34–79.

Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B., 2005. Spatial scaling factors explain eccentricity effects on face ERPs. J Vis 5 (10), 755–763.

Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B., 2007. Single-trial EEG dynamics of object and face visual processing. Neuroimage 36 (3), 843–862.

Rutherford, M. D., Clements, K. A., Sekuler, A. B., 2007. Differences in discrimination of eye and mouth displacement in autism spectrum disorders. Vision Res 47 (15), 2099–2110.

Schyns, P. G., Bonnar, L., Gosselin, F., 2002. Show me the features! Understanding recognition from the use of visual information. Psychol Sci 13 (5), 402–409.

Sekuler, A. B., Bennett, P. J., 1996. Spatial phase differences can drive apparent motion. Percept Psychophys 58 (2), 174–190.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., Bennett, P. J., 2004. Inversion leads to quantitative, not qualitative, changes in face processing. Curr Biol 14 (5), 391–396.

Tanaka, J. W., Farah, M. J., 1993. Parts and wholes in face recognition. Q J Exp Psychol A 46 (2), 225–245.

Troje, N. F., Westhoff, C., 2006. The inversion effect in biological motion perception: evidence for a "life detector"? Curr Biol 16 (8), 821–824.

Valentine, T., 1988. Upside-down faces: a review of the effect of inversion upon face recognition. Br J Psychol 79 (4), 471–491.

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

Yovel, G., Kanwisher, N., 2004. Face perception: domain specific, not process specific. Neuron 44 (5), 889–898.

Chapter 3

House learning alters neural signatures of house processing, not FFA

3.1 Abstract

The cortical mechanisms of object learning are not well understood. Several studies have examined the fusiform face area (FFA) as a potential expertise area with mixed results (e.g., Gauthier et al., 1999; Op de Beeck et al., 2006), while other studies have found training-induced changes in the lateral occipital gyrus, a generic object processing region (e.g., Grill-Spector et al., 2000; Kourtzi et al., 2005). However, there has been little focus on the effect of learning within areas that, before learning, respond preferentially to the trained object class. Given that training can alter the tuning properties of task-responsive neurons (Schoups et al., 2001), the strongest effects of learning might be expected in regions that respond preferentially to the trained object class prior to training. The current fMRI experiment focuses on houses, which preferentially activate three cortical regions (e.g., Epstein et al., 2007a): the parahippocampal place area (PPA), the retrosplenial cortex (RSC), and the temporal occipital sulcus (TOS). Houses also elicit strong behavioural perceptual learning (Husk et al., 2007). BOLD responses to house stimuli were measured during a 1-back matching task before and after 5 days of practice in a house identification task. No significant differences were observed in 1-back activity levels across sessions, nor were there any significant correlations between performance on the training task and changes in activity across sessions. However, evidence

of learning-related changes was observed in association with learning in the 1-back task itself: Behavioural improvement in the 1-back matching task correlated significantly with increases in PPA and RSC, but not TOS, activity. No such correlations were observed for activity within FFA or any retinotopically-defined visual areas. These results suggest that object discrimination training modifies activity within existing object-selective cortical networks without recruiting additional regions.

3.2 Introduction

The impact of visual experience on the organization of human cortical object regions remains poorly understood. Studies that have examined this question often focus on the middle fusiform gyrus, particularly the portion known as the Fusiform Face Area, or FFA. The FFA is reliably activated more strongly by faces than by many other stimuli, leading to the suggestion that this region is a specialized face-processing module (e.g., Kanwisher et al., 1997; Tong et al., 2000). However, the FFA is also activated by non-face objects, including houses, chairs, and a wide variety of other object classes (Haxby et al., 2001; Ishai et al., 1999; Joseph and Gathers, 2002), though not as strongly as by faces. Explanations for the differences in activation within the FFA remain varied, with some researchers favouring the original view of the FFA as a face module (Kanwisher, 2000), others proposing a model of overlapping distributed regions of object processing (Haxby et al., 2001), and still others suggesting that the fusiform gyrus is responsible for the discrimination of expertly-processed object classes (Gauthier et al., 1999). This last view rests on the idea that, unlike non-face objects, humans have a lifetime of experience with faces, and therefore are face experts. According to this theory, the apparent preference of the FFA for faces may be better understood as a more general preference for expertly-processed object classes. To test this idea, recent investigations of the effects of learning on visual cortical mechanisms have focused primarily on changes that occur in the fusiform gyrus.

Investigations of the role of the FFA in expert object processing have produced conflicting results. Training subjects on novel objects (Greebles) seems to modify FFA activity (Gauthier et al., 1999). However, Greebles have been criticized for their potentially "animate" or "face-like" quality (Kanwisher, 2000), and more importantly, were associated with quite subtle changes in FFA activity, detectable only when analyses were limited to a small, central subsection of the right FFA. Bird and car experts also exhibit expertise-related differences in right FFA activity (Gauthier et al., 2000; Xu, 2005), but

these learning effects also were small. More recent studies with Lepidoptera (butterfly and moth) experts (Rhodes et al., 2004) and car experts (Grill-Spector et al., 2004) have both failed to show increased FFA activity for experts relative to novices. However, these findings are difficult to interpret because experts failed to outperform novices on the behavioural tasks, indicating that these studies may not have tapped into the expertise of their participants. More convincingly, Op de Beeck et al. (2006) found no changes in activations of the right FFA after extensive training with novel, synthetic object classes.

Relatively few studies have examined the effects of object learning on regions beyond the FFA. Kourtzi et al. (2005) found that perceptual learning of contours defined by collinear Gabor patches induced learning-specific changes not only in the posterior fusiform region, but also in other sub-regions of the object-selective lateral occipital complex (LOC), such as LO, and in early retinotopic visual areas. These results suggest that a distributed network of regions in visual cortex mediates the perceptual learning of shape information. In another study, Op de Beeck et al. (2006) trained participants to discriminate between exemplars of synthetic object classes: Learning produced increased BOLD activation in the LOC, but no learning-related changes were evident in the FFA. Together, these two studies provide evidence of cortical plasticity in the visual system after extensive training in regions distinctly separate from the face-selective regions in the fusiform gyrus. Rather than recruiting new cortical areas, practice with a particular object category may alter the responses of neurons within cortical networks that are already involved in the processing of these objects. This hypothesis is consistent with evidence that, in primary visual cortex, orientation discrimination learning alters the orientation tuning of neurons already responsive to the task (Schoups et al., 2001). It also is consistent with psychophysical investigations showing that perceptual learning does not qualitatively change the nature of object processing, but instead results in a quantitative enhancement of the efficiency with which objects are processed (Gold et al., 1999b, 2004). Learning effects in the brain may therefore emerge as a modification of existing mechanisms, rather than the introduction of processing in new modular regions, as suggested by the FFA expertise hypothesis. Surprisingly, this hypothesis has not been examined carefully in neuroimaging studies with non-face objects that preferentially activate cortical areas outside of the FFA.

Houses, as a stimulus class, are ideal for addressing this issue. There is a solid evidence that at least three regions consistently respond more to houses and scenes than to faces and other objects: the parahippocampal place area (PPA), the retrosplenial cortex (RSC), and the temporal occipital sulcus (TOS) (Aguirre et al., 1998; Epstein et al., 1999;

Epstein and Kanwisher, 1998; Epstein et al., 2007a,b; O'Craven and Kanwisher, 2000; Grill-Spector, 2003). Before and after five days of training on a house discrimination task, we collected functional responses to the trained houses and to two control object sets (an untrained house set and a face set). Functional activity was examined within the house-responsive regions (PPA, RSC, and TOS), as well as within the FFA and early visual areas. To complement our ROI analysis, we also used a multivariate analysis technique, Partial Least Squares (McIntosh et al., 1996; McIntosh and Lobaugh, 2004; Bennett et al., 2001), to investigate learning effects across the whole brain. If object learning is driven primarily by changes within those areas most involved in processing the object class in question in the absence of learning, we should expect learning-related changes to be most pronounced in regions that respond preferentially to houses, with lesser changes expected in the fusiform gyrus, where houses elicit weaker responses. In contrast, if the FFA generally is recruited when object classes become learned, we would expect increased activation in the FFA following learning, even if that area was not preferentially activated by the object class before learning.

3.3 Method

3.3.1 Participants

Fourteen observers (mean age: 28 years; range: 18-36 years), recruited from McMaster University, completed a screening form to ensure that they met the necessary safety conditions (e.g., no metal implants) required for exposure to the strong magnetic field in the scanning room. Where necessary, vision was corrected by fitting observers with MRI-safe plastic frames and lenses: Snellen near-acuity ranged from 20/12.5 to 20/25 after correction (viewing distance in the scanner was 30 cm, in near-acuity range). All observers were paid $25 per scanning session and $10/hr for training sessions. The data from two observers were excluded due to excessive head motion (see Procedure for more detail). Of the remaining 12 observers, 8 were female (7 right-handed) and 4 were male (3 right-handed). Twelve additional observers (mean age:19; range: 18-23 years) were assigned to a control group. Snellen near-acuity ranged from 20/12 to 20/16, and far-acuity ranged from 20/10 to 20/20. All observers were paid $10/hr. Eight were female (all right-handed) and 4 were male (3 right-handed).

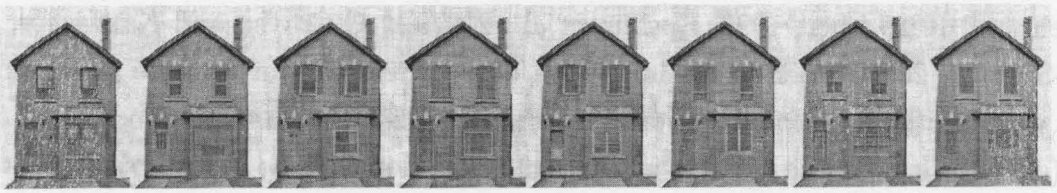

Figure 3.1: Training house set

Figure 3.2: Example stimuli. From left to right: Experimental stimuli (Trained houses, untrained houses, and faces), and Localizer stimuli (Faces, houses, and objects).

3.3.2 Stimuli

3.3.2.1 Behavioural training stimuli

The behavioural training stimuli consisted of a set of eight houses that were derived from front-view photographs of houses (Figure 5.1), modified to include the same class of basic internal features (two upper windows, a lower right-hand window, and a lower left-hand door), with individual exemplars differing in the identity of each of the internal features. A more detailed description of the stimulus generation procedure is given by Husk et al. (2007). The house stimuli were presented on a Sony Trinitron GDM-F520 monitor (800 x 600 pixels, 21 pixels/cm). At the viewing distance of 100 cm, the target houses subtended a visual angle of 5 deg x 7 deg, and the selection-screen houses subtended a visual angle of 3.4 deg x 4.8 deg. Mean luminance was $54.4cd/m2$, and images were presented with a contrast variance of 0.001. The stimulus presentation was controlled by a Macintosh G4 computer running Matlab v5.2, with software from the Psychophysics and Video Toolboxes (Brainard, 1997; Pelli, 1997). Responses were recorded on a standard QWERTY Macintosh keyboard.

3.3.2.2 Pre-testing and Post-testing stimuli

The experimental stimuli consisted of three sets of eight images: two house sets and one face set (example stimuli are presented in Figure 5.2). The first experimental house set was the same set employed during the training task. The second experimental house set was a different set of exemplars constructed in the same manner as the first experimental house set. The experimental faces were grayscale front-view photographs of faces with hair and external contours of the faces removed by placing each face within an oval mask, as originally described by Gold et al. (1999a). The amplitude spectra of the experimental faces and houses were equated separately within their object classes. This process ensures that observers cannot use global spatial frequency differences to identify individual objects. Because the structure of an image is carried largely by phase information (Oppenheim and Lim, 1981; Sekuler and Bennett, 1996), observers can still easily categorize and discriminate stimuli with averaged amplitude spectra.

Three additional sets of eight images (houses, faces, and objects; see Figure 5.2 for example stimuli) were used in ROI analyses to localize regions preferentially active for faces or houses. The localizer stimuli consisted of grayscale faces complete with hair and external contours, houses placed within their local scenic contexts, and objects isolated on a uniform background. The localizer face and house images were originally used by O'Craven and colleagues (e.g., O'Craven et al., 1999), and the objects were drawn from a database provided by Michael J. Tarr (Brown University, http://ww.tarrlab.org/). All images (experimental and localizer stimuli) were equated for average contrast.

Visual presentation of stimuli and collection of behavioural responses within the scanner were controlled by the MRIX Technologies Synchronization Control System (Thulborn Associates, Inc.). The visual stimuli were rendered as movie files, displayed through Windows Media Player (v9.0), and projected through the MRIx system to a front-projection screen mounted on a hood that was placed over the head coil within the magnet bore. Observers viewed the screen by a mirror mounted inside the hood. The effective viewing distance, the sum of the distance from the viewer's eye to the mirror and the distance from the mirror to the hood screen, was approximately 30 cm, but varied slightly according to the size of the participant's head. At this viewing distance, the visual angle of the stimuli subtended approximately 10.7 deg.

3.3.3 Procedure

3.3.3.1 Overview

The experimental procedure was designed to test whether neural processing would be altered after an intensive period of visual house discrimination training. The experiment consisted of seven sessions: a pre-testing session in the fMRI scanner, five training sessions outside the fMRI scanner, and a final post-testing session in the fMRI scanner. The procedures employed during the pre-test and post-test sessions were the same: Observers completed six functional scans in which they were asked to perform a 1-back matching task with each of the object categories shown in Figure 5.2. All six object-categories were present in each functional scan, as further detailed below. In the pre-training session only, observers also completed an additional functional scan designed to map the retinotopic organization of the cortex (e.g., Engel et al., 1997). Observers completed a high-resolution anatomical scan at the end of each session.

We chose to use a 1-back matching task during pre- and post-testing because this task has been used regularly when functionally identifying ventral temporal regions such as the FFA and PPA. Similar sequential matching tasks have been employed in earlier expertise studies (e.g., Gauthier et al., 1999, 2000). However, because the 1-back task differed from the training task, we also collected behavioural data for the training task during the pre- and post-testing sessions to ensure that the performance improvements on the training task were also present in the context of scanning sessions. These data were collected in a shortened version (104 trials) of the house discrimination training task that was performed during the anatomical scan. This shortened training task was designed only to obtain a behavioural measure of performance in the context of the scanning session; no functional data were acquired during this scan.

3.3.3.2 Behavioural Training Procedure

Observers were trained across five sessions. When possible, sessions were scheduled on consecutive days, and all training sessions were completed within a 10-day window. At the start of each session, observers were adapted to the lighting conditions of the testing room for 90 seconds. The training task was composed of 400 trials. Each trial began with a fixation point, presented for 1 s, followed by a house (randomly selected from the set of 8 possible houses) for 500 ms. Four alternatives were then presented in an array on the screen: the correct item and three other randomly selected items from the

remaining seven houses. The observer pressed one of four keys to identify the presented stimulus. The selection process was self-paced, and the stimulus array remained on the screen until the observer made a response. Once a house was selected, auditory feedback was provided: a high-pitched (600 Hz) tone in response to a correct selection, and a low-pitched (200 Hz) tone in response to an incorrect selection. Immediately after the presentation of the tone, the next trial began. Response accuracy was recorded as the proportion of correct responses.

All observers received training on the same house set (Fig. 5.1), and were exposed to the second (untrained) house set (Fig. 5.2 only during the functional imaging sessions. Previous experiments (Husk et al., 2007), where training was counterbalanced across these two sets, revealed no significant differences in the discrimination difficulty of the two experimental house sets either before ($t(14) = 1.57, p = .14$) or after ($t(14) = 1.06, p = .31$) training, and no significant differences in the amount of learning across these two house sets ($t(14) = 0.47, p = .65$).

3.3.3.3 Functional Imaging Procedure

MRI data were acquired on a 3T short bore GE Excite-HD equipped with a customized 8-channel head coil. One-back functional scans consisted of anywhere from 33 to 37 axial slices (4 mm slice thickness, 3.75 x 3.75 in-plane resolution), depending on the number of slices required to achieve full-brain coverage. We used T2* weighted gradient echo EPI scans (TE = 35 ms; TR = 3 s; flip angle = 90 deg; FOV = 24 cm; sequential acquisition; zero gap). The retinotopic scan consisted of 22 slices positioned to cover the entire occipital and most of the temporal lobes (TE = 22 ms; TR = 2 s; the other parameters were the same as those used in the 1-back functional scans). High-resolution (0.5 x 0.5 x 0.8 mm) 3D anatomical images were acquired in the axial plane using a FastIR preparation, SPGR whole-brain anatomical scan (Zip512; T1 weighted; flip angle = 12 deg; FOV = 24 cm; TE = 2.1 ms).

3.3.3.4 1-back Functional scans

The experimental and localizer stimuli were intermixed in a block randomized order within each of 6 functional runs. Figure 3.3 illustrates an example run. Each single run was composed of 12 epochs (2 epochs per object category). These epochs were blocked into 2-pseudo-randomized sequences of 6 epochs each, with a long fixation period of 24 s marking the beginning, middle, and end of the run. Epochs consisted of 24 trials, each of

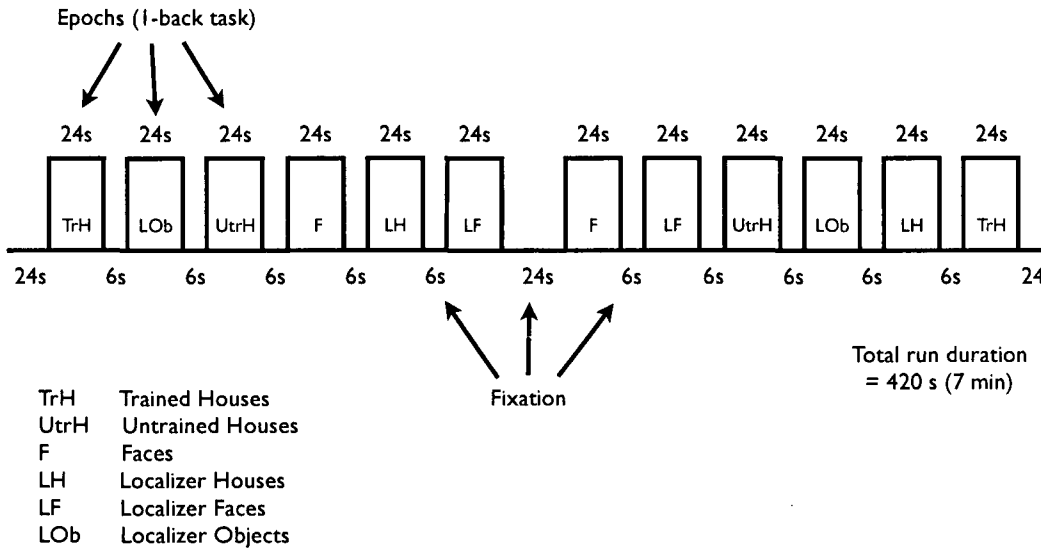

Figure 3.3: Example time-course of a single 1-back functional run. Task epochs (24 s each) were separated by 6 s fixation blocks (with longer 24 s fixation blocks at the start, end and middle of the run). Each epoch consisted of 24 trials of a 1-back task for a single object category. Each object category was presented twice during each run (once in the first half, and once in the second half of the run). Localizer epochs and experimental epochs were randomly intermixed within the run.

which was composed of 500 ms of fixation, followed by 500 ms of stimulus presentation. The total scan time for a complete run was seven minutes. Observers were required to perform a 1-back matching task to ensure that observers remained alert and attentive throughout the scanning session, and to ensure that stimuli were processed at the level of individual identity. Immediate stimulus repetitions occurred on one-quarter of the trials in each epoch.

3.3.3.5 Retinotopic Functional scans

All subjects completed a retinotopic scan (duration 520 s) during the pre-test scanning session. The scan consisted of a wedge stimulus (angular width: 45 deg; radius: 6.5 deg) that rotated around a central fixation point (diameter: 0.46 deg). The wedge changed location every 8 seconds, and covered a complete 360 rotation across eight 45 degree steps (0, 45, 90, etc.). Within the contour of the wedges, high-contrast greyscale images were presented. The image presentation alternated between a static condition and a dynamic condition, and each condition lasted for a complete wedge rotation. In the static condition, the presented images were fragments of artworks (by Escher, Brooker, Dali, Kandinsky, Vasarely, Muche, Klimt, and Klee) that changed once per second, cycling through all eight artwork fragments at each wedge position. In the dynamic condition, the images were plaids, random noise, polar gratings, and random gabor patches that cycled through six phase positions (3 times per second) to induce apparent motion. All four image-types were presented at each wedge location (2 seconds for each motion sequence).

Subjects were instructed to maintain fixation on the central fixation point. To encourage compliance, subjects were given a central distractor task: The centre of the fixation point contained a coloured segment (diameter: 0.3 deg) that flickered on and off at a rate of 1 Hz. This coloured segment was red 95% of the time, and green 5% of the time. Subjects were required to indicate whenever the coloured segment turned green with a button press.

3.3.3.6 Region of Interest Analysis

The experimental scans were imported into Brain Voyager 4.0. In-plane anatomical reference files were created from the first four volumes of each functional scan. Temporal low-frequency drifts were removed through high-pass filtering (3 cycles per time-course). Slice-scan time correction and 3D motion correction processing were also performed. Fol-

lowing a motion analysis, data runs that were contaminated by excessive motion (motion spikes $> 1mm$) were removed. When more than half the runs were contaminated in either the first or second fMRI session, the observer's data were removed entirely from the study. These criteria resulted in the exclusion of two observers. The anatomical images were transformed into standardized Talairach coordinates, and the position coordinates of the anatomical and functional data were co-registered.

The primary regions of interest (PPA, RSC, TOS, and FFA) were defined by applying a general linear model analysis contrasting the activity associated with houses to that associated with faces and objects (PPA, RSC, and TOS), and the activity associated with faces to that associated with houses and objects (FFA) with a p-value threshold of $8.51x10 - 7$. These contrasts resulted in several patches of activity across the cortex. Consistent with earlier studies, the FFA (mean Talairach coordinates: $38, -50, -18$ (R. hem.); $-35, -47, -18$ (L. hem.)) was identified as a region of contiguous voxels located in the fusiform gyrus. The PPA (mean Talairach coordinates: $25, -44, -10$ (R. hem.); $-26, -46, -9$ (L. hem.)), RSC (mean Talairach coordinates: $17, -52, 12$ (R. hem.); $-19, -58, 11$ (L. hem.)) and TOS (mean Talairach coordinates: $31, -78, 13$ (R. hem.); $-30, -80, 9$ (L. hem.)) were similarly identified as regions of contiguous voxels whose Talairach coordinates coincided with previous reports of these regions. For some subjects, house-related activity in the parahippocampal and neighbouring gyri merged. To maintain a consistent definition of the PPA across subjects, the activity in the neighbouring gyrus was excluded from the PPA definition, despite the contiguity of the voxel activations.

The time courses of activity for the experimental houses and faces were extracted within the predefined regions. Activity at each time point was calculated as the percent signal change for each condition relative to baseline (24 s blocks of fixation at the beginning, middle and end of each run). Baseline corrections were calculated individually for each voxel within the region of interest, and the percent signal change was averaged across voxels, time points (24 s blocks), and runs for each condition in all ROIs.

3.3.3.7 Retinotopic Analysis

Retinotopic mapping was carried out on an inflated map of the cortex (e.g., Engel et al., 1997; Sereno et al., 1995). To prepare for inflation, white matter was distinguished from grey matter in the anatomical images using Brain Voyager's automatic segmentation procedure, along with manual corrections where visual inspection indicated errors in the

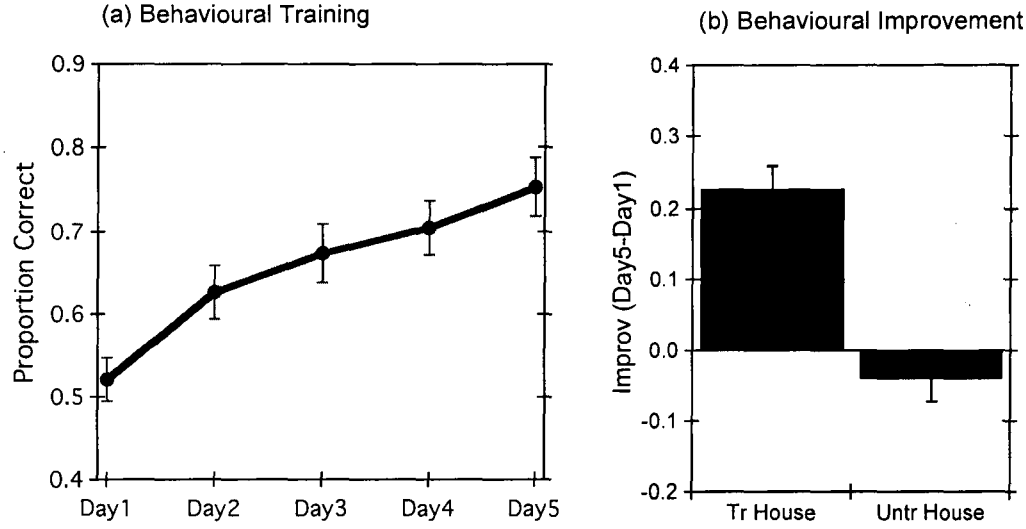

Figure 3.4: (a) Mean accuracy for 12 subjects as a function of learning across 5 days of discrimination training. (b) Transfer data from a subset of 6 subjects who were recalled for an additional experimental session approximately one week after the fifth day of training. Shown is improvement in accuracy for the trained and untrained house-sets relative to accuracy with the trained house set on Day 1.

automated solution. Functional activation was correlated with the position of the rotating wedge stimulus. These correlations were overlaid on the inflated cortex, and the borders between adjacent visual areas were manually defined by observing the phase reversals of the BOLD activation.

3.3.3.8 Partial Least Squares Analysis

The Talairach-transformed functional and structural data were converted from Brain Voyager format to ANALYZE format, and then imported into Matlab 7 and the toolbox PLSgui v5.0609151 (McIntosh et al., 1996). PLS performs a singular value decomposition of the cross-correlation between the design matrix of conditions and functional activity to derive a set of condition latent variables (each of which is a linear contrast across the design conditions) and a corresponding set of brain latent variables (each of which is a set of voxel weights). PLS is described more fully elsewhere (McIntosh et al., 1996; McIntosh and Lobaugh, 2004), but, briefly, PLS determines the unique patterns of activity across all voxels that best correspond to commonalities and differences among the experimental conditions.

3.4 Results

3.4.1 Behavioural Results

Accuracy on the 4AFC task increased from the first to the last training session (Figure 3.4a), with mean accuracy on training day 5 significantly greater than that of day 1 ($t(11) = 11.24, p < .001$). A subset of 6 subjects returned approximately one week after the final fMRI session to be tested on both trained and untrained houses. Performance on trained houses at post-test was not significantly different from the mean performance of these 6 subjects on the 5th day of training ($t(5) = 1.06, p = .34$), consistent with the notion that learning for this stimulus set asymptotes after approximately 5 days of training (Husk et al., 2007). Performance on the untrained house set was significantly worse than on the trained house set ($t(5) = 4.25, p = .008$), and did not significantly exceed Day 1 performance on the trained house set, suggesting no transfer of learning across sets (Figure 3.4b). The lack of set transfer observed here is inconsistent with an earlier experiment using these same stimuli (Husk et al., 2007); in that case, limited, but partial, transfer was observed across sets, but consistent with highly specific learning seen for other complex stimuli (Hussain et al., 2009)

To ensure that performance benefits incurred during training outside the scanner were exhibited in the context of the scanning sessions, a measure of accuracy on the training task was also evaluated during the pre-testing and post-testing in the fMRI scanner. Note that direct comparisons of absolute performance levels from the training sessions (outside the scanner) to the scanner sessions are not meaningful, as viewing conditions (e.g., stimulus contrast) in the scanner could not be equated to conditions outside the scanner. As expected, performance on the house training task rose significantly ($t(10) = 3.881, p = .003$) from pre-test ($M = 0.70, SE = 0.03$) to post-test ($M = 0.82, SE = 0.02$). The significant effect of training observed across the scanning sessions confirms that the differing environmental conditions (scanner noise, potential subject discomfort, etc.) did not interfere with exhibition of learning achieved outside the scanner.

Behavioural performance measures were also collected during the 1-back task in the two functional imaging scans (Figure 3.5. The 1-back task was intended primarily to ensure that subjects paid attention to the stimuli while performing a task that required them to differentiate the objects at an individual exemplar level (as in the training task). The 1-back task was not intended to be difficult, and therefore learning across sessions was not necessarily expected in all conditions. Repeated measures analyses of

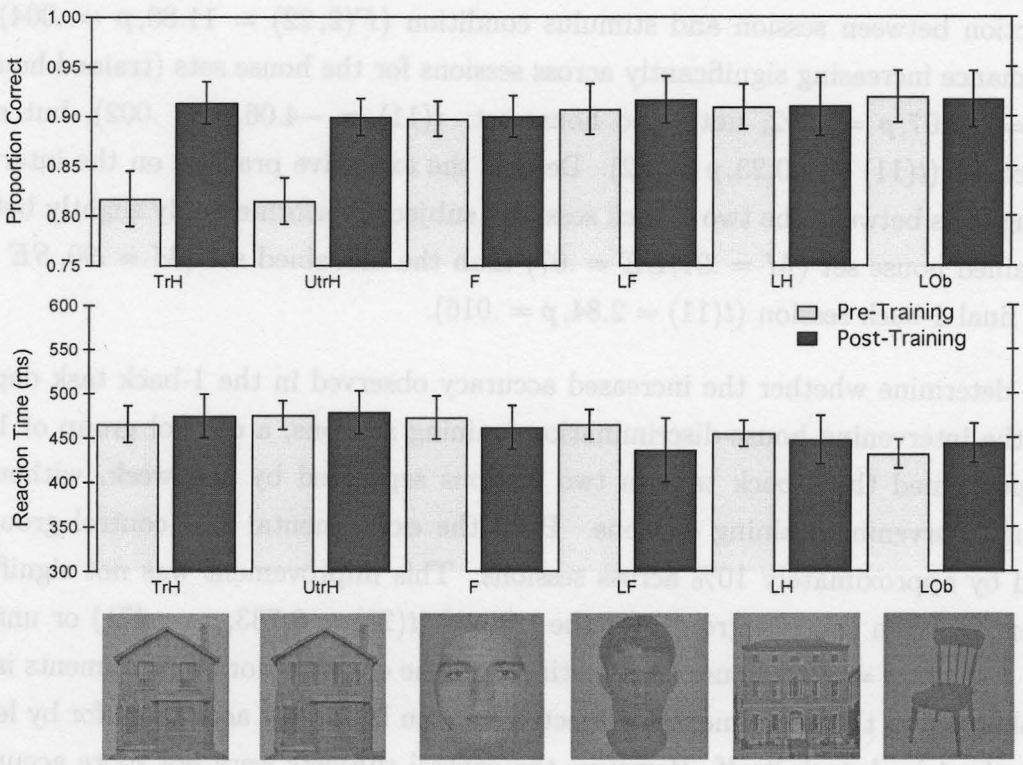

Figure 3.5: 1-back behavioural performance for 12 subjects during pre-training and post-training fMRI sessions. Performance is plotted for Experimental stimuli: Trained houses, untrained houses, and faces; and Localizer (Loc) stimuli: localizer faces, localizer houses, and localizer objects. Statistical analyses were restricted to the three experimental conditions, but behavioural performance on the localizer conditions is included here for visual comparison. Example stimuli are presented beneath each stimulus label.

variance (Greenhouse-Geisser corrected) were conducted separately on the reaction time and accuracy behavioural data. Each ANOVA was run with a 3 (stimulus condition) x 2 (session) design. Reaction times did not change significantly across sessions in any of the three experimental conditions (Session: $F(1, 11) = 0.00, p = .992$; Condition: $F(2, 22) = 0.46, p = .594$; Session x Condition: $F(2, 22) = 0.62, p = .537$). By contrast, accuracy increased overall across sessions ($F(1, 11) = 16.54, p = .002$), and varied significantly across stimulus conditions ($F(2, 22) = 9.05, p = .007$). ANOVA also revealed a significant interaction between session and stimulus condition ($F(2, 22) = 11.80, p = .004$), with performance increasing significantly across sessions for the house sets (trained house set: $t(11) = -4.07, p = .002$; untrained house set: $t(11) = -4.06, p = .002$), but not for the face set ($t(11) = -0.23, p = .82$). Despite the extensive practice on the intervening training days between the two 1-back sessions, subjects performed only slightly better on the trained house set ($M = .91, SE = .07$) than the untrained set ($M = .90, SE = .06$) in the final 1-back session ($t(11) = 2.84, p = .016$).

To determine whether the increased accuracy observed in the 1-back task depended upon the intervening house-discrimination training sessions, a control group of 12 subjects performed the 1-back task in two sessions separated by one week, without performing intervening training sessions. Both the experimental and control groups improved by approximately 10% across sessions. This improvement was not significantly different between the two groups for the trained ($t(22) = 0.733, p = .471$) or untrained ($t(22) = 0.137, p = .892$) houses, suggesting that the cross-session improvements in accuracy observed in the experimental subject group can be largely accounted for by learning within the 1-back task itself. However, the control subjects were not more accurate on the trained than untrained houses in the final 1-back session ($t(11) = 0.14, p = .892$), unlike the experimental subjects who showed a small but significant increase of 1% for the trained relative to untrained houses. Together, these results suggest that, for the experimental subjects, much of the learning observed in the 1-back task was due to prior experience with the 1-back task itself, but that some small additional gains were observed from the intervening house-training sessions.

3.4.2 Region of Interest Results

Region of interest analyses were conducted to determine whether training on house discrimination was associated with changes in activation within regions primarily responsive to houses (PPA, RSC, & TOS). Activity was also examined within FFA to further examine its role in object learning. BOLD activity was quantified as the percent signal

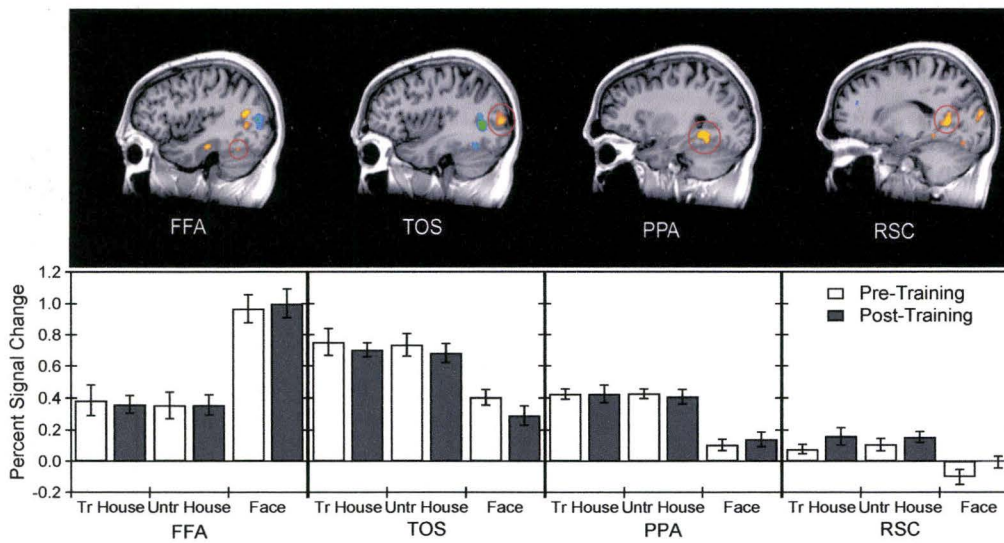

Figure 3.6: Mean percent signal change for regions FFA, TOS, PPA, and RSC (right hemisphere in all cases) in response to trained- and untrained-house, and face conditions before (light bars) and after (dark bars) 5 days of training on the trained house set. Sagittal image slices of right hemisphere of a typical observer is presented above the mean activity plots. In the sagittal image slices, regions of interest are indicated by circled regions within each slice. Note that the region colours represent correlations with GLM linear contrasts (red = positive correlations; blue = negative correlations), so for FFA, red = faces; blue = houses and objects; but for TOS, PPA, and RSC, red = houses; blue = faces and objects.

ROI	Source	Statistic	p
FFA (n=11)	Session	$F(1,10) = 0.000, p = .986$	ns
	Object Class	$F(2,20) = 94.481, p < .001$	***
	Session x Obj.	$F(2,20) = 0.443, p = .618$	ns
	Face vs. Tr H	$t(10) = 9.751, p < .001$	***
	Face vs. Untr H	$t(10) = 10.487, p < .001$	***
	Tr H vs. Untr H	$t(10) = 0.697, p = .502$	ns
TOS (n=10)	Session	$F(1,9) = 2.060, p = .185$	ns
	Object Class	$F(2,18) = 50.747, p < .001$	***
	Session x Obj.	$F(2,18) = 0.307, p = .662$	ns
	Face vs. Tr H	$t(9) = 8.222, p < .001$	***
	Face vs. Untr H	$t(9) = 6.992, p < .001$	***
	Tr H vs. Untr H	$t(9) = 0.801, p = .444$	ns
PPA (n=12)	Session	$F(1,11) = 0.023, p = .883$	ns
	Object Class	$F(2,22) = 248.929, p < .001$	***
	Session x Obj.	$F(2,22) = 0.291, p = .698$	ns
	Face vs. Tr H	$t(11) = 17.901, p < .001$	***
	Face vs. Untr H	$t(11) = 16.718, p < .001$	***
	Tr H vs. Untr H	$t(11) = 0.795, p = .444$	ns
RSC (n=7)	Session	$F(1,6) = 4.485, p = .079$	*
	Object Class	$F(2,12) = 93.414, p < .001$	***
	Session x Obj.	$F(2,12) = 0.294, p = .629$	ns
	Face vs. Tr H	$t(6) = 11.361, p < .001$	***
	Face vs. Untr H	$t(6) = 9.523, p < .001$	***
	Tr H vs. Untr H	$t(6) = 1.452, p = .197$	ns

Table 3.1: Results of repeated-measures ANOVAs (2 sessions x 3 object sets), for each ROI. Results of paired comparisons (dependent t-tests) are displayed following main effect and interaction terms.

change relative to a baseline measure collected during fixation intervals. Examples of the locations of the regions of interest for one subject, and average percent signal change for each condition and ROI are shown in Figure 3.6. Because there was high inter-subject variability in the presence of the individual regions (all four regions were clearly identifiable in only six of the twelve subjects), separate statistical analyses were conducted for each region to maximize the number of subjects included in each analysis (Table 3.1).

As expected, there were significant effects of object-class in all four regions, with greater activity for faces than for both house sets in the FFA, and greater activity for both house sets than faces in TOS, RSC, and PPA. Activity did not differ significantly between house sets in any area. Most importantly, activity levels did not change significantly after training in any region for any condition (no significant effects of session, nor session by object-class interactions). There was a trend toward an overall increase in activity from session 1 to session 2 in region RSC; however this trend did not interact with object-class. This small increase was present for faces as well as for houses, suggesting that the session effect was not specifically due to training of an object-class.

3.4.3 Retinotopic Analyses

We also looked for effects of training within retinotopically-defined early visual areas. Activity in these regions was similarly defined in terms of percent signal change relative to baseline fixation intervals, and these activity levels were compared across the two imaging sessions. No evidence for training effects was found in any early visual area (Figure 3.7). This was confirmed with a repeated measures analysis of variance (6 regions x 2 sessions x 3 object-sets) where neither the effect of session ($F(1,11) = 0.524, p = .484$), nor any of the session interaction terms were significant (Region x Session x Object interaction: $F(10,110) = 1.669, p = .195$; Session x Object interaction: $F(2,22) = 0.759, p = .480$; Session x Region interaction: $F(5,55) = 0.46, p = .657$). Learning effects might also be reflected in greater differences between the two house sets in post-test than in pre-test. To examine this hypothesis, difference scores for the two house sets were compared across sessions and visual areas. Neither the main effect of session ($F(1,11) = 1.27, p = .28$) nor the session x visual area interaction ($F(5,55) = 1.33, p = .29$) were significant, indicating that there was no significant effects of learning on the activity difference between trained and untrained houses. Finally, to ensure that subtle effects from individual regions were not being masked by these analyses, we also conducted repeated-measures ANOVAs (2 conditions (trained vs. untrained houses) x 2 sessions) separately on each early visual area. No evidence of learning was observed: There were

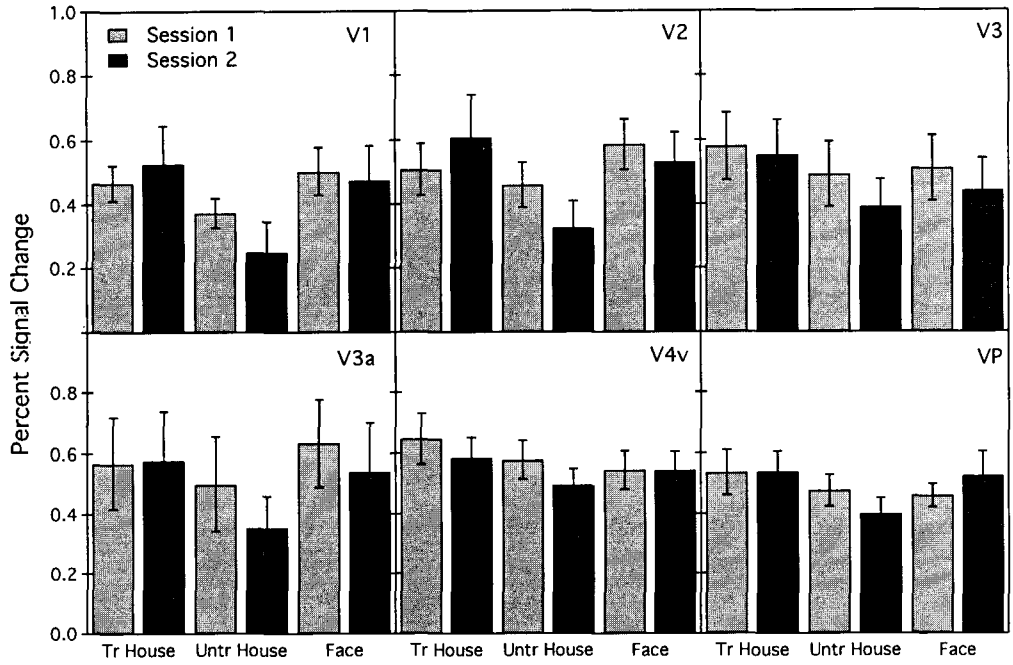

Figure 3.7: Mean percent signal change for early visual areas V1, V2, V3, V3a, V4v, and VP in response to trained houses, untrained houses and faces before (light bars) and after (dark bars) training. All regions were identified using retinotopic mapping. Error bars represent +/- 1 standard error of the mean. None of the cross-session paired comparisons reached significance.

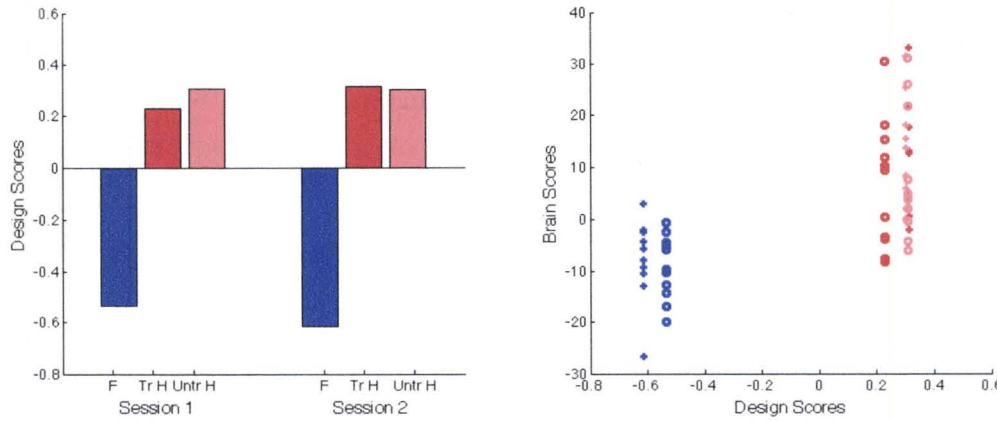

Figure 3.8: Results from partial least squares analysis. (a) The linear contrast that best describes the activity at the voxels. This latent variable contrasts house activity (trained and untrained) against face activity and has similar contributions from sessions 1 and 2, suggesting little effect of training. (b) Brain scores versus design scores (condition weights) for all subjects. Brain scores are summary values that reflect the weights across the brain voxels for each subject. In this case, negative brain scores are associated with the face condition (regardless of session), and positive brain scores are associated with the two house conditions in both sessions. Blue = faces; dark red = trained houses; light red = untrained houses. Circles = Session 1 and plusses = Session 2. (c) The voxel weights that contribute to the final brain scores. These voxel weights represent all subjects, but are overlayed on the anatomical scan for a single sample subject. Brain voxels coloured in red indicate regions that are positively correlated with house activity, while blue are positively correlated with face activity.

no significant interactions between session and house condition for any of the early visual areas (all $p's > .23$). Nor was the main effect of session significant in any early visual area (all $p's > .39$).

3.4.4 Partial Least Squares Analysis

To examine the effects of learning more broadly across the cortex as a whole, we used a data-driven, multivariate technique: partial least squares (PLS; McIntosh et al., 1996; McIntosh and Lobaugh, 2004). We conducted a task-based PLS analysis that included three experimental conditions (trained houses, untrained houses, and faces) before and after learning. The PLS analysis identified only one significant latent variable, which is shown in Figure 3.8. The design scores plotted in Fig. 3.8a can be understood as a linear contrast across experimental conditions: The height of the bars indicate the weight associated with the particular condition. The latent variable described the differ-

ence between the face and house stimuli. Trained and untrained houses were weighted approximately equally, and their relative weightings did not differ before or after learning. The similar contributions of condition weightings across the two sessions suggests that learning had very little effect on the patterns of activity described by this latent variable. The relation between the brain scores and design scores (Fig. 3.8b) indicates the relation between the brain voxel weights (which combine to form a single brain score for each subject) and the condition weights (the design scores displayed in Fig. 3.8a). In this case, faces are associated with negative brain weights, and houses are associated with positive brain weights. The singular image in Fig. 3.8c shows the weights of the voxels that, across subjects, covary across experimental conditions. The brain regions reliably associated with this latent variable include coordinates that agree with reports of the FFA (Talairach coordinates for right hemisphere: [39 -60 -15] and left hemisphere: [-30 -66 -15]) and PPA (Talairach coordinates for right hemisphere: [27 -45 -9] and left hemisphere: -24 -45 -9]).

Task	Hemisphere	FFA		TOS		PPA		RSC	
		r	p	r	p	r	p	r	p
Training task	Right	0.06	.87	0.32	.36	0.01	.99	0.22	.63
(tr. houses)	Left	-0.72	.07	0.28	.43	0.00	.99	-0.11	.84
1-back	Right	-0.14	.68	0.13	.71	0.33	.29	0.40	.37
(tr. houses)	Left	-0.46	.30	-0.11	.75	0.11	.74	0.30	.57
1-back	Right	0.16	.65	0.03	.94	0.59	.04*	0.92	.004**
(untr. houses)	Left	-0.36	.43	-0.15	.68	0.26	.41	0.36	.48
1-back	Right	-0.18	.60	0.33	.35	0.41	.18	-0.09	.84
(faces)	Left	0.05	.92	-0.31	.39	0.22	.48	0.68	.14

Table 3.2: Correlations between behavioural improvements in both the training and 1-back tasks and functional BOLD activity increases across scanning sessions (Post-Pre). For the training task, behavioural improvements were measured as the difference between the final training session and the initial training session, and were compared to the functional activity difference for the same trained house set (collected in the context of the 1-back task). For the 1-back task, correlations were computed separately for functional and behavioural improvements for each of the trained house, untrained house, and face conditions.

3.4.5 Behavioural correlations

The behavioural results suggested some variability in the amount of learning observed across subjects. These individual differences might mask learning related changes in

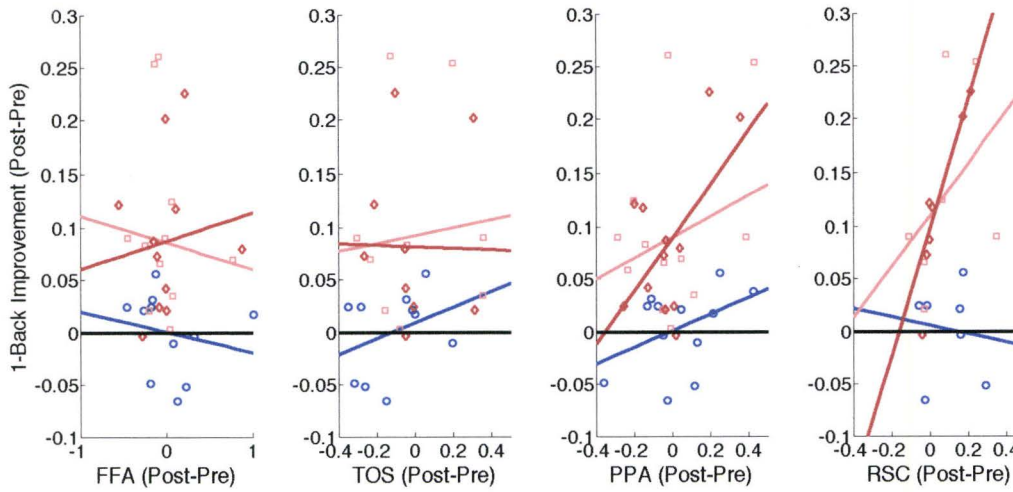

Figure 3.9: Scatter plots display the relation between improvement on the 1-back behavioural task and increases in BOLD activity in FFA, TOS, PPA, and RSC across scanning sessions. Data points are individual subject data. Data and lines of best-fit for faces (blue circles), trained houses (light red squares) and untrained houses (dark red diamonds) are displayed. Corresponding Pearson correlation coefficients are listed in Table 3.2.

BOLD activation that were not observable in comparisons of mean functional activity across sessions. Therefore, correlations were examined between the amount of learning in both the training and 1-back tasks, and changes in mean functional activation within our regions of interest from pre-test to post-test. The amount of learning was defined as the accuracy difference from the first to last day of training. However, no significant correlations were observed between activity levels in any region and the extent of learning on the training task (Table 3.2).

Because the behavioural analysis of the 1-back task also suggested that learning of the experimental house sets had occurred between pre-test and post-test, we examined correlations learning on the 1-back task (i.e., accuracy difference between pre-test and post-test) and the change in functional activation across these two sessions. Interestingly, despite the relatively constant level of mean activity across sessions in all regions of interest, significant correlations between the amount of learning and the change in activation were found in RSC and PPA (Table 3.2; Figure 3.9). These correlations were significant only for houses not used in the training task, although the trends were similar with the other house set. Although it may seem surprising that the correlations were significant only for the untrained house set, it is important to note that much of the

learning in the 1-back task was due to learning in the context of the 1-back task itself. In that context, the two experimental house sets received equal training (from the 1-back practice at pre-test), so larger effects for the "trained" experimental house set would not necessarily be expected here. Correlations between 1-back improvement and behaviour were also measured for the faces. Although subjects did not show significant learning for faces from pre-test to post-test, there was enough variability in performance change across sessions to permit this comparison; nevertheless, significant correlations were not observed for the faces in any region.

Task	Hem.	V1		V2		V3		V3a		V4v		VP	
		r	p	r	p	r	p	r	p	r	p	r	p
Training task	Right	0.40	.19	0.33	.29	0.46	.13	0.45	.14	0.29	.36	0.30	.35
(tr. houses)	Left	0.39	.22	0.54	.07	0.10	.76	0.25	.43	0.43	.16	0.42	.18
1-back	Right	-0.02	.94	-0.13	.68	-0.30	.34	-0.26	.42	-0.17	.60	-0.18	.58
(tr. houses)	Left	-0.25	.42	-0.15	.64	0.05	.87	0.06	.85	-0.19	.56	-0.34	.28
1-back	Right	-0.04	.91	-0.01	.98	-0.12	.72	-0.09	.78	0.12	.71	0.06	.85
(untr. houses)	Left	-0.26	.41	-0.12	.71	-0.02	.94	-0.13	.70	-0.25	.44	-0.33	.29
1-back	Right	0.60	.04*	0.47	.12	0.31	.32	-0.20	.53	0.05	.88	0.28	.37
(faces)	Left	0.26	.42	0.23	.47	0.18	.58	0.30	.35	0.42	.18	0.41	.18

Table 3.3: The relation between improvement on the 1-back behavioural task and increases in BOLD activity in early visual areas V1, V2, V3, V3a, V4v, and VP across scanning sessions. Data points are individual subject data.

Correlations between improvement on both the training task and the 1-back task and differences in the BOLD signal across sessions were also examined within retinotopic visual areas. No significant correlations were observed for either experimental house set in any of these areas. Improvements in face performance on the 1-back task were moderately correlated with changes in V1 activity across sessions (Table 3.3).

3.5 Discussion

Five days of training on a house 4AFC discrimination task was sufficient to produce a substantial change in behavioural performance, with accuracy in the training task rising 23% relative to baseline performance on the first day of training. This result is consistent with previous behavioural experiments examining training with the same stimuli (Husk et al., 2007). These earlier behavioural experiments also demonstrated the emergence of strong inversion effects with the same amount of training, suggesting that five days of training is sufficient to induce at least some characteristics of expertly processed object classes (Husk et al., 2007). Previous experiments also demonstrated that additional days of training beyond five days provide only minimal additional increases in performance (Husk et al., 2007, Exp 1). Despite the substantial behavioural learning in the 4-AFC training task, no corresponding significant functional changes were observed in our fMRI measures: No significant activation differences were observed across sessions within the whole-brain multivariate analyses, nor within the more localized ROI analyses of face preferential areas (FFA), house preferential areas (PPA, RSC, & TOS), or retinotopically-defined early visual areas. Nor were there any significant correlations between behavioural performance (or improvement) in these tasks and changes in activation within these areas. The functional activity in these ROIs appeared quite stable across sessions and minimally influenced by the training task.

Because the 1-back task was performed both at pre-test and at post-test, there was opportunity for learning to occur within the context of this task as well, and performance did improve for both experimental house sets across sessions. Significant correlations were observed between the amount of improvement on the 1-back task across sessions and activity increases in both the RSC and PPA for one of the two experimental house sets: those subjects who showed the greatest improvement in the identification of the experimental house sets across sessions exhibited the largest changes in activity within both the house responsive regions PPA and RSC, with the strongest effects in the RSC. No such correlations were observed in either TOS or FFA, nor in any of the early visual

areas.

The regions that emerged in the current study as particularly responsive to houses (PPA, RSC, & TOS) have been previously implicated, more generally, as important for tasks involving scene processing (Aguirre et al., 1998; Epstein et al., 1999; Epstein and Kanwisher, 1998; Epstein et al., 2007a,b; O'Craven and Kanwisher, 2000; Grill-Spector, 2003), and the sensitivity of RSC and PPA to learning is consistent with some previous reports of the effect of familiarity in scene processing (Epstein et al., 2007a,b). Of these regions, RSC seems to be reliably more responsive to familiar scenes than unfamiliar scenes (Epstein et al., 2007a,b), consistent with the stronger effect of learning in the RSC observed in the current study. The effect of familiarity on PPA activity is less clear, as effects of familiarity have been observed (Epstein et al., 2007a), but not consistently (Epstein et al., 2007b). Epstein has suggested that the stronger effect of familiarity on the RSC than PPA is indicative of differences in the areas' relative roles in scene processing. He proposed that the RSC is associated with retrieval of topographical memory, whereas the PPA is responsible for visual representations of spatial structure. Our findings suggest that RSC shows greater sensitivity to learning than PPA, with larger RSC activation changes for those who demonstrated the most task-learning. Thus, even in the context of a task where all stimuli were familiar to the subjects, RSC activity changes scaled with degree of learning. However, note that the stronger correlations we observed for the untrained house-set pose a challenge to this interpretation: subjects had greater exposure to the trained houses (across tasks). If the RSC is responsive to familiarity, one would expect stronger activity for the trained house set. Nonetheless, if Epstein's interpretation of the role of RSC is correct, the correlations between behavioural learning and RSC may reflect changes in memory retrieval for the house exemplars with learning.

The current study found no clear evidence for the involvement of FFA in learning under the conditions of this study, despite the large behavioural improvements in the task, and the presence of markers of expertise, such as the emergence of an inversion effect after training of these stimuli (Husk et al., 2007). These results, combined with other contradictory reports of FFA's role in learning suggest that, although the fusiform may play a role in the learning of some object classes under some conditions, the FFA seems not to be universally involved in object learning. Unfortunately, the broad diversity in methodology across all previous studies, in terms of the type of object classes studied, the stimulus set-sizes, and the duration and method of training, makes it difficult to pinpoint those conditions under which the fusiform may be involved in object learning. This will be an important goal for future studies aimed at characterizing the function of

the fusiform gyrus in object learning.

Early visual areas were also unresponsive to learning of this task. This is interesting given that learning in the context of contour completion tasks does produce changes in early visual areas (Kourtzi et al., 2005; Li et al., 2008). However, other studies examining object recognition have also failed to find early visual differences in object learning (Op de Beeck et al., 2006). Contour completion learning tasks may tap into lower-level processes than object learning tasks. The greater involvement of high-level areas (such as RSC) over lower-level visual areas is consistent with the reverse-hierarchy model of learning (Ahissar and Hochstein, 1997; Ahissar, 1999). Overall, these results, combined with those of previous studies, suggest that there is no single common neural area that represents all aspects of visual learning. The areas involved in visual learning likely depend both on the stimulus and the type of training employed. The role of FFA may be more limited than previously thought, and other areas may play larger roles in different contexts. The RSC and PPA, in particular seem to play an important role in the learning of house discrimination tasks, acting in the context of a larger, more distributed, cortical network.

3.6 Acknowledgements

The localizer object stimuli were provided courtesy of Michael J. Tarr (Brown University, http://ww.tarrlab.org/). This work was supported by Canada Research Chairs (PJB & ABS), by NSERC Discovery Grants 42133 & 105494, and by NSERC PGS-D awards (JSH & LRB).

References

Aguirre, G. K., Zarahn, E., D'Esposito, M., 1998. An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. Neuron 21 (2), 373–383.

Ahissar, M., 1999. Perceptual learning. Current Directions in Psychological Science 8 (4), 124–128.

Ahissar, M., Hochstein, S., 1997. Task difficulty and the specificity of perceptual learning. Nature 387 (6631), 401–406.

Bennett, P. J., Sekuler, A. B., McIntosh, A. R., Della-Maggiore, V., 2001. The effects of aging on visual memory: evidence for functional reorganization of cortical networks. Acta Psychol (Amst) 107 (1-3), 249–273.

Brainard, D. H., 1997. The psychophysics toolbox. Spat Vis 10 (4), 433–436.

Engel, S. A., Glover, G. H., Wandell, B. A., 1997. Retinotopic organization in human visual cortex and the spatial precision of functional mri. Cereb Cortex 7 (2), 181–192.

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.

Epstein, R. A., Higgins, J. S., Jablonski, K., Feiler, A. M., 2007a. Visual scene processing in familiar and unfamiliar environments. J Neurophysiol 97 (5), 3670–3683.

Epstein, R. A., Parker, W. E., Feiler, A. M., 2007b. Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. J Neurosci 27 (23), 6141–6149.

Gauthier, I., Skudlarski, P., Gore, J. C., Anderson, A. W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci 3 (2), 191–197.

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. Nat Neurosci 2 (6), 568–573.

Gold, J., Bennett, P. J., Sekuler, A. B., 1999a. Identification of band-pass filtered letters and faces by human and ideal observers. Vision Res 39 (21), 3537–3560.

Gold, J., Bennett, P. J., Sekuler, A. B., 1999b. Signal but not noise changes with perceptual learning. Nature 402 (6758), 176–178.

Gold, J., Sekuler, A., Bennett, P., 2004. Characterizing perceptual learning with external noise. Cognitive Science 28 (2), 167–207.

Grill-Spector, K., 2003. The neural basis of object perception. Curr Opin Neurobiol 13 (2), 159–166.

Grill-Spector, K., Knouf, N., Kanwisher, N., 2004. The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci 7 (5), 555–562.

Grill-Spector, K., Kushnir, T., Hendler, T., Malach, R., 2000. The dynamics of object-selective activation correlate with recognition performance in humans. Nat Neurosci 3 (8), 837–843.

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.

Husk, J. S., Bennett, P. J., Sekuler, A. B., 2007. Inverting houses and textures: investigating the characteristics of learned inversion effects. Vision Res 47 (27), 3350–3359.

Hussain, Z., Sekuler, A. B., Bennett, P. J., 2009. Contrast-reversal abolishes perceptual learning. J. Vis. 9, 1–8.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., Haxby, J. V., 1999. Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci U S A 96 (16), 9379–9384.

Joseph, J. E., Gathers, A. D., 2002. Natural and manufactured objects activate the fusiform face area. Neuroreport 13 (7), 935–938.

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

Kanwisher, N., McDermott, J., Chun, M. M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17 (11), 4302–4311.

Kourtzi, Z., Betts, L. R., Sarkheil, P., Welchman, A. E., 2005. Distributed neural plasticity for shape learning in the human visual cortex. PLoS Biol 3 (7), e204.

Li, W., Piech, V., Gilbert, C. D., 2008. Learning to link visual contours. Neuron 57 (3), 442–451.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., Grady, C. L., 1996. Spatial pattern analysis of functional brain images using partial least squares. Neuroimage 3 (3 Pt 1), 143–157.

McIntosh, A. R., Lobaugh, N. J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 23 Suppl 1, S250–63.

O'Craven, K. M., Downing, P. E., Kanwisher, N., 1999. fMRI evidence for objects as the units of attentional selection. Nature 401, 584–587.

O'Craven, K. M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. J Cogn Neurosci 12 (6), 1013–1023.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., Kanwisher, N. G., 2006. Discrimination training alters object representations in human extrastriate cortex. J Neurosci 26 (50), 13025–13036.

Oppenheim, A., Lim, J., 1981. The importance of phase in signals. Proceedings of the IEEE 69 (5), 529–541.

Pelli, D. G., 1997. The videotoolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10 (4), 437–442.

Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., Clifford, C. W. G., 2004. Orientation-contingent face aftereffects and implications for face-coding mechanisms. Curr Biol 14 (23), 2119–2123.

Schoups, A., Vogels, R., Qian, N., Orban, G., 2001. Practising orientation identification improves orientation coding in v1 neurons. Nature 412 (6846), 549–553.

Sekuler, A. B., Bennett, P. J., 1996. Spatial phase differences can drive apparent motion. Percept Psychophys 58 (2), 174–190.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., Tootell, R. B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268 (5212), 889–893.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., Kanwisher, N., 2000. Response properties of the human fusiform face area. Cognitive Neuropsychology 17 (1-3), 257–279.

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

Chapter 4

House-identity training systematically decreases strength of late ERP components

4.1 Abstract

Experts differ from novices in their ERP responses (Tanaka and Curran, 2001; Gauthier et al., 2003). In particular, the N1, a component often associated with face processing, is larger for experts than novices when viewing objects drawn from their classes of expertise. This component is also larger following explicit training of object recognition (Scott et al., 2006, 2008). However, few studies have tracked the emergence of ERP training across multiple training sessions. We trained subjects on a house discrimination task across 5 sessions. ERP recordings were collected during all training sessions. ERP activity declined significantly across sessions, particularly in the time-range of 200-300 ms. Throughout all sessions, reaction time was correlated with the amplitude of early time-ranges, and accuracy was correlated with the amplitude in later time-ranges. These results imply that training can have strong, systematic effects on the strength of the ERP response. However, training does not always increase the strength of the N1 component.

4.2 Introduction

Event-related potentials (ERPs) permit the precise investigation of the timing of events, and therefore may provide important insights about how the development of

expertise alters the timing of neural processing. Interestingly, although a growing number of studies have addressed questions of how processing differs for familiar versus unfamiliar stimuli (e.g., Guillaume and Tiberghien, 2001; Itier and Taylor, 2002), or have compared the EEG signatures of real world experts to novices (e.g., Tanaka and Curran, 2001; Gauthier et al., 2003), few studies have directly tracked the emerging changes in neural processing across a series of consecutive training sessions.

Training modifies several ERP components including the N1 (e.g., Scott et al., 2006, 2008) the N250 (Tanaka et al., 2006) and the N400 (Speer and Curran, 2007). Most of the expertise research has focussed on the N1, a component that occurs around 150-200 ms after stimulus onset, which has a particularly large amplitude following the presentation of faces relative to the presentation of other classes of objects (Bentin et al., 1996; Eimer, 2000a; Itier and Taylor, 2004; Jeffreys and Tukmachi, 1992). Because faces frequently are encountered in the environment, and often are presumed to be expertly processed, the sensitivity of the N1 to faces has resulted in a wealth of interest in the sensitivity of this component to other expertly-processed object classes.

Expertise in a variety of areas has been linked to changes in the N1. In some cases, expertise has been linked to an increase in the amplitude of the N1. For example, dog and bird experts (Tanaka and Curran, 2001) and car experts (Gauthier et al., 2003) produce larger N1 responses to stimuli depicting objects drawn from their respective classes of expertise. Similarly, English readers exhibit larger N1 responses to Roman lettering than to Chinese characters, while English/Chinese bilinguals produce similarly sized N1 responses to both character sets (Wong et al., 2005). In other cases, expertise has been linked to changes in the strength of the N1 inversion effect. For faces, the N1 component is both larger and delayed when faces are presented inverted relative to upright (e.g., Rossion et al., 2000; Itier et al., 2006; Rousselet et al., 2007, 2008), and this sensitivity to stimulus inversion has become a hallmark of expert processing. Both fingerprint experts (Busey and Vanderkolk, 2005), and subjects trained with novel objects (Rossion et al., 2002) exhibit N1 inversion effects similar to those observed for faces.

N1 responses have also been compared across varying degrees of expertise within the category of faces. For example, humans exhibit larger N1 responses to human than ape faces (Carmel and Bentin, 2002) and monkeys exhibit the reverse pattern of larger N1 and P2 responses to monkey faces than to human faces (Peissig et al., 2007). Similarly, humans may have differential levels of experience discriminating amongst faces of one's race relative to those of other races. There is some evidence that the N1 may be larger

for own-race faces (Ito and Urland, 2005), however race effects on the N1 appear to be highly variable: Some studies have found exactly the opposite effect of smaller N1 in response to own race faces (Stahl et al., 2008), while others have failed to find any significant effect of race (Ito et al., 2004). It may be that some of the discrepancies across studies of race-effects could be explained by differential exposure to faces of other races across subject groups. Surprisingly few studies have examined changes in N1 produced by carefully controlled amounts of training. The most direct testing involved training on owls (Scott et al., 2006) and cars (Scott et al., 2008), wherein both the N1 and N250 components were enhanced by in-lab training (with the N250 sensitive only to subordinate-level training, in the case of owl training). The effect of training on the N250 amplitude was robust, persisting for at least one week after training, but the effect on N1 amplitude disappeared following the week delay. Training subjects to categorize novel objects also results in earlier and larger N1 responses to both trained objects and objects derived from the same prototype as the trained objects (Curran et al., 2002 Mar).

The current study tracked EEG signatures of training across each of five consecutive days of training on a house discrimination task, which will enable us to examine changes in the N1 and other components across the gradual process of training a visual discrimination task. In addition to standard ERP analyses, global changes across time and space will be examined through partial least squares, a data-drive multivariate statistical analysis.

4.3 Methods

4.3.1 Subjects

Eight subjects (M = 23.1 years; range = 20-27) were drawn from the McMaster University Vision and Cognitive Neuroscience Lab subject pool. All were right-handed, and six were male. All subjects had normal or corrected-to-normal acuity. Subjects were paid \$10/hr for participating in the experiment.

4.3.2 Stimuli

The training stimuli were a set of 8 houses (Figure 4.1). The individual houses in the training set were visually similar, and were constructed by selecting a single house to act as a common frame. House exemplars were differentiated on the basis of the doors and windows, with each individual house having a unique door, lower window and upper

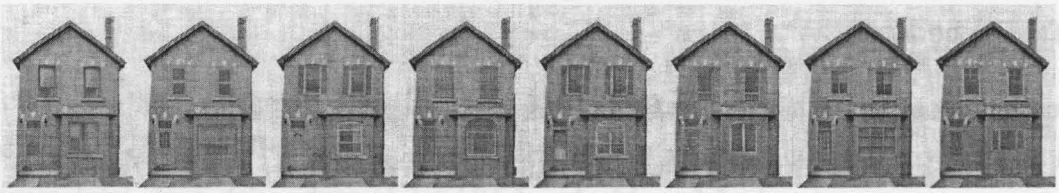

Figure 4.1: Training houses.

window pair (see Husk et al., 2007, for additional details). Target stimuli were presented at low contrast (contrast variance = 0.001) and were sized to 7.3 x 5.2 deg visual angle. Target and distractor stimuli were presented at higher contrast on the response screen (contrast variance = 0.5), and were reduced in size to 3.7 x 2.6 deg visual angle.

4.3.3 Apparatus

All stimuli were presented on a Sony Trinitron GDM-F520 monitor (resolution: 800 x 600 pixels; 20 pixels/cm; refresh rate: 85 Hz). The experiment was conducted within a dimly-lit sound-attenuated booth. Viewing distance (1 m) was controlled by use of a chin rest. Average luminance was 41.84 cd/m^2.

Electrophysiological recordings were collected using the 256 channel Geodesic sensor net system (Electrical Geodesics Inc., Eugene, Oregon; Tucker, 1993) and acquisition software NetStation 4.2.4. The ground electrode was located along the midline, anterior to Fz. Electrode impedances were maintained within 50 micro-ohms. The analog signal was digitized at 500 Hz and band-pass filtered between 0.1 and 200 Hz. EEG signal was referenced online to Cz, but re-referenced offline to the average across electrodes. The re-referenced EEG was band-pass filtered between 1 and 30 Hz. Bad channels were identified using routines in BESA 5.0, then removed from further processing. Baseline correction was performed using a baseline of 300 ms, and artifacts were rejected, removing any trials where the amplitude exceeded ±100 microvolts, or any trials wherein the amplitude changed by more than 75 microvolts across two consecutive time points. Only correct trials were retained for further analysis.

EEG analyses were performed using EEGLAB (Delorme and Makeig, 2004) with extraction of ERPs and additional statistical testing conducted with additional in-lab Matlab routines. Partial least squares analyses were performed with Matlab routines PLSgui (McIntosh and Lobaugh, 2004).

4.3.4 Procedure

Subjects completed five sessions of a 4-AFC house discrimination training task, with no more than one session per day and all sessions occurring within a two-week interval. Training sessions consisted of four blocks of 104 trials each, with each trial composed of approximately 1.5 s of fixation (range: 1.14-1.85 s), followed by a 500 ms stimulus presentation, after which a response screen was presented. The response screen displayed the target stimulus along with three distractor stimuli. The target and distractors were randomly selected from the set of eight houses, with the constraint that each house was presented an equal number of times per session. The target position was randomized within the response array across trials. This screen remained in place until a response was made. Subjects were instructed to respond by pressing one of four keyboard buttons to indicate the location of the target image on the response screen. High- (600 Hz) and low-pitched (200 Hz) tones followed each correct and incorrect response, respectively. At the start of each session, subjects were instructed to minimize blinking and head movements.

4.4 Results

4.4.1 Behavioural results

The behavioural training task improved house-discrimination performance both in terms of accuracy and reaction time for all subjects (Figure 4.2). Mean accuracy on the house discrimination task rose significantly from the first day ($M = 0.72, SEM = 0.03$) to the final day ($M = 0.93, SEM = 0.01$) of training ($t(7) = 8.60, p < .001$). In concert with the change in accuracy, reaction time fell from the first ($M = 1799, SEM = 121$) to last ($M = 1035, SEM = 44.8$) day of training ($t(7) = 7.811, p < .001$). One subject performed substantially worse than the others (both in terms of lower accuracy and longer reaction times). Nonetheless, exclusion of this subject had minimal impact on the results of both behavioural and electrophysiological analyses, therefore, this subject was not excluded from any of the presented analyses.

4.4.2 Event-related Potentials Analysis

Our first analysis compared trial-averaged ERPs measured during the five training sessions. Figure 4.3 plots ERPs across sessions for four representative subjects. Most subjects exhibited clear P1-N1-P2 responses. As identified by peak analyses, the exact

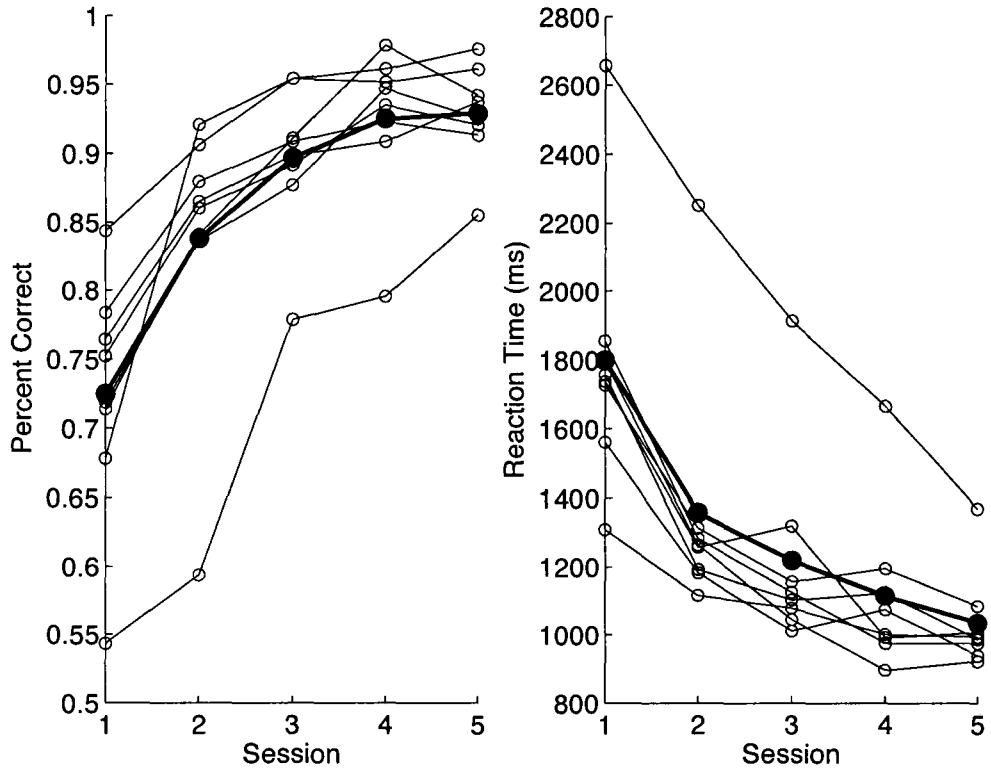

Figure 4.2: Behavioural training results on the 4-AFC house training task: (a) Accuracy (b) Reaction time across training sessions. Thin lines (open symbols) reflect individual subject data, while thick lines (filled symbols) reflect the mean across subjects.

Peak	Hemisphere	Amplitude			Latency		
		Day5-Day1	t	p	Day5-Day1	t	p
P1	Right	-0.09	-0.38	.71	-4.5	-1.03	.34
	Left	-0.92	-0.99	.35	0.25	0.09	.93
N1	Right	-0.45	-1.06	.32	1.75	0.30	.77
	Left	0.29	1.43	.19	7.25	2.01	.08
P2	Right	-0.6	-1.09	.31	-1.75	-0.37	.72
	Left	-0.18	-0.34	.75	-2.75	-1.01	.35

Table 4.1: Comparison of peak amplitude and latency for P1, N1, and P2 components from Day 1 to Day 5 of training.

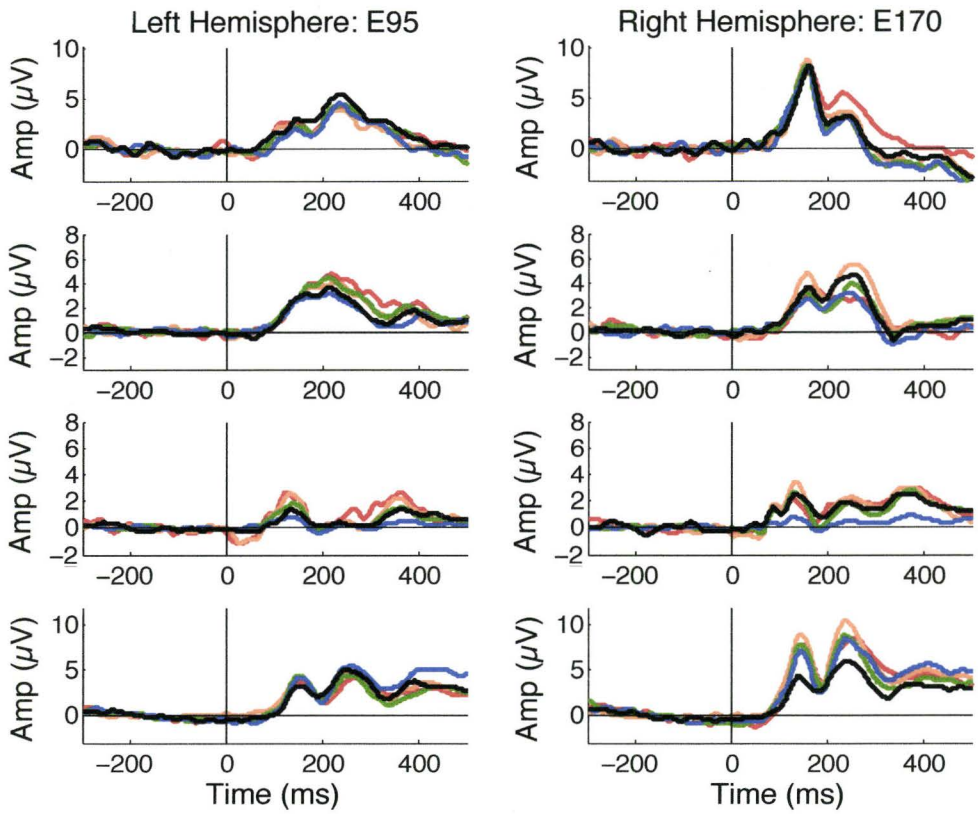

Figure 4.3: Training data: ERPs (averaged across trials) in response to house stimuli for 5 training sessions (Line colours from first to last session: Red, Orange, Green, Blue, & Black) for four representative subjects (left panel = left hemisphere; right panel = right hemisphere).

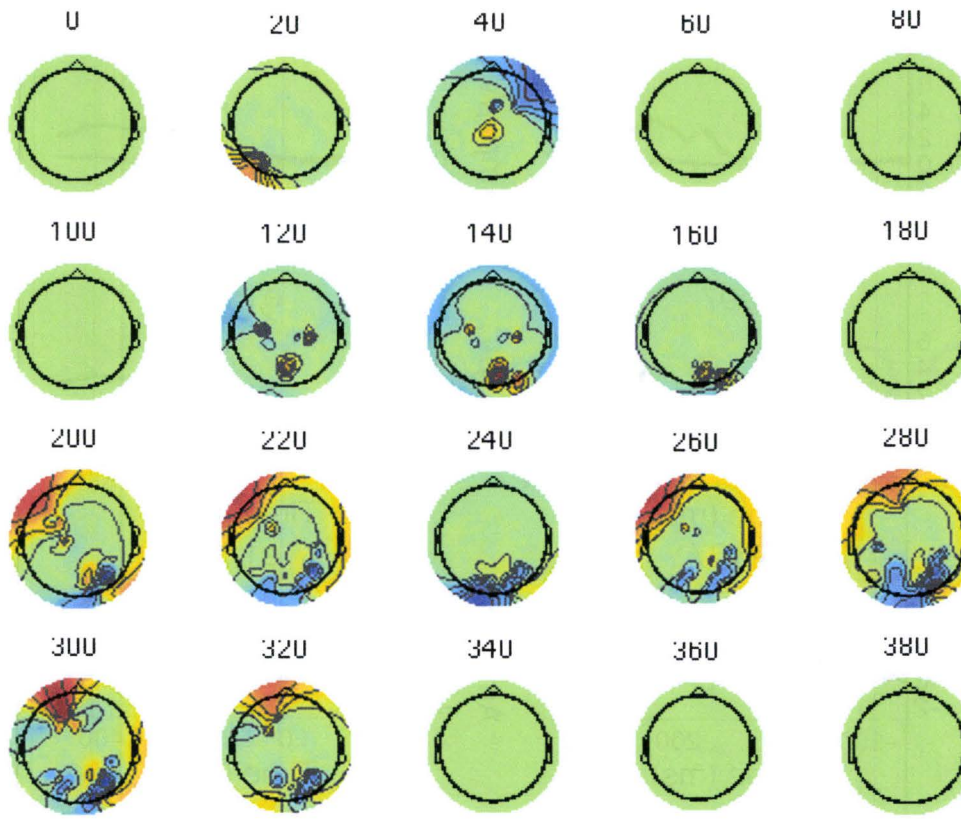

Figure 4.4: Results of t-tests that compared whether difference ERPs (Avg Days 4-5 minus Avg Days 1-2) were significantly greater than zero (at each electrode and time-point). Red = large positive t-scores; Blue = large negative t-scores; green = non-significant t-scores

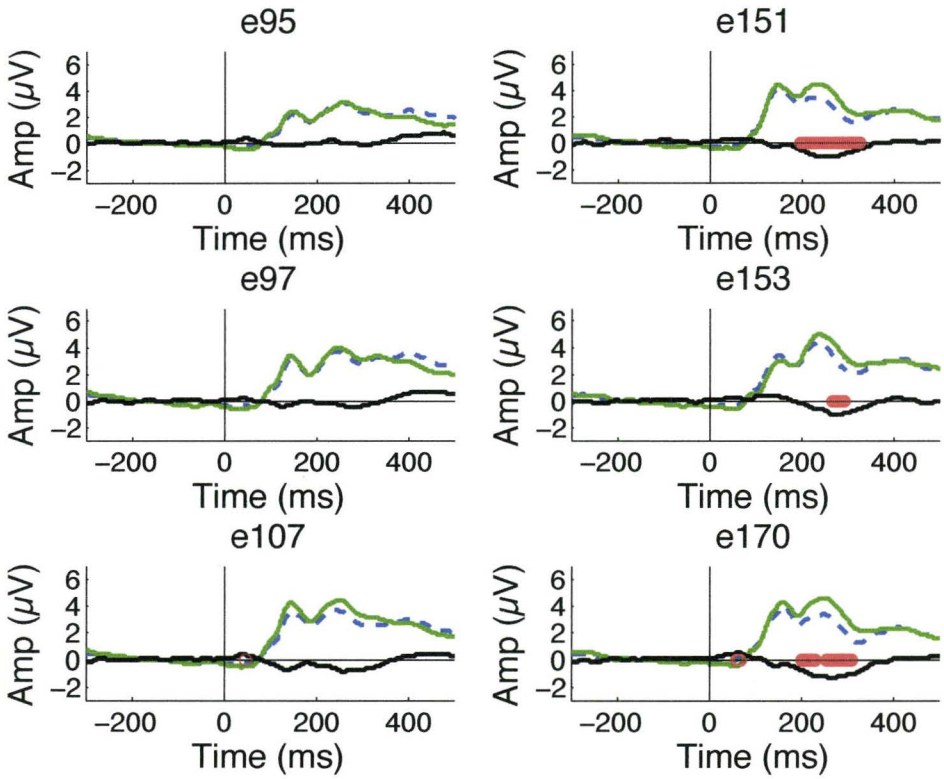

Figure 4.5: Sample electrodes from left hemisphere (left panel) and right hemisphere (right panel). ERPs of average of Days 1-2 (green solid lines) and Days 4-5 (blue dashed lines) are plotted as the average across subjects. Difference waves (Avg Days 4-5 minus Avg Days 1-2) are plotted (black solid lines). Time-points where the difference waves differ significantly from zero are plotted as red dots.

timing of these peaks varied across subjects and sessions. The P1 was generally located in the time-range of 130-170 ms, the N1 in the time-range of 170-238 ms, and the P2 in the time-range of 206-260 ms. The N250 component was not consistently present across subjects and sessions, so was not systematically investigated. Neither the amplitude nor the latency of any of any of the P1- N1- or P2- peaks differed significantly from the first to last session of training (Table 4.1).

Although the peak analysis did not reveal any significant effects of training, Figure 4.3 hints that amplitude in late parts of the ERP (i.e., > 200 ms) declined with training, particularly in the right hemisphere. To examine changes across the entire time-range of the ERP traces, the difference between the average of the first two training sessions and the average of the last two training sessions was calculated at each electrode and time-point. These differences were tested to determine whether they differed significantly from zero. The resulting t-values are presented in Figure 4.4. This figure presents topographic maps of the significant t-test values ($p < .01$) across the scalp plotted at 20 ms time intervals from 0-380 ms after stimulus onset. Values range from red (positive t-values) to blue (negative t-values). Training is associated with significant changes predominantly in the time-range of 200-320 ms. These differences are expressed as increases in the amplitude of left frontal electrodes, and decreases in the amplitude of right posterior electrodes. An additional early difference is visible as a decrease in the right frontal electrodes at approximately 40 ms after stimulus onset. To examine the time-course of these session differences in more detail, Figure 4.5 illustrates these differences at three electrodes in the posterior left and right hemispheres. These electrodes were chosen because they fall within the region of peak posterior activity for the P1-N1-P2 complex. The ERP and difference waves for these electrodes indicate a fall-off in amplitude after training that differs significantly (as indicated by red dots along x-axis) from zero at right hemisphere electrodes between approximately 200 and 300 ms after stimulus onset. The early part of this time-range overlaps with the peak of the N1 in some subjects. This time-range also overlap with the peak of the P2 for all subjects.

4.4.3 Partial Least Squares

To complement the ERP analysis, changes across the five training sessions were examined with partial least squares (PLS) (McIntosh et al., 1996; McIntosh and Lobaugh, 2004). PLS is a multivariate technique that determines the sets of weights that maximize the covariance between ERP activity and task conditions. The output of a PLS analysis is a series of latent variables, each of which corresponds to sets of weights (saliences) applied

to both the conditions (in this case, the training sessions) and the ERP responses (across time). The significance of each latent variable was assessed using permutation tests (500 samples). The reliability of the saliences for significant latent variables were assessed with bootstrap estimates of the standard error around the saliences (500 samples).

A PLS analysis that included each training session as a separate condition revealed a single significant latent variable ($p = .006$). This latent variable contrasted activity during the first two sessions of training against activity in the last two sessions (Figure 4.6). Activity on the third (middle) session contributed minimally to the contrast. These condition weights (design scores) are suggestive of a linear trend across sessions. Figure 4.7 illustrates the electrode saliences for this latent variable at two electrodes from the left and right hemispheres that typically form the focal points of N1 activity. The red dots along the x-axis indicate time-points that are reliable. Because these estimates of reliability are not the result of hypothesis tests, they have not been corrected for multiple comparisons. The plotted salience values in Figure 4.7 indicate the relative contribution of the ERP activity across time to the condition difference illustrated in Figure 4.6. For the right-hemisphere electrodes, the difference between the early and late sessions is maximal around 300 ms after stimulus onset. A summary of the electrode saliences across the entire scalp is shown in Figure 4.8, which plots the sum of the significant electrode salience time-points in 25 ms windows ranging from 0-400 ms after stimulus onset. Scalp points plotted in red indicate electrodes that strongly express the condition weights described in Figure 4.6 (i.e., electrodes where the amplitude falls of with progressive sessions). Scalp points plotted in blue indicate electrodes that strongly express the inverse of the condition weights (i.e., electrodes where amplitude increases with progressive sessions). The effect described by this latent variable appears to be concentrated in the posterior right hemisphere. Thus, both the ERP and PLS analyses suggest that training causes a decline in amplitude starting about 200 ms after stimulus onset.

Because ERP responses vary in both amplitude and latency across subjects, a second PLS analysis was conducted that factored out these cross-subject differences. Difference ERP waves were calculated relative to the first training session, for each of the subsequent training sessions. A PLS analysis was then conducted with training sessions 2-5 as conditions, and the difference waves (always relative to training session 1) as brain activity. Even more clearly than the results of the first PLS analysis, the condition weights for this latent variable (Figure 4.9) describe a decreasing linear trend across sessions, indicating that the amplitudes of the difference ERPs declined approximately linearly across

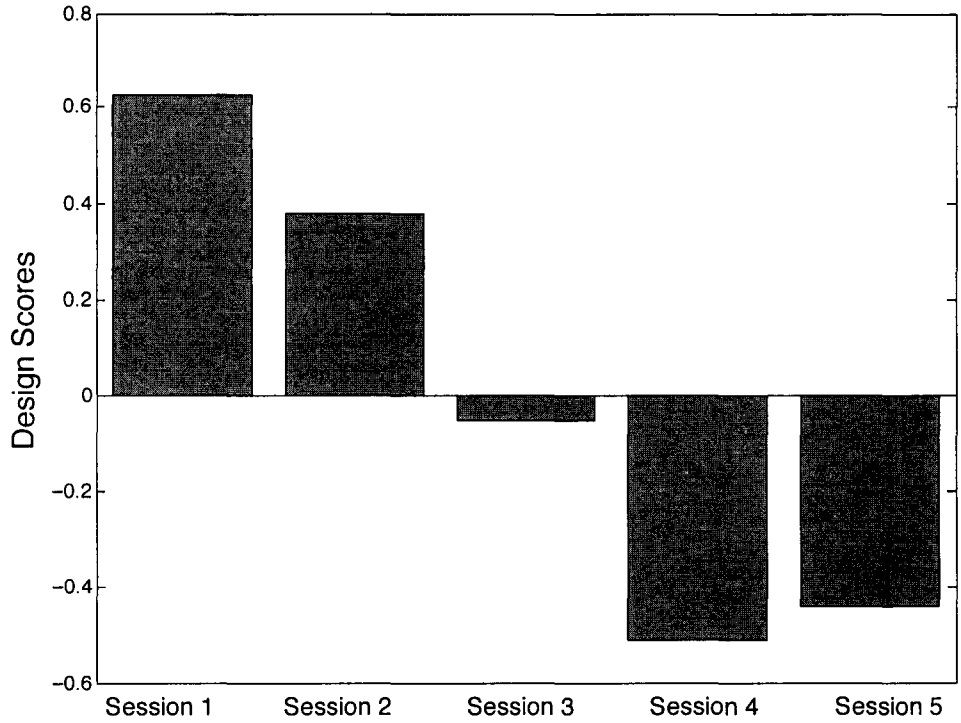

Figure 4.6: PLS analysis: Training sessions. Condition weights for the significant latent variable: The bars represent weights in a linear contrast that best distinguish between the five conditions (sessions).

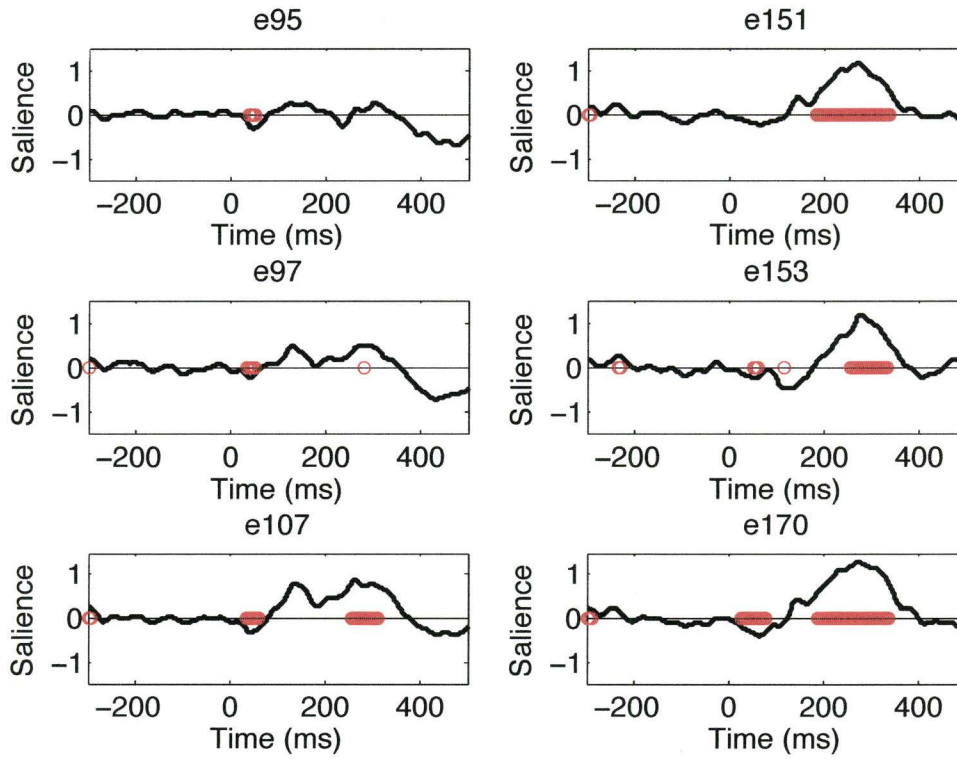

Figure 4.7: PLS analysis: Training sessions. Saliences at two electrodes that typically exhibit large N170 amplitudes in the left (1st column) and right (2nd column) hemispheres. Red dots indicate time-points where saliences have reliably non-zero values (as assessed by a 500 sample bootstrap).

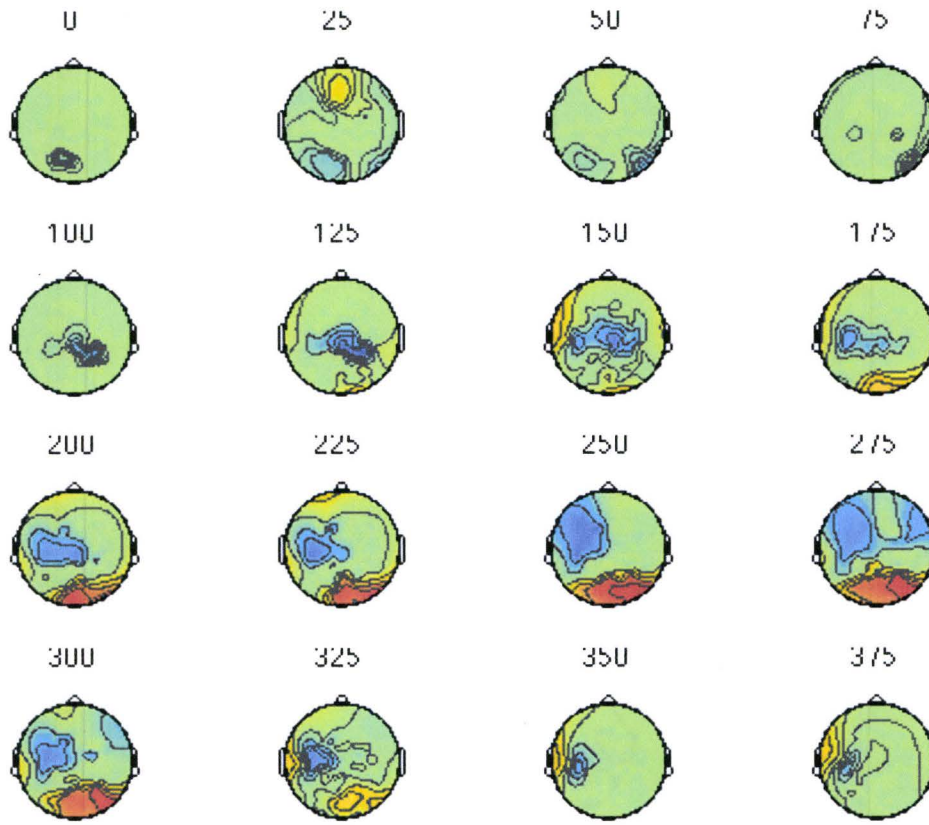

Figure 4.8: PLS analysis: Training sessions. A summary of electrode saliences across the scalp. The time points at which the salience was reliably non-zero were summed within 25 ms windows (from 0-400 ms) for each electrode. The resulting window sums are plotted across the scalp: red values indicate electrodes that strongly exhibit the pattern of the condition weights in the given time-window (i.e., where amplitude falls off with training); blue values indicate electrodes that strongly exhibit the reverse pattern of condition weights during the time-window (i.e., where amplitude rises with training).

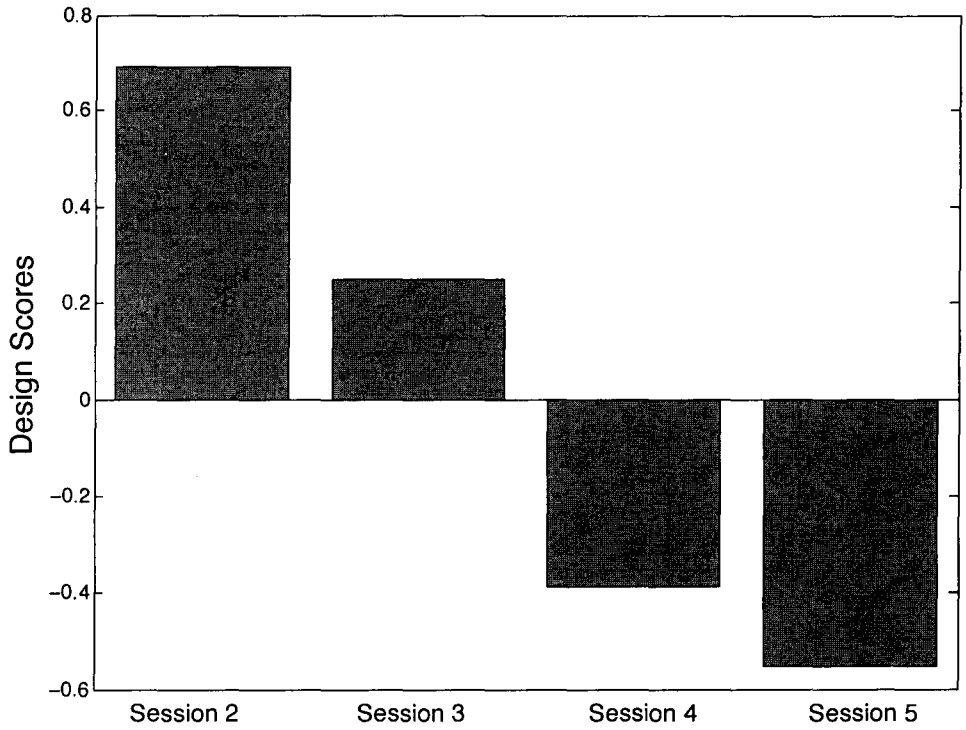

Figure 4.9: PLS analysis: Difference from first day of training. Condition weights for the significant latent variable: The bars represent weights in a linear contrast that best distinguish between the five conditions (sessions).

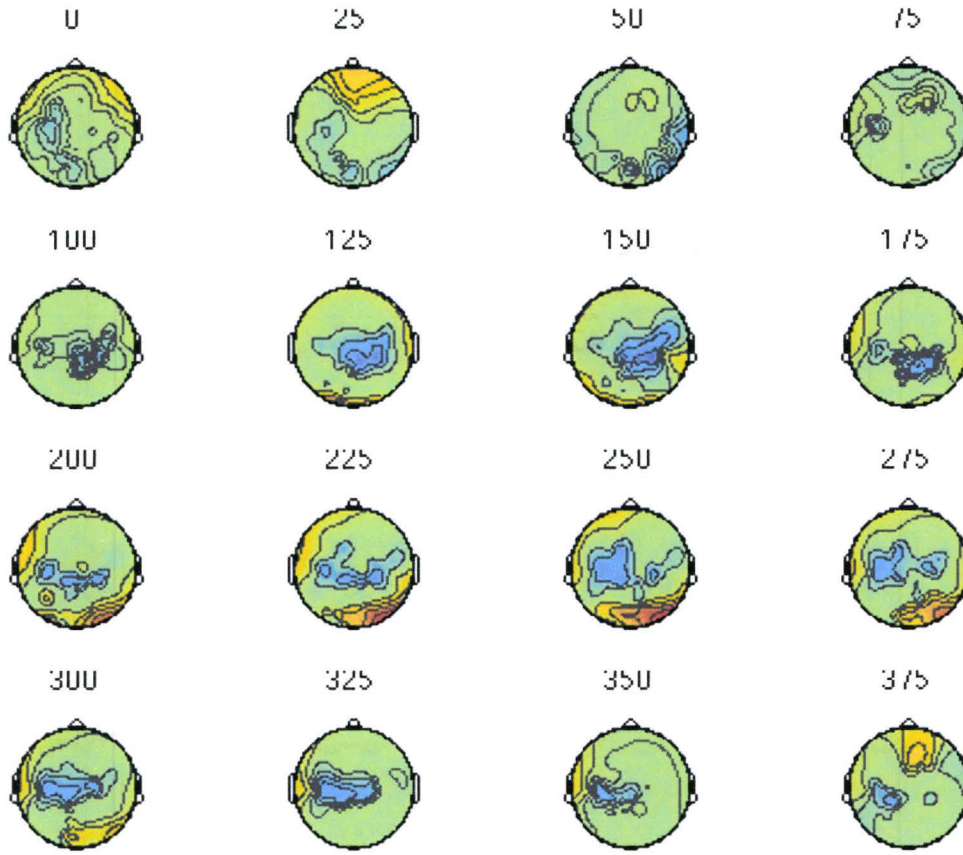

Figure 4.10: PLS analysis: Difference from first day of training. To summarize the corresponding electrode saliences across the scalp, the time points at which the salience was reliably non-zero were summed within 25 ms windows (from 0-400 ms) for each electrode. The resulting window sums are plotted across the scalp: red values indicate electrodes that strongly exhibit the pattern of the condition weights in the given time-window (i.e., where amplitude falls off with training); blue values indicate electrodes that strongly exhibit the reverse pattern of condition weights during the time-window (i.e., where amplitude rises with training).

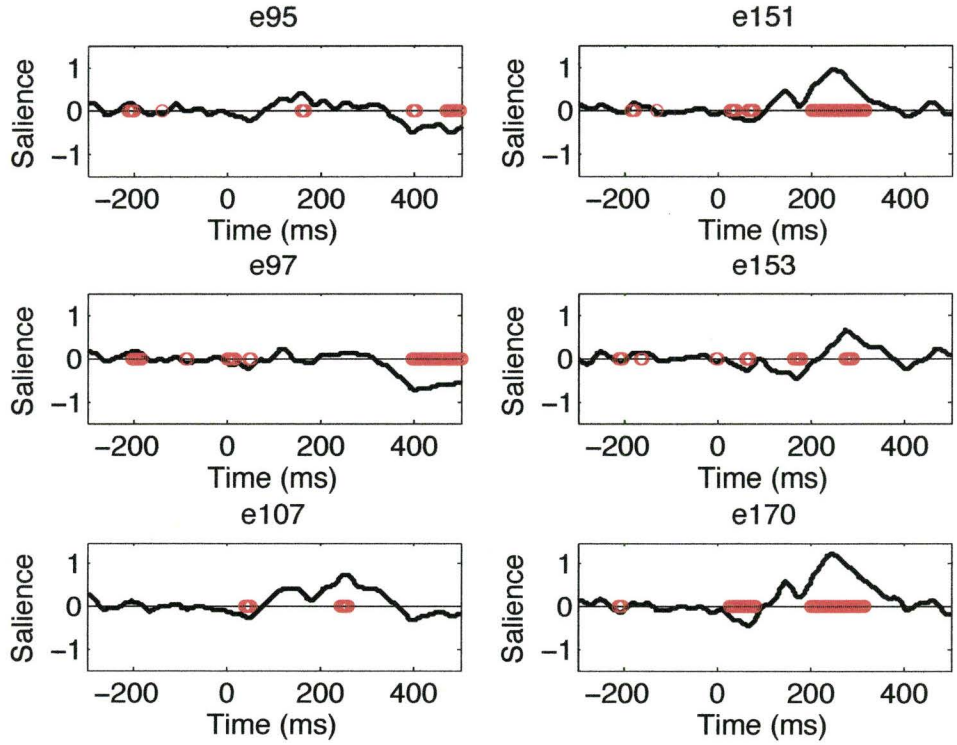

Figure 4.11: PLS analysis: Difference from first day of training. Electrode saliences for latent variable 1 for the left (1st column) and right (2nd column) hemispheres. Electrodes are typical peak N170 electrodes.

training sessions. Figure 4.10 indicates the electrodes and time window within which these effects are reliable. The strongest effects (as depicted by regions of red on the scalp) are centered around the right posterior electrodes in the time window of 200-300 ms, which is consistent with the results obtained with the previous PLS analysis. Figure 4.11 illustrates the time-course of the latent variable for sample posterior electrodes in the left and right hemisphere. Note that saliences are again, more reliable for right than left-hemisphere electrodes.

Training effects on the ERP signatures may be modulated by the amount of learning exhibited by subjects. To examine this possibility, we employed behavioural PLS analyses. Unlike Task PLS analyses that contrast brain activity (ERPs) across experimental conditions (sessions), behavioural PLS analyses contrast the correlation between brain and behaviour across different experimental conditions (McIntosh and Lobaugh, 2004). Separate behavioural PLS analyses were performed on the accuracy and RT data. Figure 4.14 plots the design scores across sessions, as well as the correlations across sessions. The design scores indicate the strength of the expression of the task/behaviour correlations across sessions, and can be thought of as contrasts across the conditions. In this case, the design scores are positive for all sessions, indicating that a similar correlation is being expressed across sessions (note: unlike Task PLS, Behavioural PLS analyses are not mean-centred, therefore the design scores do not necessarily sum to zero). In particular, both the accuracy and reaction time correlations appear to change only moderately across sessions (Figure 4.14), with correlations dropping off somewhat by the last day of training. Like the task PLS results, the correlations between ERP and accuracy (Figure 4.12) were most reliable in the time range of 200-300 ms in the posterior right hemisphere. By contrast, the ERP reaction time correlations (Figure 4.13) were most reliable at earlier latencies, centred around 100 ms, and most strongly evident in the posterior left hemisphere.

4.5 Discussion

Five days of training on a visual house discrimination task resulted in substantial learning as measured by both accuracy and reaction time measures. Both response measures changed substantially across the first couple sessions, and then continued to change at a slower rate for the remaining training sessions. These behavioural changes were accompanied by changes in the ERP traces: amplitude within a time window of 200-300 ms after stimulus onset decreased significantly from early to late sessions. This

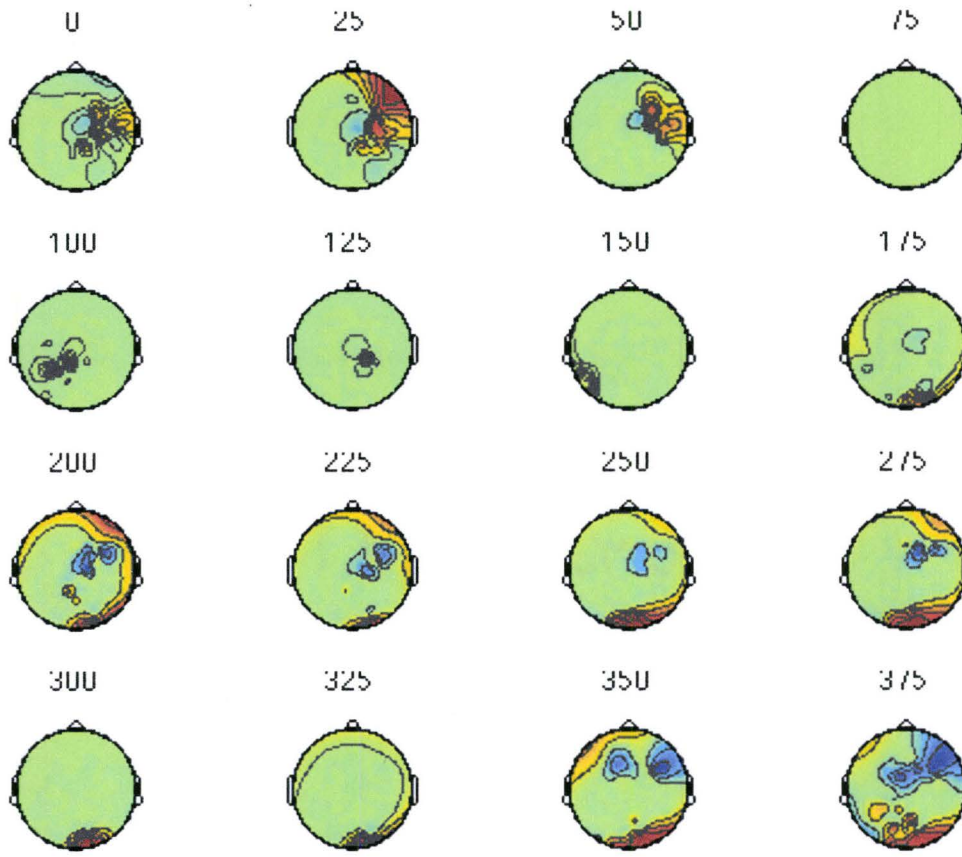

Figure 4.12: Behavioural PLS analysis (Accuracy): Difference from first day of training. To summarize the corresponding electrode saliences across the scalp, the time points at which the salience was reliably non-zero were summed within 25 ms windows (from 0-400 ms) for each electrode. The resulting window sums are plotted across the scalp: red values indicate electrodes that strongly exhibit the pattern of the condition weights in the given time-window (i.e., positive correlations with accuracy); blue values indicate electrodes that strongly exhibit the reverse pattern.

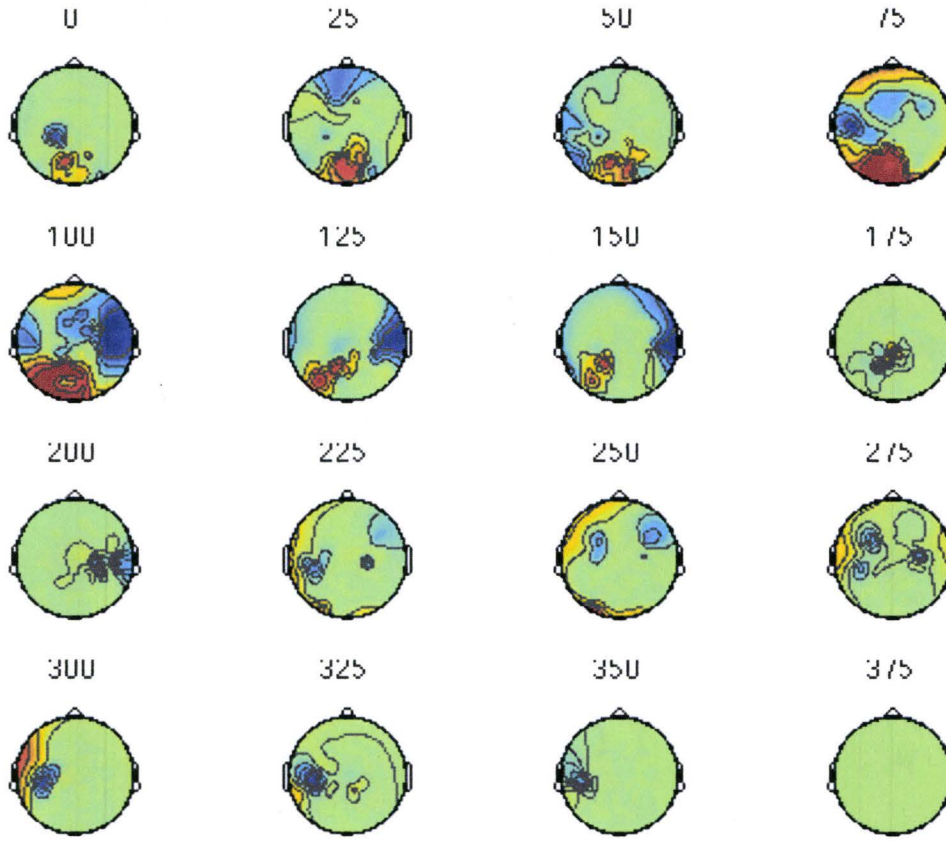

Figure 4.13: Behavioural PLS analysis (RT): Difference from first day of training. To summarize the corresponding electrode saliences across the scalp, the time points at which the salience was reliably non-zero were summed within 25 ms windows (from 0-400 ms) for each electrode. The resulting window sums are plotted across the scalp: red values indicate electrodes that strongly exhibit the pattern of the condition weights in the given time-window (i.e., negative correlations with reaction time); blue values indicate electrodes that strongly exhibit the reverse pattern.

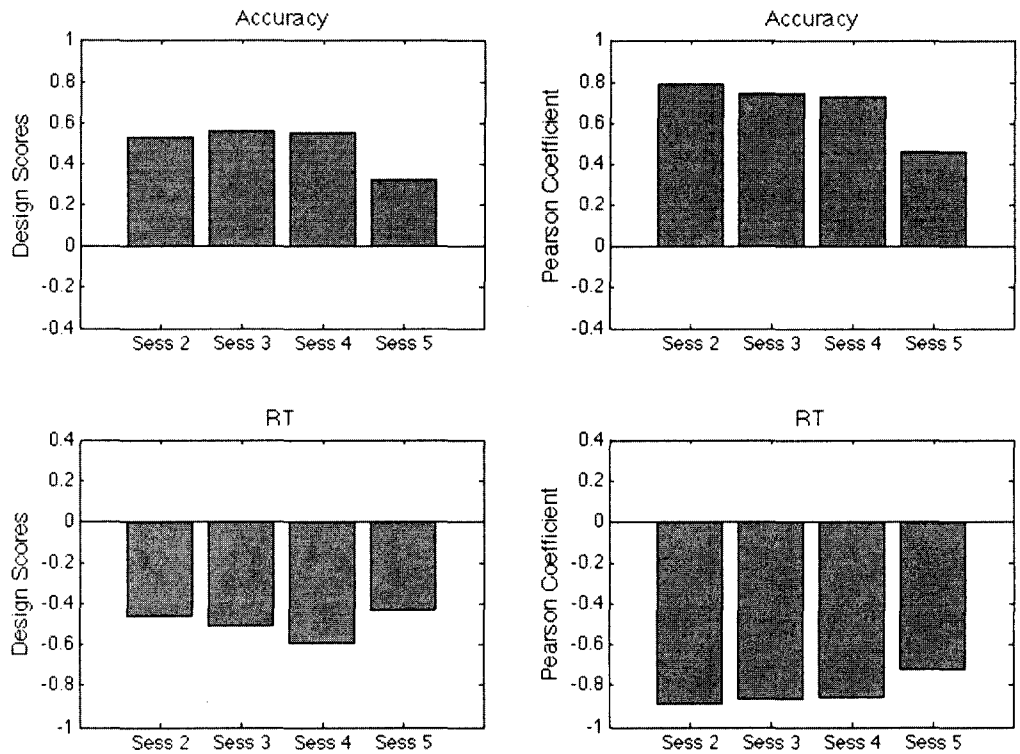

Figure 4.14: Behavioural PLS analysis (Accuracy & RT): Difference from first day of training. The latent variable is described by condition weights and electrode saliences across time.

decrease was approximately linear across training sessions, particularly when measured relative to performance on the first day of training. Throughout training, changes in the early components ERP activity (i.e., $\approx$ 100 ms after stimulus onset), measured relative to the first session, were correlated with decreased reaction time, whereas changes in the later components (i.e., 200-300 ms) were correlated with increased response accuracy.

We did not find any evidence that N1 amplitude increased across training sessions. Indeed, we found the opposite: that amplitude decreased in the time range of the N1 and P2. This result is surprising in the context of previous studies that have reported an increase in the N1 with the acquirement of task expertise (Gauthier et al., 2003; Tanaka and Curran, 2001; Wong et al., 2005; Scott et al., 2006, 2008), and in some cases, with an increase in the N250 (Scott et al., 2006, 2008). Scott et al. (2006) found the N250 component was especially sensitive to the training of individual exemplars, therefore, we might have expected similar findings in the context of learning to discriminate amongst houses. We were unable to explicitly examine training effects on this component because most subjects in the current study did not elicit clear N250 responses. However, our results are inconsistent with an increase in the N250 with training, as we found that amplitude decreased universally within the time window of 200-300 ms.

The current study differed in at least two potentially important ways from those studies that reported increased N1 and N250 responses with training. The first difference was stimulus set size: the current study examined the effects of training subjects to become highly familiar with a small number of stimuli, rather than the approach taken in these other expertise studies of presenting a larger number of less-familiar stimuli from a highly trained class. This explanation would be consistent with previous reports of decreased N1 amplitudes for familiar relative to novel stimuli (Guillaume and Tiberghien, 2001; Itier and Taylor, 2002), and of decreased N1 responses after immediate stimulus repetitions (e.g., Campanella et al., 2000; Heisz et al., 2006).

The current study also differed from most other expertise studies in making use of a single 4-AFC visual discrimination task. By contrast, Scott et al. (2006, 2008) had subjects learn names for individual items and Gauthier et al. (2003) trained subjects in several cognitive tasks. The type of training can have a substantial impact on whether the N1 is modulated by training. For example, in a visuo-tactile task, subjects were either trained to pantomime interactions with the objects, or to point at the action-regions (e.g., handles) of the object (Kiefer et al., 2007). Only the pantomime training group resulted in larger N1 responses to objects that shared common functional interactions

(e.g., placement of handles), than objects that shared a common overall shape. No effect of training was observed for the pointing group.

The use of a single, repetitive, visual training task is more commonly employed in lower-level perceptual learning studies, and EEG changes with learning have also been examined in the context of this type of training task. For example, EEG training effects have been studied in the context of vernier acuity judgments (Skrandies et al., 1996; Ludwig and Skrandies, 2002; Skrandies et al., 2001; Shoji and Skrandies, 2006), and judgments of line orientation (Ding et al., 2003; Song et al., 2005), grating orientation (Song et al., 2007), and arrow direction (Song et al., 2002, 2005). Our results are consistent with reports of decreased N1 amplitude in the context of learning line and grating orientation judgments, and arrow direction judgments (Song et al., 2002; Ding et al., 2003; Song et al., 2005), but are inconsistent with learning in the context of vernier acuity judgments, where negative components occurring around 200-300 ms increased with learning (Skrandies et al., 1996). Later positive components have generally also exhibited an increase in amplitude with training (Song et al., 2002; Ding et al., 2003; Shoji and Skrandies, 2006; Landis et al., 1984), but a decrease similar to that observed in the current study has been reported in the context of line orientation judgments (Song et al., 2005).

We also found correlations between behaviour and ERP amplitude: early time ranges were related to reaction time, while later time ranges were related to response accuracy. These relations held true throughout training. The differential timing of the reaction time and accuracy correlations suggests the strength of early neural responses contribute to the speed of object recognition, while later time ranges (corresponding typically to the P2/P3 components) contribute more to task accuracy. The late effects of accuracy are consistent with other findings that stimulus familiarity tends to have greater impacts on later components (e.g., Eimer, 2000b; Bentin and Deouell, 2000, but see Jacques and Rossion (2006)).

We know of no other studies that have examined these types of accuracy/RT correlations across training sessions, however a similar correlation between behaviour and P1/N1 strength has been reported in the context of change detection task (Curran et al., 2009). In that case, an amalgamated measure of the P1 and N1 components was correlated with accuracy. It is interesting to note that the correlations reported by Curran et al. (2009) were similarly unaffected by expertise: the relation between accuracy and signal strength was comparable across both expert image analysts and novice participants. Thus, while

training affects overall signal strength, it does not appear to alter the relation between signal strength and behavioural measures.

The current results contribute to a growing literature examining electroencephalographic changes across training, and is the first that we know of to track ERP responses across a large sequence of training sessions. The amplitude decrease with learning that we have reported here is consistent with some reports from other visual training paradigms, but inconsistent with the typical reports of increased amplitude responses for field experts compared to novices, a discrepancy that might be explained by differences in familiarity with the presented stimuli between lab-trained experts and real-world experts.

References

Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience 8 (6), 551–565.

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

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

Campanella, S., Hanoteau, C., Depy, D., Rossion, B., Bruyer, R., Crommelinck, M., Guerit, J., 2000. Right N170 modulation in a face discrimination task: An account for categorical perception of familiar faces. Psychophysiology 37 (6), 796–806.

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

Curran, T., Gibson, L., Horne, J. H., Young, B., Bozell, A. P., 2009. Expert image analysts show enhanced visual processing in change detection. Psychon Bull Rev 16 (2), 390–397.

Curran, T., Tanaka, J. W., Weiskopf, D. M., 2002 Mar. An electrophysiological comparison of visual categorization and recognition memory. Cogn Affect Behav Neurosci 2 (1), 1–18.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-

trial EEG dynamics including independent component analysis. J Neurosci Methods 134 (1), 9–21.

Ding, Y., Song, Y., Fan, S., Qu, Z., Chen, L., 2003. Specificity and generalization of visual perceptual learning in humans: an event-related potential study. Neuroreport 14 (4), 587–590.

Eimer, M., 2000a. Attentional modulations of event-related brain potentials sensitive to faces. Cognitive Neuropsychology 17 (1-3), 103–116.

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

Gauthier, I., Curran, T., Curby, K. M., Collins, D., 2003. Perceptual interference supports a non-modular account of face processing. Nat Neurosci 6 (4), 428–432.

Guillaume, F., Tiberghien, G., 2001. An event-related potential study of contextual modifications in a face recognition task. Neuroreport 12 (6), 1209–1216.

Heisz, J. J., Watter, S., Shedden, J. M., 2006. Progressive N170 habituation to unattended repeated faces. Vision Res 46 (1-2), 47–56.

Husk, J. S., Bennett, P. J., Sekuler, A. B., 2007. Inverting houses and textures: investigating the characteristics of learned inversion effects. Vision Res 47 (27), 3350–3359.

Itier, R. J., Latinus, M., Taylor, M. J., 2006. Face, eye and object early processing: what is the face specificity? Neuroimage 29 (2), 667–676.

Itier, R. J., Taylor, M. J., 2002. Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using erps. Neuroimage 15 (2), 353–372.

Itier, R. J., Taylor, M. J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cereb Cortex 14 (2), 132–142.

Ito, T. A., Thompson, E., Cacioppo, J. T., 2004. Tracking the timecourse of social perception: the effects of racial cues on event-related brain potentials. Pers Soc Psychol Bull 30 (10), 1267–1280.

Ito, T. A., Urland, G. R., 2005. The influence of processing objectives on the perception of faces: an ERP study of race and gender perception. Cogn Affect Behav Neurosci 5 (1), 21–36.

Jacques, C., Rossion, B., 2006. The speed of individual face categorization. Psychol Sci 17 (6), 485–492.

Jeffreys, D. A., Tukmachi, E. S., 1992. The vertex-positive scalp potential evoked by faces and by objects. Exp Brain Res 91 (2), 340–350.

Kiefer, M., Sim, E.-J., Liebich, S., Hauk, O., Tanaka, J., 2007. Experience-dependent plasticity of conceptual representations in human sensory-motor areas. J Cogn Neurosci 19 (3), 525–542.

Landis, T., Lehmann, D., Mita, T., Skrandies, W., 1984. Evoked potential correlates of figure and ground. Int J Psychophysiol 1 (4), 345–348.

Ludwig, I., Skrandies, W., 2002. Human perceptual learning in the peripheral visual field: sensory thresholds and neurophysiological correlates. Biol Psychol 59 (3), 187–206.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., Grady, C. L., 1996. Spatial pattern analysis of functional brain images using partial least squares. Neuroimage 3 (3 Pt 1), 143–157.

McIntosh, A. R., Lobaugh, N. J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 23 Suppl 1, S250–63.

Peissig, J. J., Singer, J., Kawasaki, K., Sheinberg, D. L., 2007. Effects of long-term object familiarity on event-related potentials in the monkey. Cereb Cortex 17 (6), 1323–1334.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., Crommelinck, M., 2002. Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. Psychol Sci 13 (3), 250–257.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., Crommelinck, M., 2000. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. Neuroreport 11 (1), 69–74.

Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B., 2007. Single-trial EEG dynamics of object and face visual processing. Neuroimage 36 (3), 843–862.

Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B., 2008. Time course and robustness of ERP object and face differences. J Vis 8 (12), 3.1–18.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., Curran, T., 2006. A reevaluation of the electrophysiological correlates of expert object processing. J Cogn Neurosci 18 (9), 1453–1465.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., Curran, T., 2008. The role of category learning in the acquisition and retention of perceptual expertise: a behavioral and neurophysiological study. Brain Res 1210, 204–215.

Shoji, H., Skrandies, W., 2006. ERP topography and human perceptual learning in the peripheral visual field. Int J Psychophysiol 61 (2), 179–187.

Skrandies, W., Jedynak, A., Fahle, M., 2001. Perceptual learning: psychophysical thresholds and electrical brain topography. Int J Psychophysiol 41 (2), 119–129.

Skrandies, W., Lang, G., Jedynak, A., 1996. Sensory thresholds and neurophysiological correlates of human perceptual learning. Spat Vis 9 (4), 475–489.

Song, Y., Ding, Y., Fan, S., Chen, L., 2002. An event-related potential study on visual perceptual learning under short-term and long-term training conditions. Neuroreport 13 (16), 2053–2057.

Song, Y., Ding, Y., Fan, S., Qu, Z., Xu, L., Lu, C., Peng, D., 2005. Neural substrates of visual perceptual learning of simple and complex stimuli. Clin Neurophysiol 116 (3), 632–639.

Song, Y., Peng, D., Lu, C., Liu, C., Li, X., Liu, P., Qu, Z., Ding, Y., 2007. An event-related potential study on perceptual learning in grating orientation discrimination. Neuroreport 18 (9), 945–948.

Speer, N. K., Curran, T., 2007. ERP correlates of familiarity and recollection processes in visual associative recognition. Brain Res 1174, 97–109.

Stahl, J., Wiese, H., Schweinberger, S. R., 2008. Expertise and own-race bias in face processing: an event-related potential study. Neuroreport 19 (5), 583–587.

Tanaka, J. W., Curran, T., 2001. A neural basis for expert object recognition. Psychol Sci 12 (1), 43–47.

Tanaka, J. W., Curran, T., Porterfield, A. L., Collins, D., 2006. Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. J Cogn Neurosci 18 (9), 1488–1497.

Wong, A. C. N., Gauthier, I., Woroch, B., DeBuse, C., Curran, T., 2005. An early electrophysiological response associated with expertise in letter perception. Cogn Affect Behav Neurosci 5 (3), 306–318.

Chapter 5

ERP differences between faces and houses are unaffected by house-identity training

5.1 Abstract

Previously we demonstrated that training subjects to identify houses decreases the amplitude of the ERP signals in the time-range of 200-300 ms. Here, we present data collected before and after these training sessions that compare the ERP responses to the trained houses and other object classes (including untrained houses, faces and objects). Session differences were observed for all classes of stimuli. Amplitude increased across sessions at late time ranges, selectively for the untrained stimuli, resulting in an increase in the difference between the ERP signatures of the trained houses and faces after training. No training-related amplitude changes were observed in the time-range of the N1, however the N1 response decreased marginally in latency in the right hemisphere for trained houses, suggesting that training resulted in slightly faster processing of the stimuli.

5.2 Introduction

Faces elicit a large negative event-related potential response about 170 ms after stimulus onset (e.g., Bentin et al., 1996; Eimer, 2000a; Itier and Taylor, 2004; Jeffreys and Tukmachi, 1992). Despite the reliability of this N1 response, its function remains unclear. The N1 can be elicited by a wide range of scenes and objects, in addition to faces, but

tends to be larger and earlier for faces than for other object categories. An interesting exception to this general trend is the N1 response to eyes presented in isolation, which actually is larger than the N1 produced by whole faces (Bentin et al., 1996; Taylor et al., 2001). Despite the amplitude and latency differences between the N1 for objects and for faces, there is some evidence that the N1 responses elicited by objects are produced by the same mechanisms that produce the N1 responses evoked by faces (Rossion et al., 2007, 2004). If the same mechanisms are responsible for both the face N1 and the object N1, then the face N1 may only differ quantitatively, not qualitatively, from the object N1. For example, the face N1 and object N1 might differ only as a function of expertise.

Early studies found that similar N1 responses were evoked by unfamiliar and famous faces (Eimer, 2000a,b; Bentin and Deouell, 2000), which suggested that the N1 is not affected by stimulus familiarity. More recent findings challenge this view. For example, observers exhibit larger N1 responses to human faces than to ape faces (Carmel and Bentin, 2002), and experts exhibit larger N1 responses to objects drawn from their area of expertise (Tanaka and Curran, 2001; Gauthier et al., 2003). Other studies have shown that practice in object categorization tasks can alter the N1 evoked by objects drawn from the class of objects seen during training (Scott et al., 2006, 2008), although it is unclear if such training results in N1 responses that are as large as those produced by faces. If larger, earlier N1 responses are produced as a result of expertise, one might expect the ERP response of trained objects to similarly increase in amplitude and decrease in latency, reducing the difference between the face and object N1 responses.

We recently demonstrated that five days of practice in a house discrimination task systematically alters the amplitude of the ERP response evoked by houses (Chapter 4). Specifically, training reduced the amplitude of the ERP 200-300 ms after stimulus onset, and the effect increased approximately linearly across the five days of practice. Unlike previous training studies (Scott et al., 2006, 2008), we found no evidence for an increase in N1 amplitude. However, the training task we employed was unusual for ERP studies. For example, the stimuli were presented at low contrast, and contrast is known to affect the strength of ERP signals (Mace et al., 2005). Furthermore, the ERP measurements were made in the context of a 4-AFC discrimination task, rather than a matching task with simpler go/no-go response demands, as is more typically employed in ERP studies.

Here, we examine the effects of training on performance in a one-back matching task that was administered before and after the previously-reported house-training sessions. Unlike the 4-AFC task used during training, the one-back task used high-contrast pat-

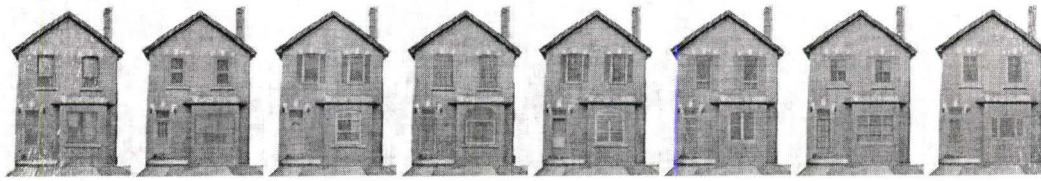

Figure 5.1: Training houses.

terns drawn from several sets of stimuli. If the larger N1 response to faces is a result of expertise, we would expect that N1 responses to the trained houses should increase with training, and that the difference ERP between faces and houses should decrease after training. Previous work has shown that the effects of training in a house discrimination task partially transfer to novel exemplars of the training set (Chapter 2). Here, ERP effects of learning might also be expected to show partial transfer of learning to the untrained similar houses.

5.3 Methods

5.3.1 Subjects

Eight subjects (mean age = 23.1; range = 20-27) participated in this experiment. All subjects were right handed and had normal or corrected-to-normal acuity. Six of the subjects were male. Subjects were paid $10/h for participating in the experiment.

5.3.2 Stimuli

The training stimuli were a set of 8 houses (Figure 5.1). All of the houses consisted of a single door, a single lower window, and a pair of upper windows, all placed within a common outer frame taken from one house. Individual houses were constructed by using different doors and windows for each exemplar, and the resulting items were visually similar. See (Chapter 2) for additional details about the stimuli. Training stimuli were presented at low contrast (contrast variance = 0.001).

The pre-training and post-training one-back task used six stimulus sets (see examples in Figure 5.2): the training house set, consisting of items that were seen during training, an untrained house set consisting of novel exemplars of houses that were similar to the trained set, a second set of untrained houses that differed greatly in appearance from the trained set (from Chapter 2), two sets of faces (one that retained external features

Figure 5.2: Example stimuli from the one-back task. From left to right: Trained houses, similar untrained houses, faces 1, faces 2, dissimilar untrained houses, objects.

such as hair and original face contour (from O'Craven et al., 1999), and another set of different faces that excluded these features by cropping the faces within a common oval frame (from Gold et al., 1999), and a set of miscellaneous common objects (drawn from an online database provided by Michael J. Tarr (Brown University, http://ww.tarrlab.org/). These six stimulus sets were selected to parallel the experimental paradigm employed in a complementary functional magnetic resonance imaging analysis of this training task (see Chapter 3). One-back stimuli were presented at high contrast (contrast variance = 0.01).

5.3.3 Apparatus

All stimuli were presented on a Sony Trinitron GDM-F520 monitor (resolution: 800 x 600 pixels; 20 pixels/cm; refresh rate: 85 Hz) in a dimly-lit, sound-attenuated booth. Average luminance was 41.8 cd/m^2. Subjects viewed the display binocularly from a distance of 1 m. A chin rest was used to stabilize the viewing position.

Electrophysiological recordings were collected using the 256 channel Geodesic sensor net system (Electrical Geodesics Inc., Eugene, Oregon; Tucker, 1993), and acquisition software NetStation v4.2.4. The ground electrode was located along the midline, anterior to Fz. Electrode impedances were maintained within 50 micro-ohms. The analog signal was digitized at 500 Hz and band-pass filtered between 0.1 and 200 Hz. EEG signal was referenced online to Cz, but re-referenced offline to the average across electrodes. The re-referenced EEG was band-pass filtered between 1 and 30 Hz. Bad channels were identified using routines in BESA 5.0, then removed from further processing. Baseline correction was performed using a baseline of 300 ms, and artifacts were rejected, removing any trials where the amplitude exceeded +/- 100 microvolts, or any trials wherein the amplitude changed by more than 75 microvolts across two consecutive time points. Only correct trials were retained for further analysis.

EEG analyses were performed using EEGLAB (Delorme and Makeig, 2004) with extraction of ERPs and additional statistical testing conducted with additional in-lab Matlab routines. Partial least squares (PLS) analyses were performed with Matlab routines PLSgui (McIntosh and Lobaugh, 2004).

5.3.4 Procedure

5.3.4.1 Experiment Overview

The data reported in this study were collected in the context of a 7-session training study. During the first and last session subjects completed the one-back matching task, and during the intervening five sessions they completed the 4-AFC house discrimination training task. The current paper presents the data from the first and last session of the 7-session study (for detailed methods and results from the intervening training sessions, see Chapter 4). At the start of each session, subjects were instructed to minimize blinking and head movements.

5.3.4.2 Training Procedure

The training procedure has been described in detail elsewhere (Chapter 4), but will be summarized here briefly. The training task was a 4-AFC discrimination task. The target house was presented for 500 ms, after which subjects were given unlimited time to select amongst four house exemplars, one of which was a target-match. On each trial, the target and distractors were chosen randomly from the set of eight houses, with the constraint that each of the eight houses were viewed an equal number of times within each session. Task difficulty was controlled by presenting the target houses at low contrast (contrast variance = 0.001). Subjects completed 416 trials with auditory feedback on each of five days.

5.3.4.3 One-back Procedure

The one-back matching task was designed to allow comparison of EEG signals across object categories before and after the training sessions. As such, the task used the training stimuli (houses) and comparison stimuli from a variety of other object categories. The six conditions were block randomized, with 10 blocks per condition for a total of 60 blocks per session. Each block comprised 24 trials, wherein each trial started with approximately 1.5 second of fixation, followed immediately by the 500 ms stimulus presentation. The exact

duration of the fixation period was jittered randomly from 1.14 to 1.85 s to ensure some unpredictability in the appearance of the stimulus, resulting in a total trial duration of 1.64-2.35 s. Within each block, target events (i.e., immediate repetitions of the previous stimulus) occurred on one quarter of the trials. Subjects were instructed to press a key to indicate the presence of these target events. Responses were accepted between stimulus onset and the onset of the following stimulus (i.e. responses were still accepted through the fixation period leading to the following stimulus; there was no inter-stimulus interval aside from fixation), and a failure to respond during this interval was recorded as a miss.

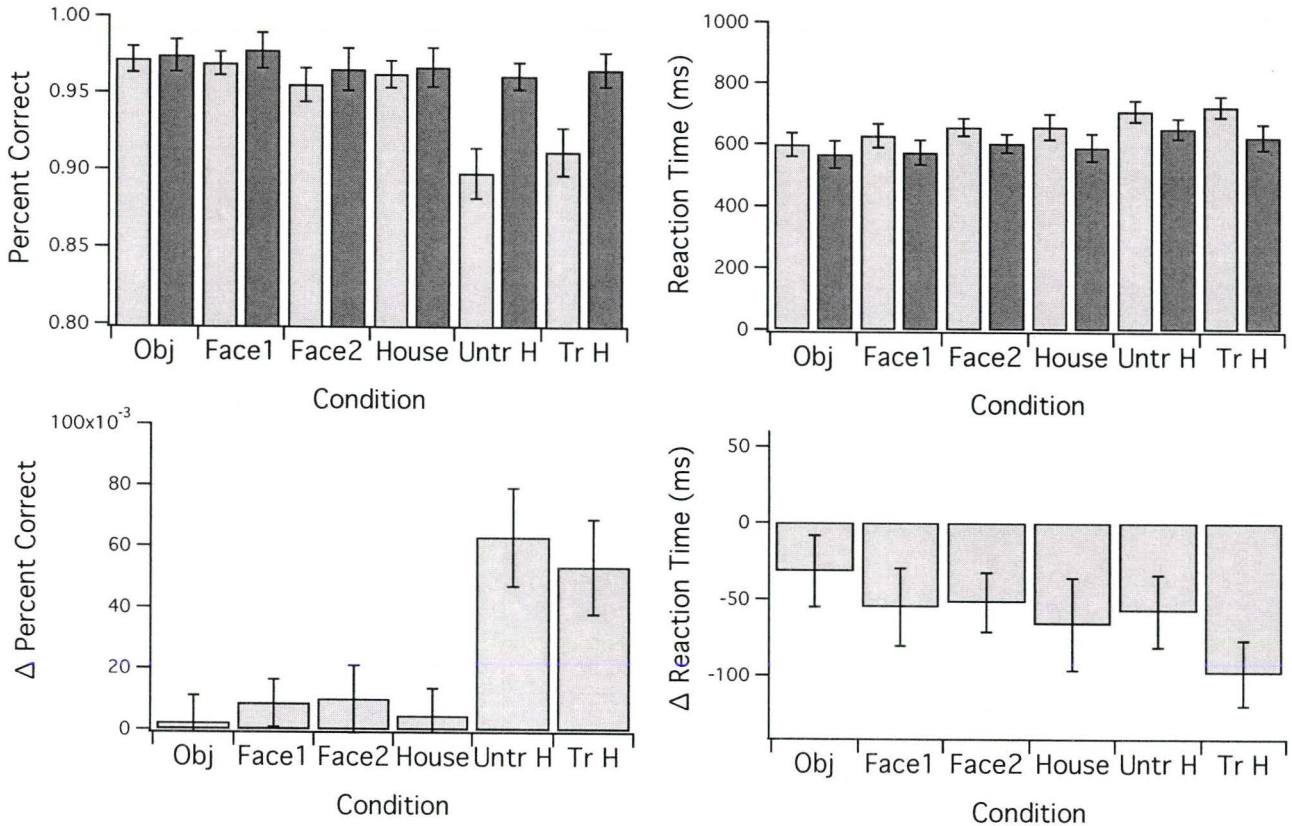

Figure 5.3: (a) Accuracy and (b) reaction time on the one-back task: Pre- and post-training results are shown by light and dark bars, respectively. Improvement across sessions (post-training minus pre-training) is plotted for (c) accuracy and (d) reaction time. Error bars represent +/- 1 standard error of the mean.

5.4 Results

5.4.1 One-back behaviour

As expected, performance in the one-back task was near ceiling for nearly all sets of stimuli in both the pre- and post-test sessions (Figure 5.3a). Accuracy for the trained houses and the similar untrained house set was lower than for the other sets, though accuracy was still quite high (i.e., $\approx 90\%$ correct for both sets in the pre-training session). For these two sets of houses, accuracy increased significantly from pre-test to post-test (trained houses: $t(7) = 2.43, p = .045$; untrained houses: $t(7) = 2.98, p = .021$; Figure 5.3c) and reaction time decrease significantly from pre- to post-test (trained houses: $t(7) = -2.44, p = .045; t(7) = -4.61, p = .002$; Figure 5.3b). The decrease in reaction time was larger for the trained house set than for the untrained house set ($t(7) = 2.37, p = .049$; Figure 5.3d).

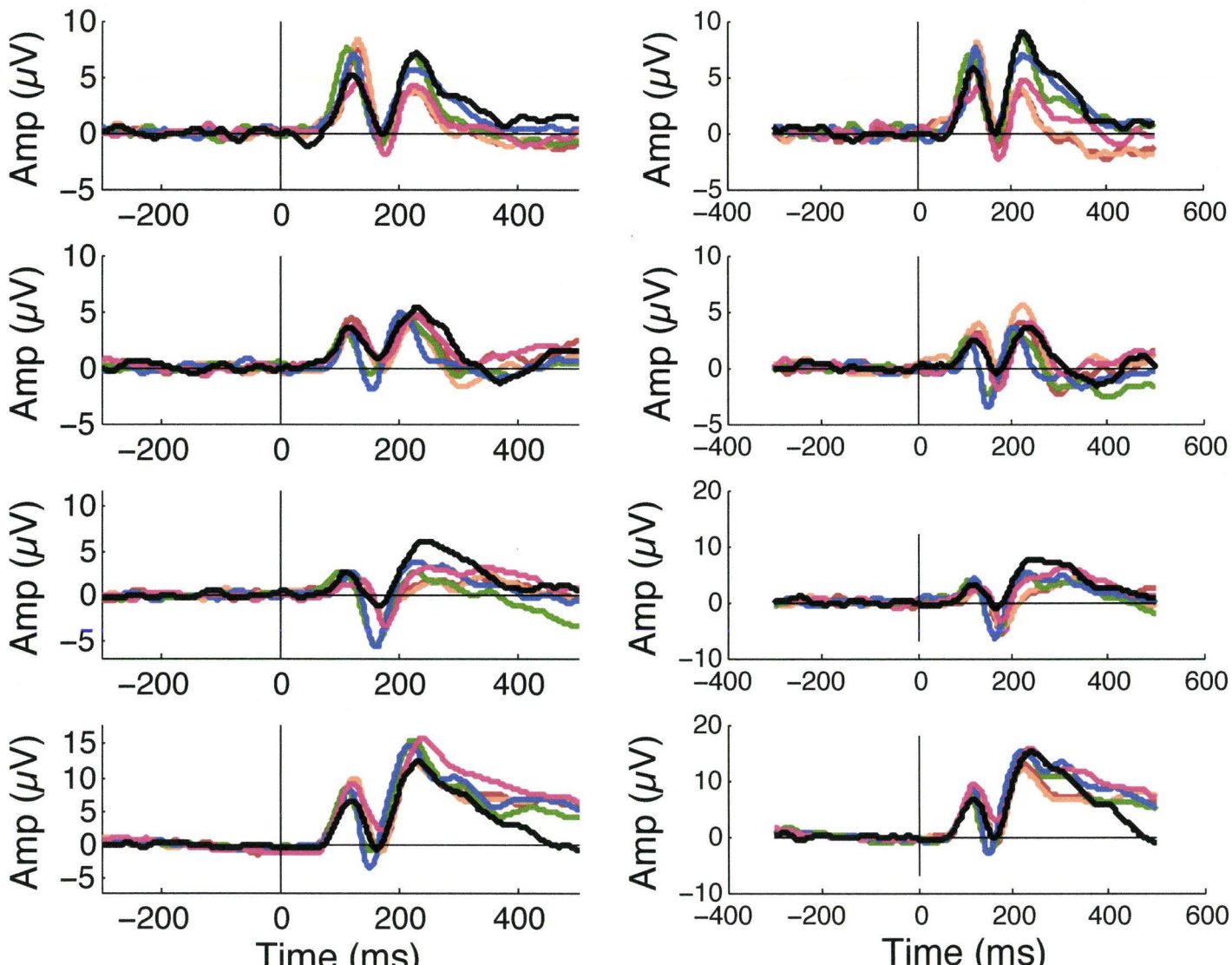

Figure 5.4: ERPs for pre-training (left column) and post-training (right column) for four representative subjects. Legend: Red = trained houses; orange = untrained similar houses; pink = untrained dissimilar houses; Blue/green = faces, and black = objects.

Peak	Hemisphere	Amplitude			Latency		
		Post-Pre	t	p	Post-Pre	t	p
P1	Right	-0.22	-0.51	.62	1.75	1.43	.19
	Left	-0.52	-1.21	.27	-0.25	-0.24	.81
N1	Right	-0.61	-1.50	.18	-3.50	-2.14	.06
	Left	-0.25	-0.52	.62	-1.75	-0.86	.41
P2	Right	-0.06	-0.13	.90	7.00	1.21	.27
	Left	0.31	0.56	.59	3.50	0.37	.73

Table 5.1: Comparison of peak amplitude and latency for P1, N1, and P2 components for trained houses from Pre-training to Post-training.

5.4.2 ERP responses

Activity associated with each of the stimulus conditions and sessions was averaged across trials. Figure 5.4 presents these trial-averaged ERPs for four representative subjects. A peak analysis identified P1, N1, and P2 components for all subjects and conditions. The N250 could not be isolated in the context of this task, so this component was not explicitly analyzed. The P1 was identified in the time-range of 88-132 ms, the N1 was identified in the time-range of 160-210ms, and the P2 was identified in the time-range of 226-284ms. The obseved ERP patterns are typical of other results reported in other studies (e.g., Rousselet et al., 2007, 2008): The N1 in response to faces was similar across both face sets, and, in both cases, larger and earlier than N1 responses to the house sets. The responses to the mixed object set were fairly similar in amplitude and latency to that of faces.

To examine session-differences for each object condition, t-tests were performed to test whether session differences for each condition were significantly different from zero. Table 5.1 lists the results of these session differences for the peak locations of the P1, N1, and P2. Neither amplitude nor latency varied significantly across sessions for any peak. However, in the right hemisphere, the N1 peak was marginally earlier after training ($p = .06$), suggesting that training may have some impact on the speed of neural object processing.

In addition to the peak analysis, session differences were examined systematically across electrodes and time-points. T-tests were performed to determine where session differences differed significantly from zero. Figure 5.5 plots the significant t-values ($p <$

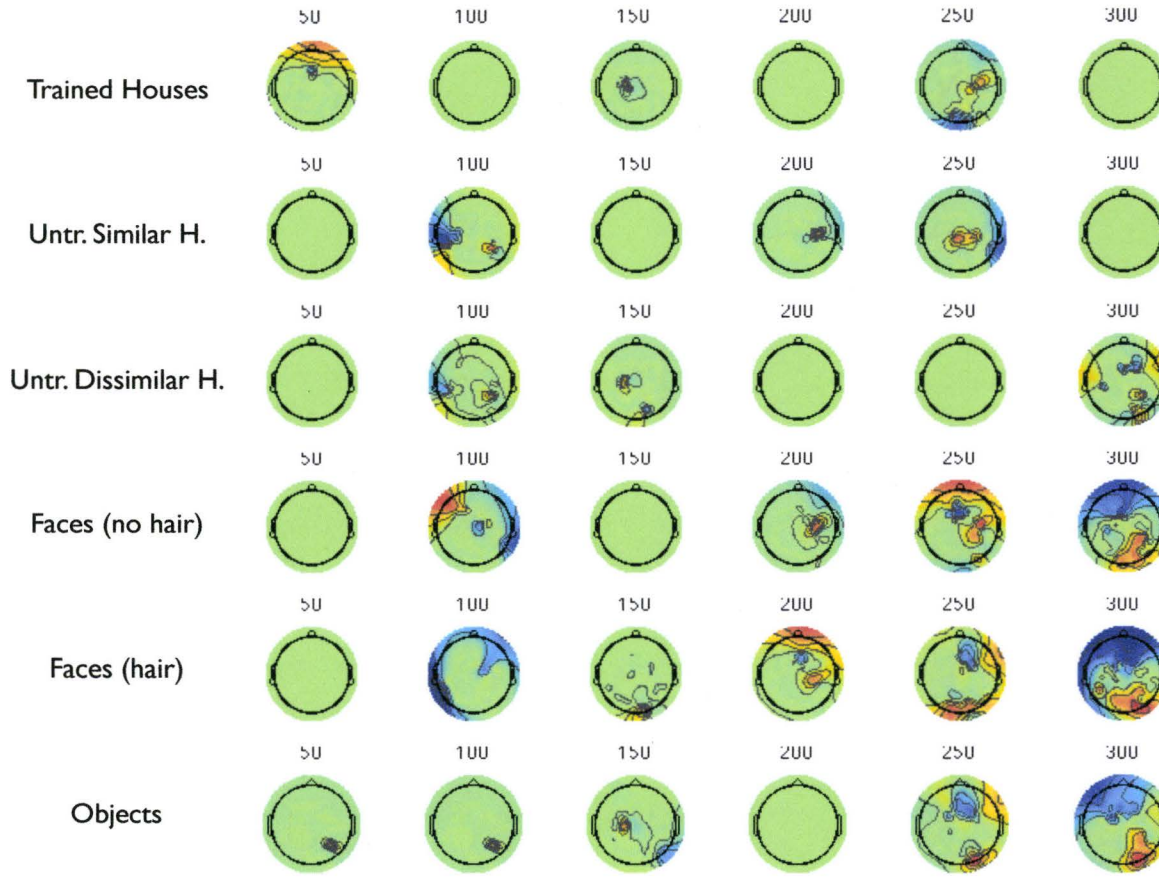

Figure 5.5: Results of t-tests that compared whether difference ERPs (post-training - pre-training) were significantly greater than zero (at each electrode and time-point) for each of the six object conditions. Red = large positive t-scores; Blue = large negative t-scores; green = non-significant t-scores

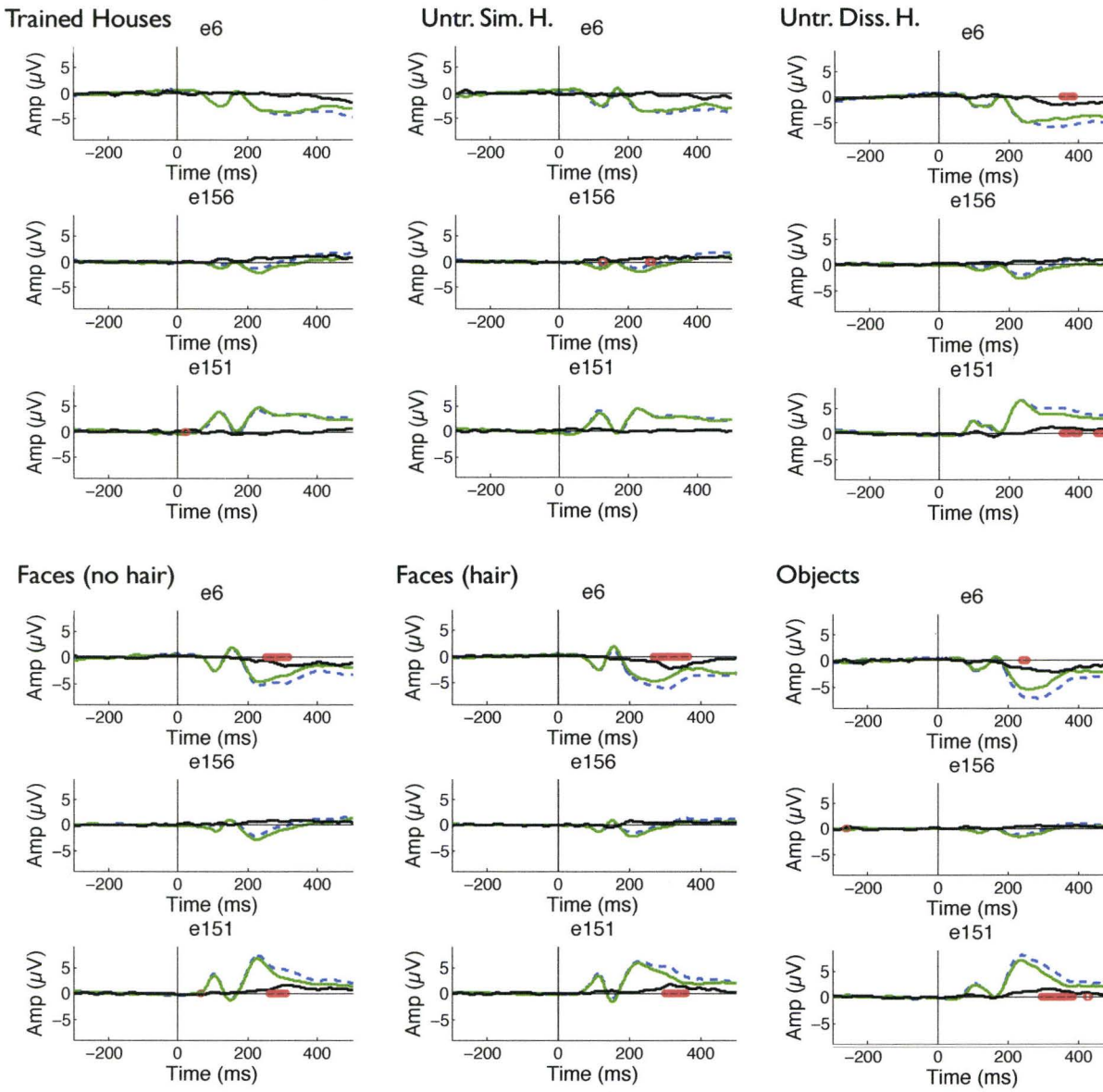

Figure 5.6: Plots of session differences (black solid lines) between Pre-test (green solid lines) and post-test (blue dashed lines). Significant time-points are indicated by red dots along the x-axis. Top row: trained houses, untrained similar houses, and untrained dissimilar houses. Bottom row: faces (no external features), faces (hair and external features), and objects.

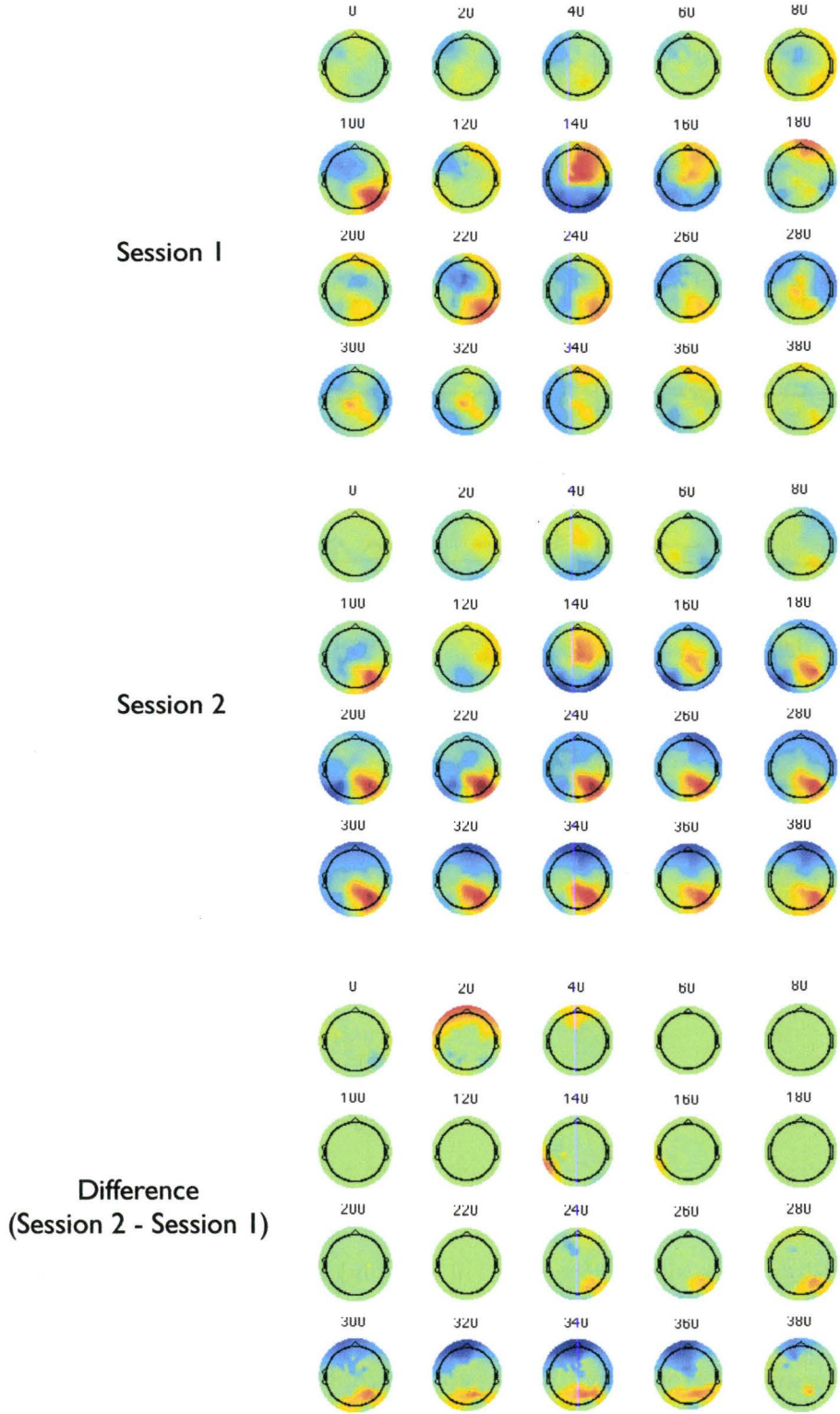

Figure 5.7: Differences between trained houses and faces compared across sessions. The first two panels illustrate the difference between faces and houses in sessions 1 and 2 (Red = Face greater than House; Blue = House greater than Face). The bottom panel is the session difference (Session 2 minus Session 1) for these difference scores, masked to display only significant session differences.

.01) across the scalp. Significant session differences are present, in varying amounts, for all object conditions. Surprisingly, the trained houses do not appear to exhibit greater session differences than other object conditions. In fact, the faces and objects appear to exhibit greater session differences than any of the house sets. T-test results for example electrodes (right hemisphere only) from frontal, mid, and posterior scalp locations are presented in Figure 5.6. Late signficant session differences are present in the time range of 300-400 ms for all conditions except the trained houses and untrained similar houses. It is apparent, both from the overall scalp patterns, and from the example electrodes, that the trained houses, and the untrained similar houses are the only two object sets that do not exhibit an increase in amplitude in the range of 300-400 ms after stimulus onset.

Finally, we examined whether the differences between the trained houses and faces were affected as a result of training. If training of houses increases the similarity of their ERP signatures to that of faces, we should expect smaller differences between faces and houses across sessions. Figure 5.7 illustrates changes in the difference ERPs between houses and faces across sessions. Difference waves were calculated between faces (no hair) and the trained houses. These difference waves were calculated both pre- and post-training (displayed in top and middle panels), and t-tests were conducted to determine whether these difference waves differed across sessions. The bottom panel of Figure 5.7 plots only the significant differences across sessions. The topographic distributions indicate that differences between faces and houses increased after training. This was especially prominent in late time-periods (between 300-360 ms). This increase in differences between the faces and houses is likely a result of the cross-session increase in amplitude observed for faces, but not for houses. In earlier time-ranges, no significant differences were observed across sessions. Thus, we found no evidence that differences between faces and houses lessened after training in the time-range of the N1.

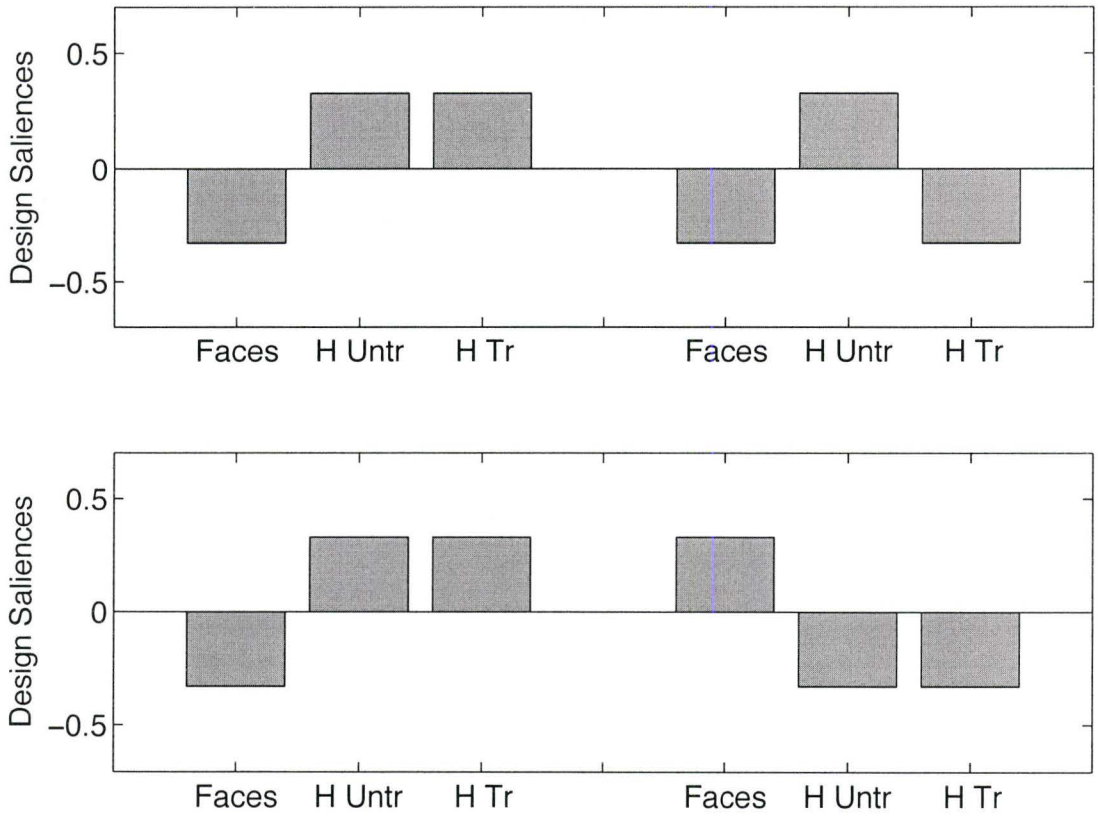

Figure 5.8: PLS predictions. Top: This pattern of design saliences would be expected if training increased the similarity of ERP activity between the trained houses and faces, without transfer to the untrained house set. Bottom: This pattern of design saliences corresponds to a training × stimulus interaction, and indicates that training had opposite effects on houses and faces.

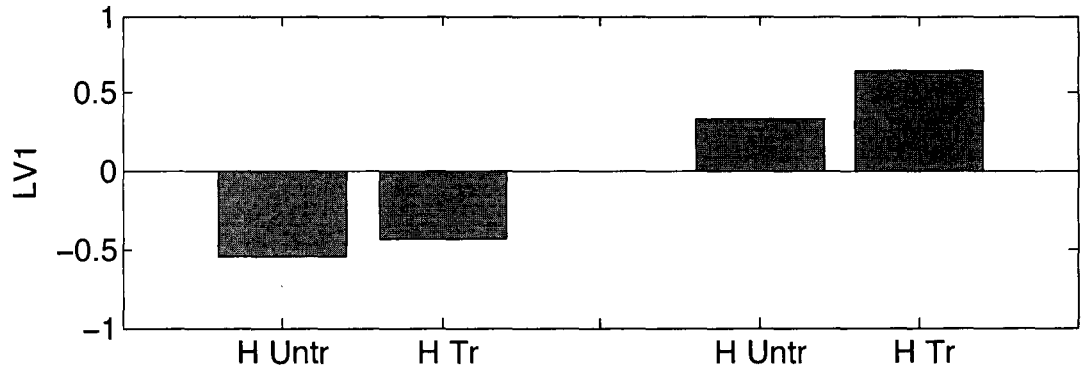

Figure 5.9: PLS analysis: Trained vs. Untrained houses. Design saliences for latent variable LV1 ($p < .016$). The pattern of weights distinguishes between responses in pre- and post-training sessions

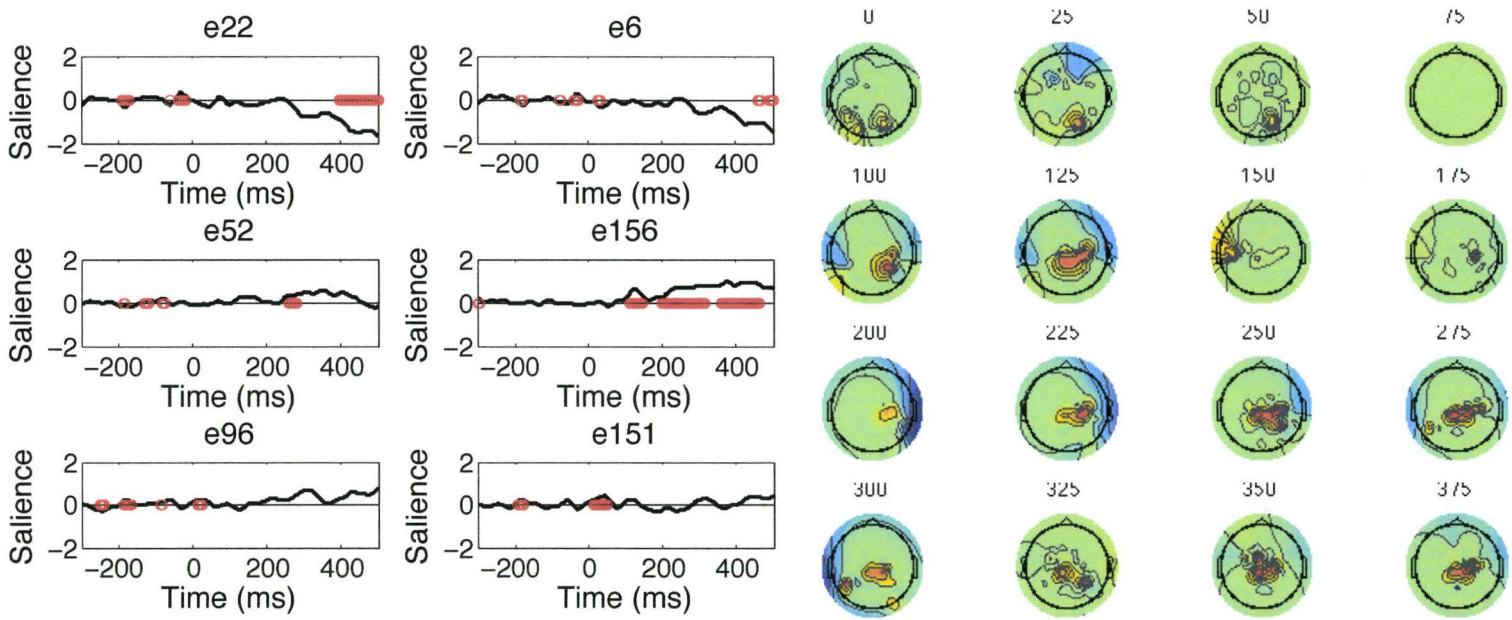

Figure 5.10: PLS analysis: Trained vs Untrained houses. Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

5.4.3 PLS analyses

In addition to the ERP analyses, a series of partial least squares (PLS) analyses were conducted to examine the effects of training on ERP activity across sessions and object conditions. PLS is a multivariate statistical technique, giving it an advantage over more traditional univariate ERP analyses in identifying patterns of activation across electrodes and conditions. We employed PLS to determine whether training effects might be present for the trained houses that were missed by more traditional univariate analyses.

PLS determines the sets of weights that maximize the covariance between ERP activity and task conditions (McIntosh et al., 1996; McIntosh and Lobaugh, 2004). The output of a PLS analysis is a series of latent variables that correspond to sets of weights, or saliences, across both the experimental conditions (in this case, the object conditions and sessions) and the ERP responses across time. The significance of each latent variable was assessed with permutation tests (500 samples), and the reliability of non-zero saliences were determined through bootstrap estimates (500 samples) of the salience standard errors.

Because PLS is a data-driven technique (unlike analyses of variances, where contrasts of interest are explicitly chosen), the resultant latent variables must be interpreted post-hoc, and do not always correspond to simple experimental factors. Figure 5.8 illustrates two hypothetical PLS results that would be indicative of training effects. In the first example (top panel), the contrast formed by the design saliences (condition weights) illustrates a case where both trained- and untrained-houses are initially differentiated from faces in Session 1 (left side), but only the untrained-houses are differentiated from faces in Session 2 (right side). This pattern of results might be expected if house-identification training increased the amplitude of house ERP activity so that it differed less from that of faces, after training. In this example, the effect of training on the ERP was assumed not to transfer to the untrained house set. In the second example (bottom panel), the design saliences illustrate a case where the effect of session is equal and opposite for the houses and faces. This pattern of results might be expected in house-identification training increased the amplitude of house ERP activity, while face ERP activity declined across sessions. This example also illustrates a case of transfer of learning from the trained houses to the untrained houses.

The subsequent PLS analyses will contrast the activity of trained and similar-untrained houses first, with each other, and then with each of: dissimilar-untrained houses, objects,

and faces. If training effects were present, we would expect results similar to those illustrated in Figure 5.8.

5.4.3.1 Untrained vs Trained Houses

The first PLS analysis was designed to determine whether any changes occurred across sessions for the trained houses, and whether such effects generalized to the similar untrained houses. In previous studies, we found that training with these same houses increases identification accuracy on the trained house set and that the effect partially generalizes to untrained houses (Chapter 2). Learning-related changes in ERP activity for the trained houses might also be expected to transfer to the untrained house set.

This first PLS analysis revealed only one significant latent variable (LV1; $p < .016$). The design saliences for this latent variable (Figure 5.9) contrast the pre-training and post-training sessions, indicating that ERP activity differed across sessions. The timing of the session effect is indicated in the left-hand panel of Figure 5.10, which shows the ERP saliences (weights) for several frontal, central, and posterior electrodes from the left and right hemisphere. Time-points where the non-zero saliences are reliable are indicated by red dots. The right hand panel of this figure presents a topography of the ERP saliences that was created by first summing reliable saliences within 25 ms windows (from 0-400 ms) for each electrode, and then plotting the sums across the scalp: In each time window, red values indicate electrodes where responses varied across conditions in a way that matched the design saliences (i.e., post-training activity was greater than pre-training activity), and blue values indicate electrodes where responses varied in the opposite way. The topography of LV1 suggests that the session difference was expressed most strongly in central electrodes starting around 100 ms after stimulus onset and persisting through 400 ms after stimulus onset.

Because the session effect represented by LV1 was present for both trained and untrained houses, it is not clear whether this effect is a result of training that generalized from trained to untrained houses (Chapter 2). The effect could, for example, represent a generalized increase in activity for all classes of stimuli. Additional PLS analyses presented in the following sections address this question.

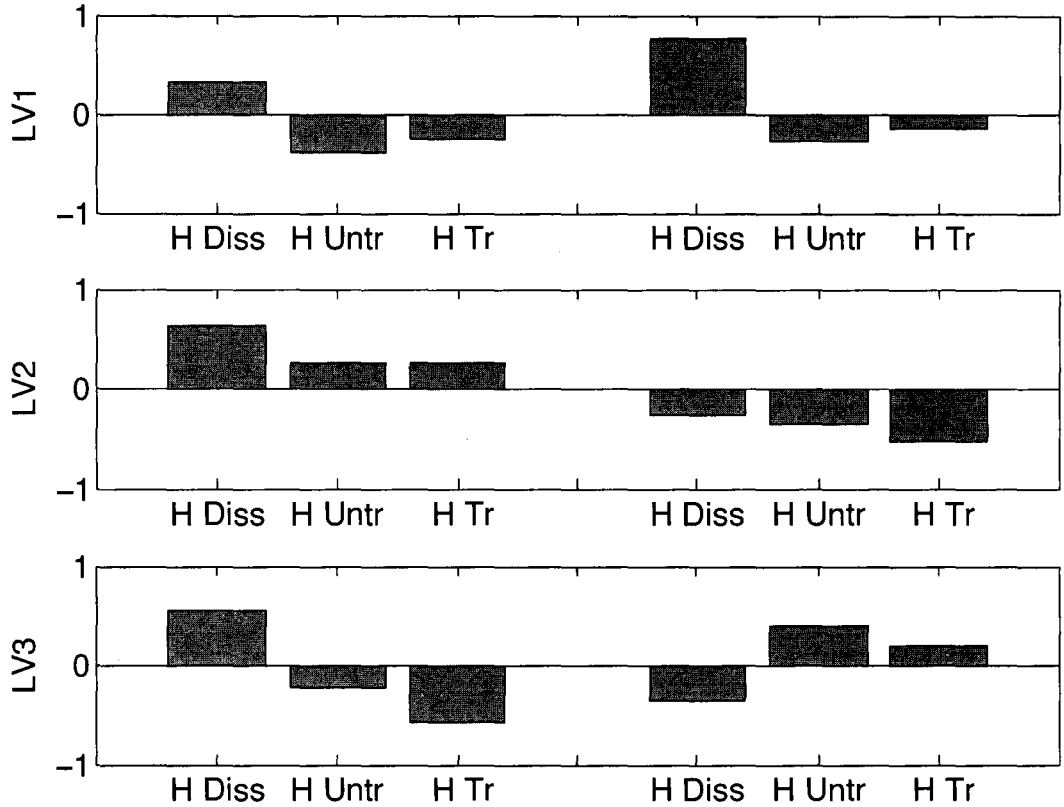

Figure 5.11: PLS analysis: All houses. Design saliences for latent variables LV1 ($p < .001$), LV2 ($p < .08$), and LV3 ($p < .82$). LV1 contrasted the ERP responses to the trained- and similar-untrained houses with the ERP responses to the dissimilar-untrained houses. LV2 distinguished between responses in pre- and post-training sessions, and LV3 represented an interaction between session and house-type.

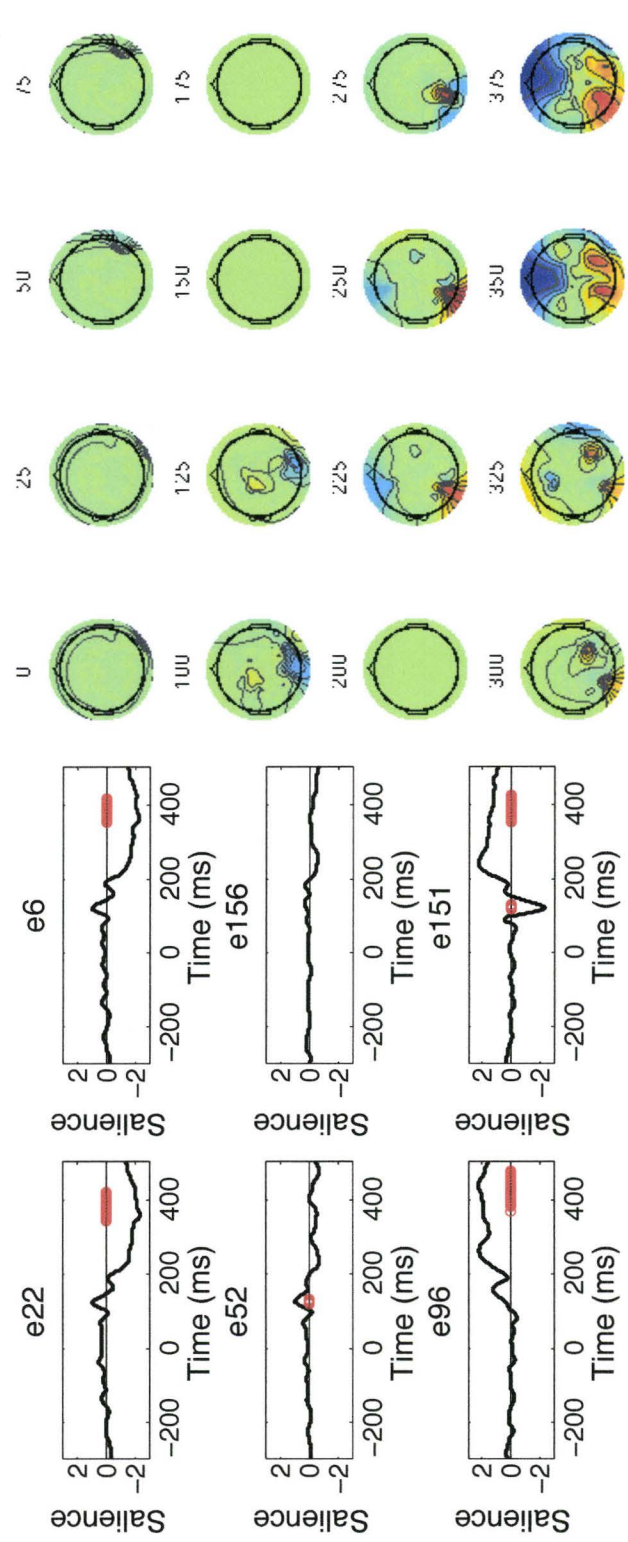

Figure 5.12: PLS analysis: All houses (LV1). Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

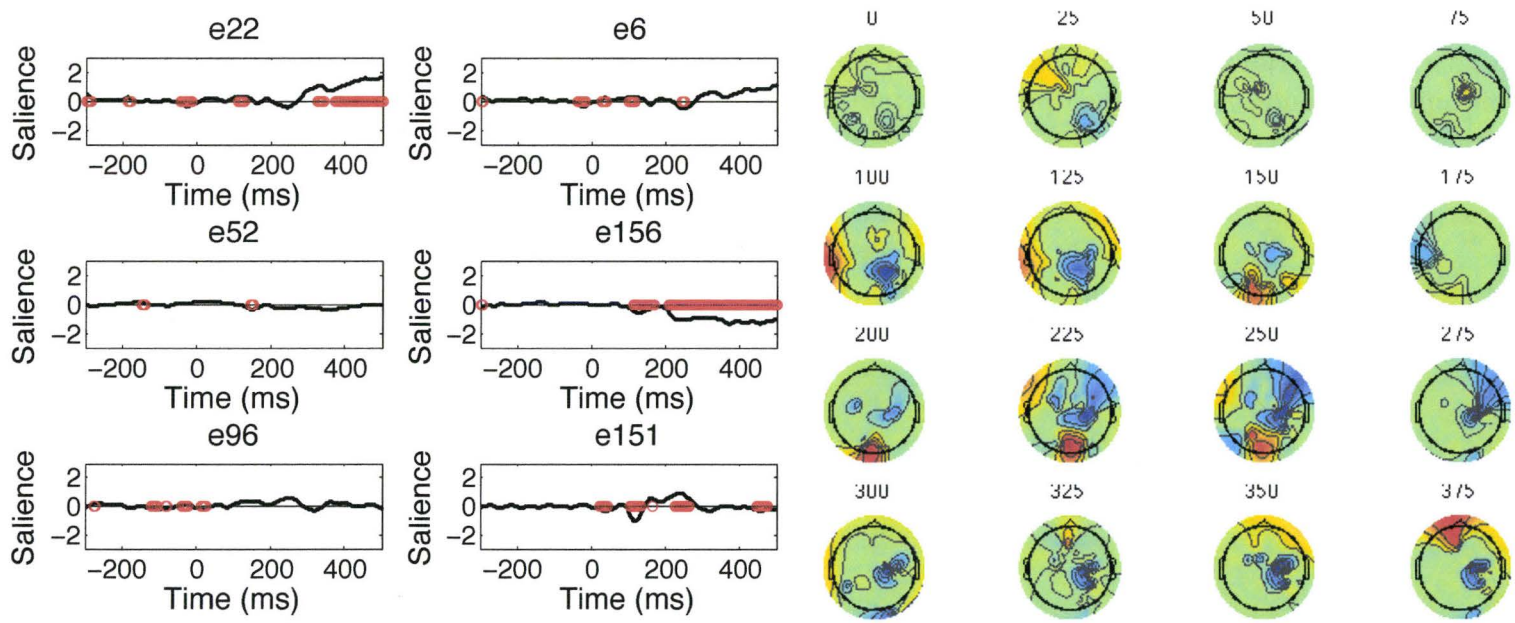

Figure 5.13: PLS analysis: All houses (LV2). Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

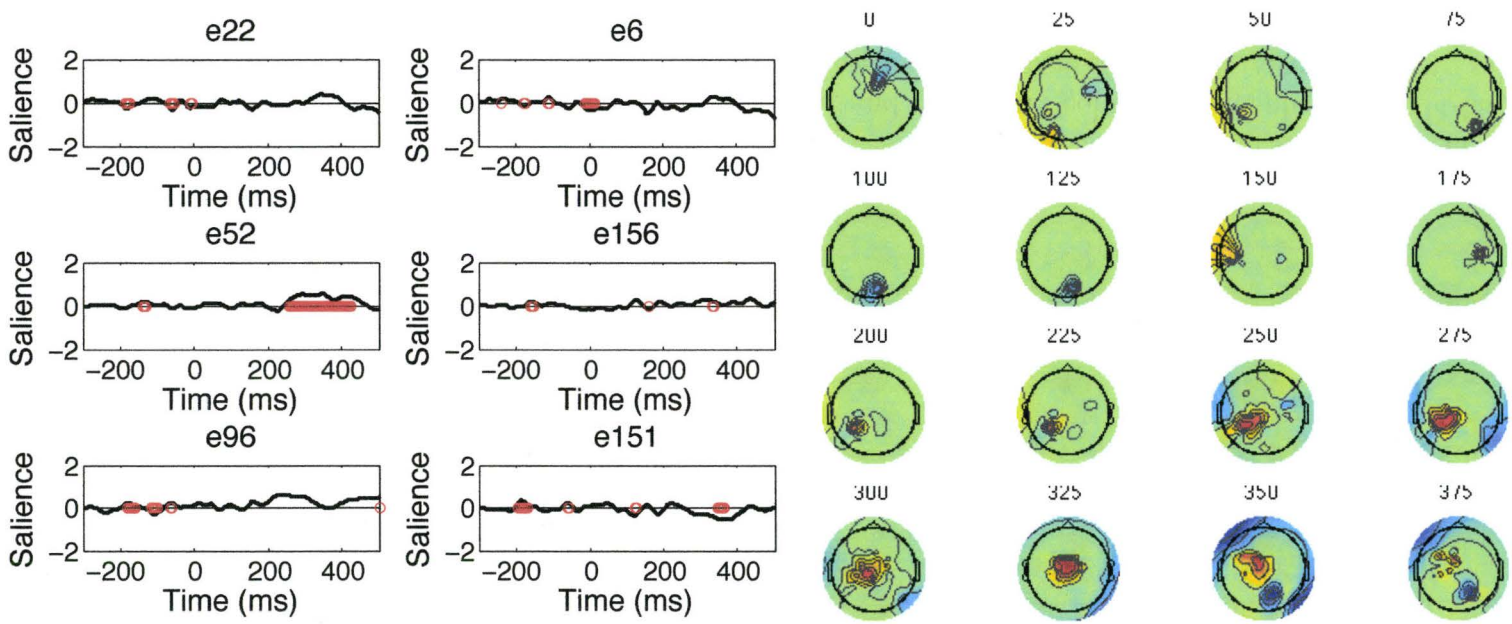

Figure 5.14: PLS analysis: All houses (LV3). Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

5.4.3.2 All Houses

We next compared activity across all three house sets (i.e., trained, untrained-similar, and untrained-dissimilar houses). This analysis was designed to help clarify the interpretation of the session effect observed in the first PLS analysis. If the cross-session increase in activity observed for the trained and untrained houses in the first PLS analysis was due to learning that generalized from the trained to the untrained houses, we might expect this effect to be weaker for the dissimilar house set. On the other hand, if activity was universally higher during the second session (and unrelated to the intervening training), the session effect might be present for all three house sets.

This analysis produced one significant latent variable (Figure 5.11), LV1 ($p < .001$), that distinguished between the dissimilar houses and the other two house sets. This difference among stimulus categories was expressed in the late components of the ERPs (i.e., $\approx$ 350 ms after stimulus onset): responses to dissimilar houses had higher amplitude on bilateral posterior electrodes and lower amplitude on frontal electrodes (Figure 5.12). A similar, but left lateralized effect is observable around 250 ms. The second latent variable, LV2, was marginally significant ($p < .08$), and replicates the session effect observed in the first PLS analysis. Just as observed in the first PLS analysis, the ERP activity of central electrodes increased from pre- to post-training (Figure 5.13), with no interaction across conditions. The third latent variable represents exactly the type of training-related interaction that would clearly indicate a training effect, with ERP activity increasing for dissimilar houses, but decreasing for trained and similar houses in the range of 200-400 ms for central electrodes (Figure 5.14). However, LV3 was not statistically significant ($p < .82$).

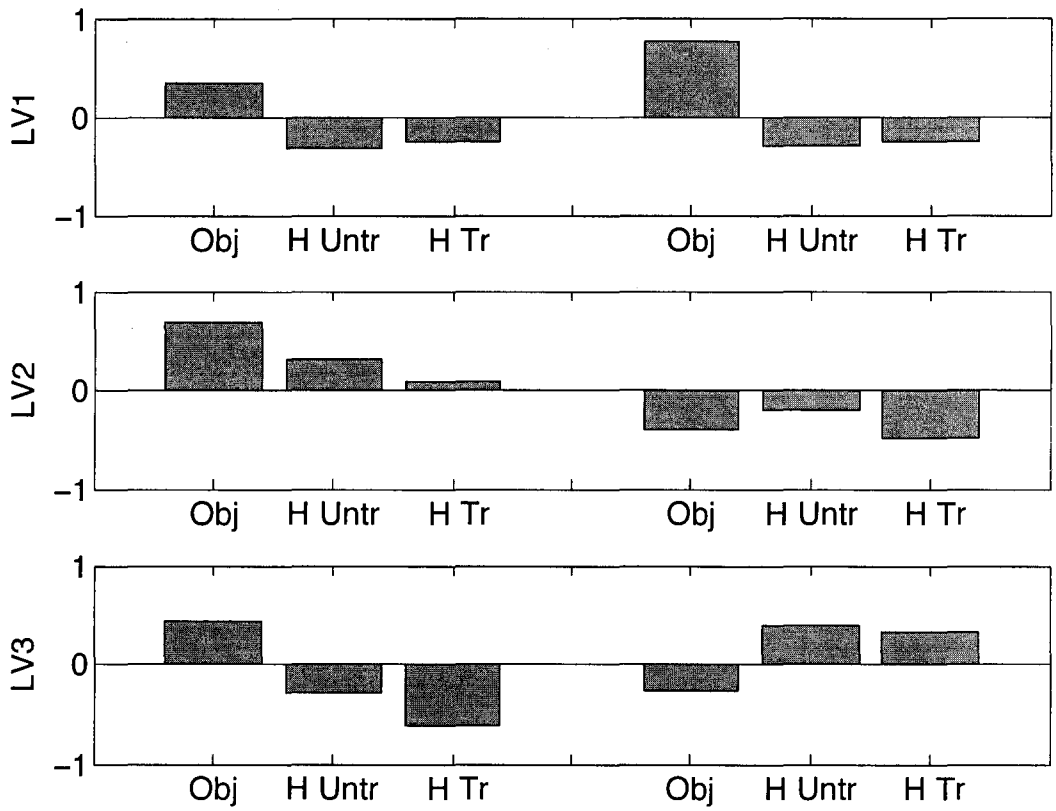

Figure 5.15: PLS analysis: Houses vs Objects. Design saliences for latent variables LV1 ($p < .001$), LV2 ($p < .12$), and LV3 ($p < .76$). LV1 distinguished between ERP responses to objects and houses, LV2 distinguished between responses in pre- and post-training sessions, and LV3 represented an interaction between session and object-type

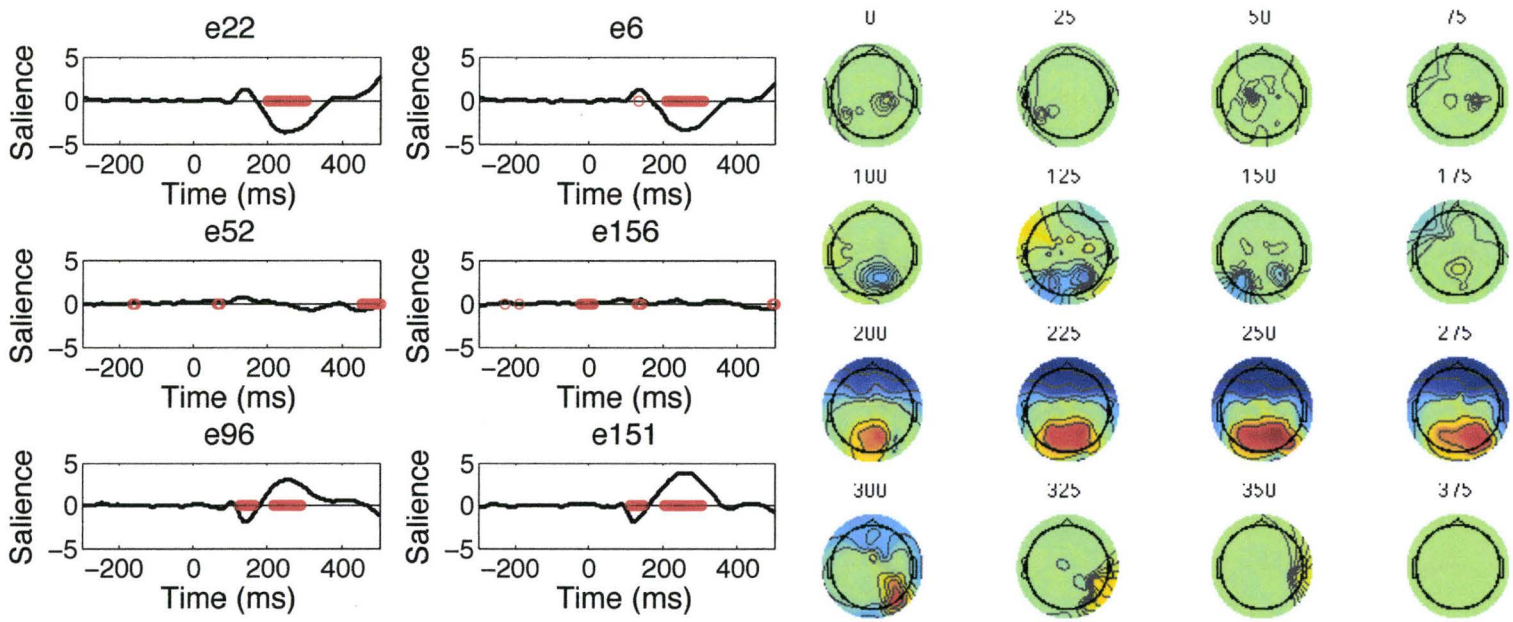

Figure 5.16: PLS analysis: Houses vs Objects (LV1). Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

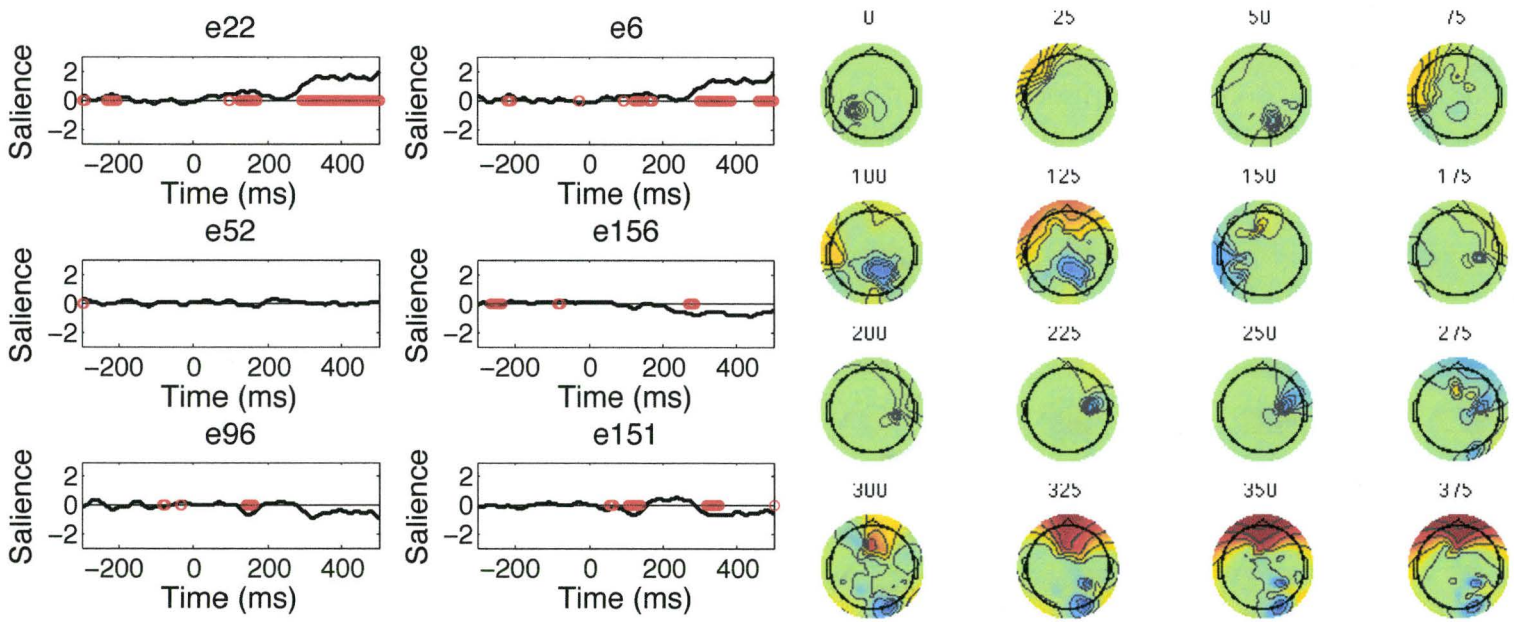

Figure 5.17: PLS analysis: Houses vs Objects (LV2). Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

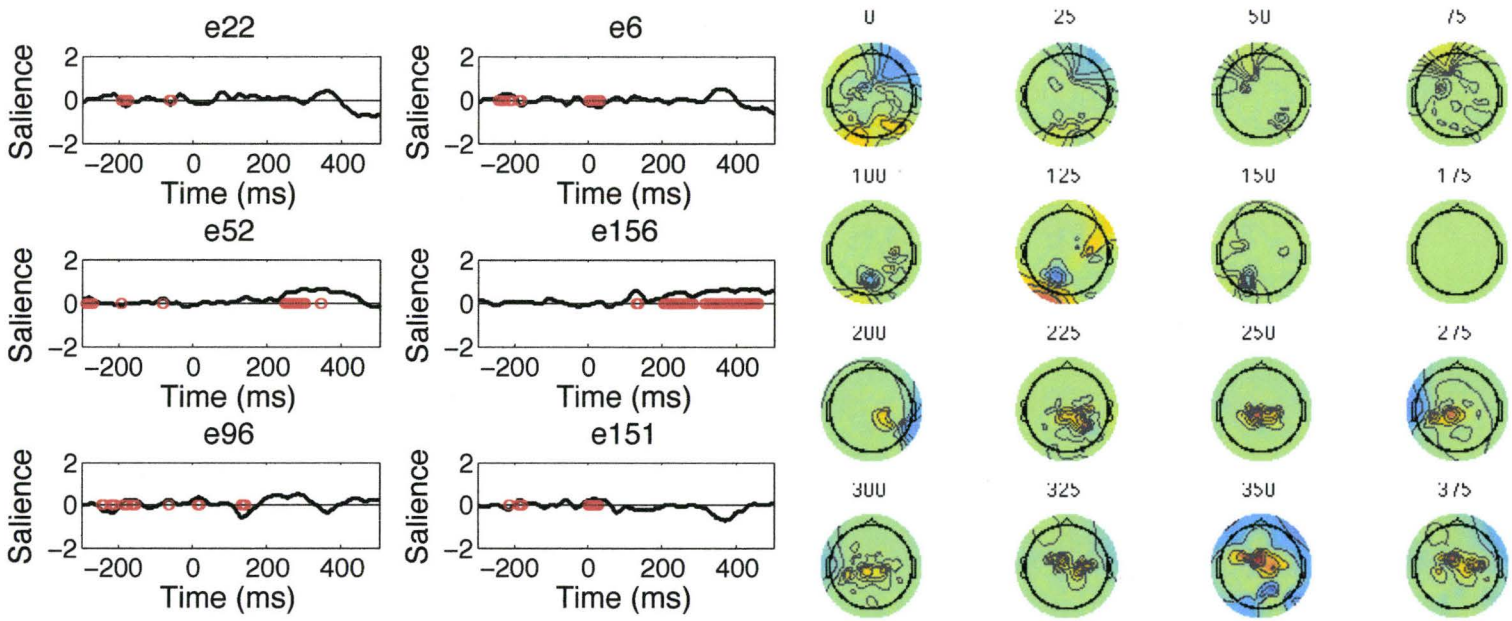

Figure 5.18: PLS analysis: Houses vs Objects (LV3). Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

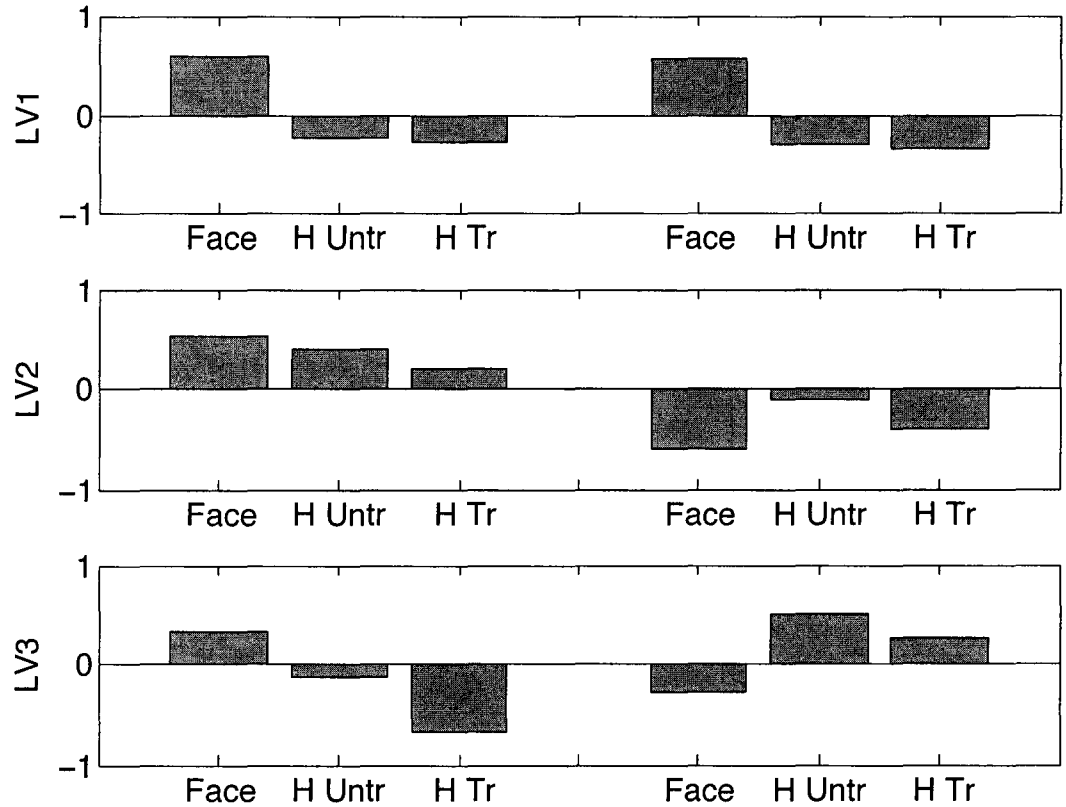

Figure 5.19: PLS analysis: Houses vs Faces. Design saliences for latent variables LV1 ($p < .001$), LV2 ($p < .02$), and LV3 ($p < .79$). LV1 distinguished between ERP responses to faces and houses, LV2 distinguished between responses in pre- and post-training sessions, and LV3 represented an interaction between session and object-type.

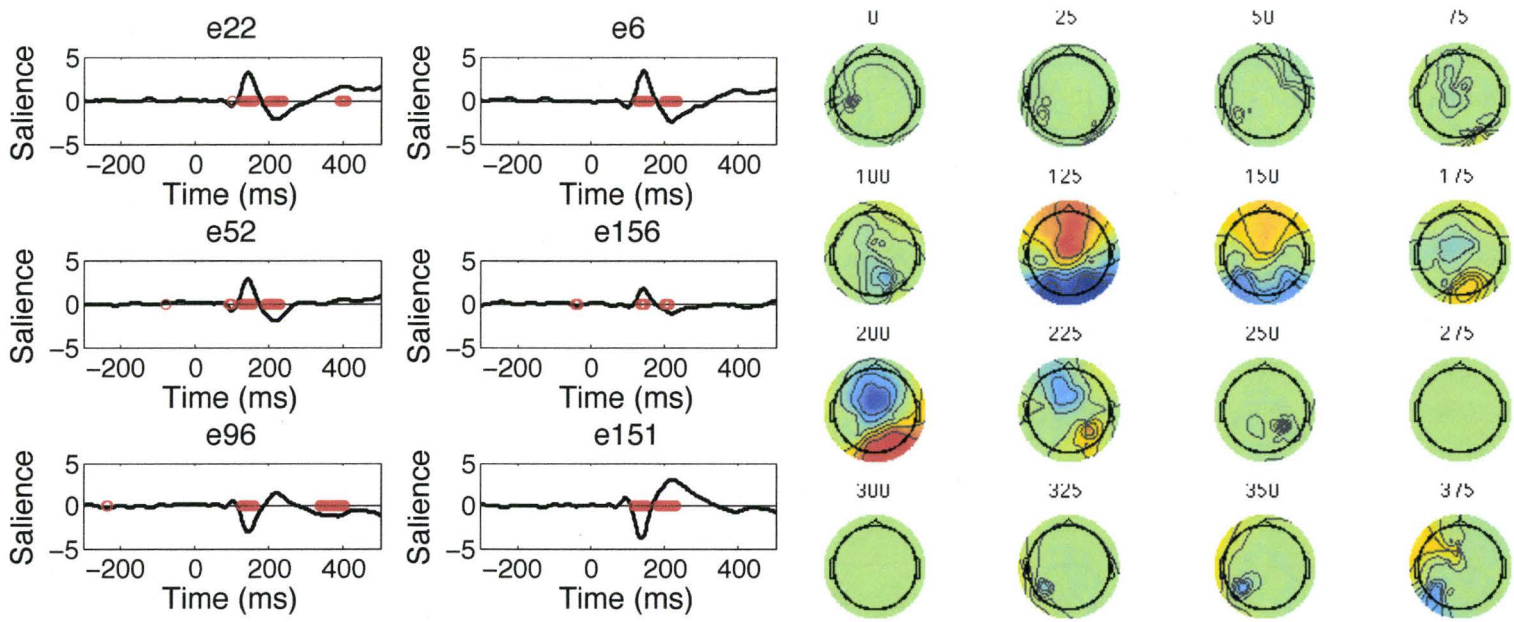

Figure 5.20: PLS analysis: Houses vs Faces (LV1). Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

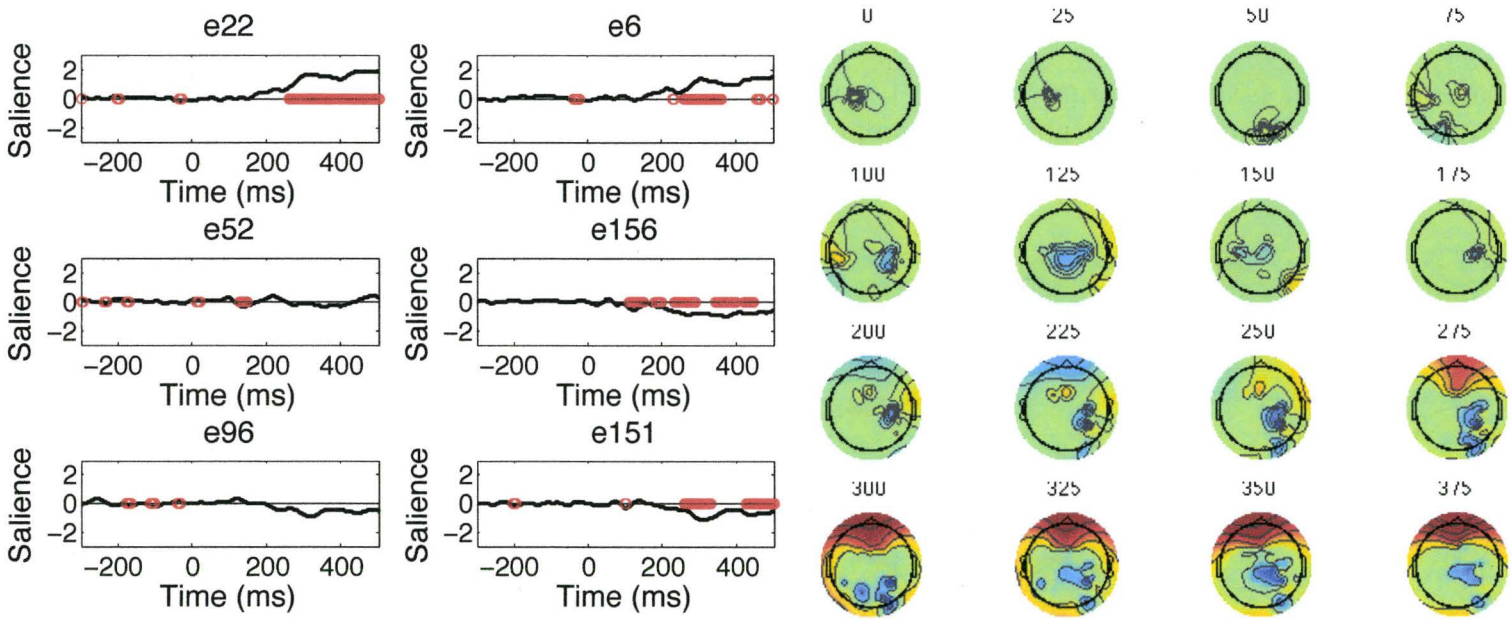

Figure 5.21: PLS analysis: Houses vs Faces (LV2). Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

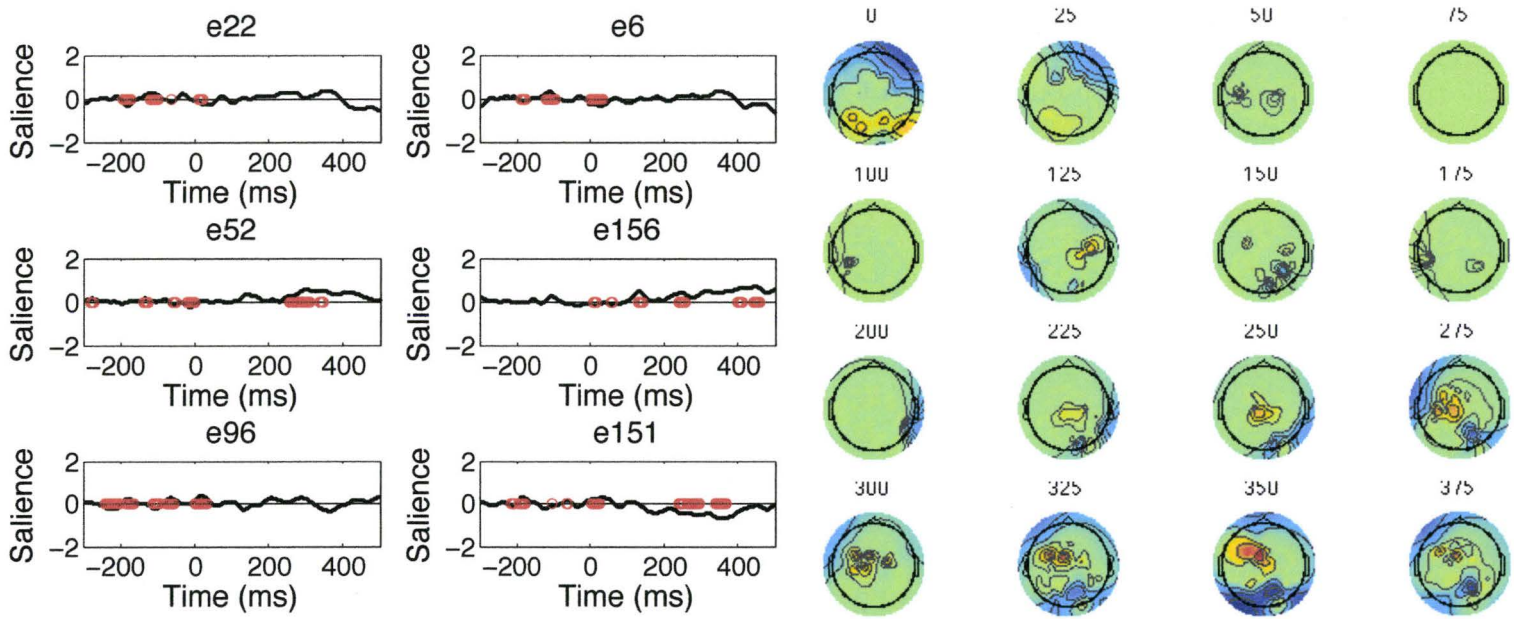

Figure 5.22: PLS analysis: Houses vs Faces (LV3). Left hand panel: Electrode saliences for sample frontal (top row), central (middle row) and posterior (bottom row) electrodes in the left (1st column) and right (2nd column) hemispheres. Right-hand panel: Topographic plots of electrode saliences: Values range from red (strong positive saliences) to blue (strong negative saliences)

5.4.3.3 Houses vs. Objects

The first two PLS analyses indicated that the ERP activity in central electrodes increased during the post-training session for both the trained and untrained houses, and marginally, for all three house sets. We ran two additional PLS analyses to compare the activity of the trained and similar-untrained houses to increasingly different stimulus types: a class of mixed objects, and a class of faces. The trained and similar-untrained house sets were first compared with the mixed class of objects. This analysis resulted in only one significant latent variable (LV1; $p < .001$) that indicated that ERPs for mixed objects differed from ERPs for houses (Figure 5.15). In particular, objects induced stronger posterior (and weaker frontal) activity than houses, between 200-300 ms after stimulus onset. As in the previous analysis, LV2 corresponded to a marginally significant ($p < 0.12$) effect of session that did not differ across conditions. However, the topography of this session effect differed from that obtained in the first two PLS analyses (Figure 5.17): The session effect was associated with lower activity at post-training in the frontal electrodes approximately 300 ms after stimulus onset. Interestingly, the third latent variable, though non-significant, represented the same form of interaction observed when comparing the three house sets, with a similar topography (Figure 5.18). Although the session-difference observed in the trained- and similar-untrained houses is weakened by the addition of the object condition, this analysis still provides no firm evidence of a training effect because no clear interaction term between session and object-class was observed.

5.4.3.4 Houses vs. Faces

The final PLS analysis, comparing the trained and untrained houses to faces also produced two significant latent variables (Figure 5.19). LV1 ($p < .001$) contrasted the ERP activity for faces with houses, particularly in 125-150 ms after stimulus onset, when activity for faces was higher at frontal electrodes and lower at posterior electrodes, and again at 200 ms after stimulus onset, when activity for faces was lower at frontal electrodes and frontal electrodes and higher at posterior electrodes (Figure 5.20). LV2 ($p < .02$) was a non-specific session effect, similar to those found in the previous PLS analyses. This session effect corresponded to a decrease in frontal activity in the post-training session for both houses and faces, starting approximately 275 ms after stimulus onset (Figure 5.21). Finally, the third, non-significant LV (LV3, $p < .79$) represented an increase in activity at central electrodes for the trained and untrained houses after 250 ms, with a

corresponding decrease in activity for faces (Figure 5.22)

This analysis, more clearly than the last two, indicates that the session effect observed with the trained- and similar-untrained houses was not unique to these two conditions, as similar effects were observed with faces in this analysis, and marginally observed with both dissimilar-untrained houses and objects in the previous PLS analyses. This generalized session effect, combined with the failure to find significant interaction terms in all of the analyses, suggests that although session effects were present, they were unrelated to training in the house-identification task.

5.5 Discussion

The aim of this study was to compare the relative ERP activity induced by houses and other object classes (including faces) before and after training on a house discrimination task. Five days of training on a house discrimination task is sufficient to induce substantial improvement in accuracy and reaction time across subjects, and to induce a house inversion effect (improved performance on upright versus inverted stimulus presentations), a phenomenon frequently attributed to expert processing (Chapters 2, 3, and 4).

Here, we proposed that house discrimination training might alter the ERP signature to the trained houses, and that such effects might generalize to similar houses, though untrained. ERP analyses revealed a marginal decrease in N1 latency with training, suggesting that house-discrimination training may increase the speed of processing associated with these houses. Session differences in the amplitude of responses circa 300-400 ms were present for all conditions except the trained and untrained houses. The absence of an amplitude increase across sessions in this task could be a result of the overall decrease in amplitude for these houses observed during the training sessions (Chapter 4). This would be consistent with previous reports of decreased amplitude of late components with increased stimulus familiarity (Eimer, 2000b; Bentin and Deouell, 2000).

PLS analyses revealed session differences for the trained and untrained houses missed by the univariate ERP analyses, however subsequent PLS analyses indicated that these session differences were not isolated to these conditions. Similar session effects were observed for the dissimilar houses, objects, and faces. Because the observed session differences were universal across conditions, they are unlikely to reflect an effect of house training. These session differences were consistent across subjects, therefore, it is unlikely

that they were due to random day effects in ERP strength, however, they could reflect more general changes associated with task and stimulus familiarity in the one-back task itself. In a previous study that employed this same one-back task (Chapter 3) we observed that a control group that performed the one-back task across two sessions without intervening training sessions, increased in accuracy across sessions. This suggests that session effects, non-specific to the trained houses, might reflect learning effects within the one-back task itself. We did not observe any conclusive evidence of a selective change for the trained houses, that might be indicative of an effect of the intervening training sessions. However, it is interesting to note that we consistently observed, across three separate analyses, a non-significant latent variable that was indicative of a selective session change for the trained (and similar) houses. Although these LVs did not approach significance, it is tempting to wonder whether these effects might have emerged more reliably with a greater sample size.

Although we observed session effects for both houses and faces, we observed no evidence that the difference between ERP signatures for houses and faces was decreased from pre-training to post-training. Rather, the only change across sessions was an increase in the difference between faces and houses, as a result of a selective amplitude increase for faces in late time-windows around 300-400 ms. We found no evidence that earlier time-windows, such as those including the N1 and N250, were modified by training in this task. This is consistent with earlier reports that the face N1 is not noticeably affected by the familiarity of the presented faces (Eimer, 2000a,b; Bentin and Deouell, 2000), but in contrast to reports of greater N1 strength for cars than faces in car experts (Gauthier et al., 2003) and of greater N1 strength for birds than dogs in bird experts and the reverse for dog experts (Tanaka and Curran, 2001).

As we were unable to clearly identify the N250 for our subjects, we were unable to directly test the involvement of this component in our task. Previous reports have indicated that this component is involved in learning to differentiate between individual exemplars (Scott et al., 2006, 2008), thus an increase in N250 amplitude would be expected in the context of training individual house-discrimination. The PLS analyses did identify some session differences that appeared as early 250 ms after stimulus onset, overlapping with the typical timing of the N250, however, these session differences were not specific to the trained conditions, so it is unlikely that these differences reflect training effects.

In sum, we found no evidence of selective session differences for trained stimuli, in contrast to previous reports of increased N1 amplitude with expertise (Gauthier et al.,

2003; Tanaka and Curran, 2001). Real world expertise presumably involves training that is both longer in duration and more varied in context relative to the house training our subjects received, where training was restricted to a single visual perceptual task. Future research will need to determine whether either of these factors might account for the difference between the sensitivity of the N1 to expertise in real-world contexts, relative to the stability of the N1 in the current context.

References

Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience 8 (6), 551–565.

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

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

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134 (1), 9–21.

Eimer, M., 2000a. Attentional modulations of event-related brain potentials sensitive to faces. Cognitive Neuropsychology 17 (1-3), 103–116.

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

Gauthier, I., Curran, T., Curby, K. M., Collins, D., 2003. Perceptual interference supports a non-modular account of face processing. Nat Neurosci 6 (4), 428–432.

Gold, J., Bennett, P. J., Sekuler, A. B., 1999. Identification of band-pass filtered letters and faces by human and ideal observers. Vision Res 39 (21), 3537–3560.

Itier, R. J., Taylor, M. J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cereb Cortex 14 (2), 132–142.

Jeffreys, D. A., Tukmachi, E. S., 1992. The vertex-positive scalp potential evoked by faces and by objects. Exp Brain Res 91 (2), 340–350.

Mace, M. J.-M., Thorpe, S. J., Fabre-Thorpe, M., 2005. Rapid categorization of achromatic natural scenes: how robust at very low contrasts? Eur J Neurosci 21 (7), 2007–2018.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., Grady, C. L., 1996. Spatial pattern analysis of functional brain images using partial least squares. Neuroimage 3 (3 Pt 1), 143–157.

McIntosh, A. R., Lobaugh, N. J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 23 Suppl 1, S250–63.

O'Craven, K. M., Downing, P. E., Kanwisher, N., 1999. fMRI evidence for objects as the units of attentional selection. Nature 401, 584–587.

Rossion, B., Collins, D., Goffaux, V., Curran, T., 2007. Long-term expertise with artificial objects increases visual competition with early face categorization processes. J Cogn Neurosci 19 (3), 543–555.

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. Proc Natl Acad Sci U S A 101 (40), 14521–14526.

Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B., 2007. Single-trial EEG dynamics of object and face visual processing. Neuroimage 36 (3), 843–862.

Rousselet, G. A., Husk, J. S., Bennett, P. J., Sekuler, A. B., 2008. Time course and robustness of ERP object and face differences. J Vis 8 (12), 3.1–18.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., Curran, T., 2006. A reevaluation of the electrophysiological correlates of expert object processing. J Cogn Neurosci 18 (9), 1453–1465.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., Curran, T., 2008. The role of category learning in the acquisition and retention of perceptual expertise: a behavioral and neurophysiological study. Brain Res 1210, 204–215.

Tanaka, J. W., Curran, T., 2001. A neural basis for expert object recognition. Psychol Sci 12 (1), 43–47.

Taylor, M. J., Edmonds, G. E., McCarthy, G., Allison, T., 2001. Eyes first! Eye processing develops before face processing in children. Neuroreport 12 (8), 1671–1676.

Chapter 6

General Discussion

6.1 Overview

As outlined in Chapter 1, the purpose of this thesis is not only to learn more about the effect of experience on object processing, but also, by extension, to consider the extent to which the processing of faces differs from that of other object classes. To achieve this aim, three studies were conducted:

The first study, summarized in Chapter 2, examined the inversion effect, a behavioural phenomenon that has been studied extensively in the context of face perception. Our study demonstrated that inversion effects can be generated in stimuli that initially do not exhibit this effect, with as little as five days of in-lab training. Importantly, we demonstrated that these inversion effects, like those found with faces, generalize to members of the object class that were not seen during training. Finally, we demonstrated that in-lab training produces an inversion effect that is statistically equivalent (in magnitude) to that observed with faces. We concluded that inversion effects can be obtained as a result of greater experience with a particular stimulus orientation, the benefits of which transfer only partially to other orientations. In addition, the size of the resulting inversion effect is affected both by the characteristics of the object set and by the amount of exposure to that set. Finally, we argued that the results implied that the absolute size of the face inversion effect is insufficient evidence to conclude that faces are processed by a separate mechanism from other objects.

The second study, described in Chapter 3, explored the specificity of the fusiform gyrus to face processing, by attempting to replicate earlier reports of increased fusiform

activity in response to in-lab training with non-face objects (Gauthier and Tarr, 1997). We extended this earlier paradigm by making use of a stimulus class (houses) that is associated with a pre-existing cortical area (the parahippocampal gyrus), in order to consider how experience affects the FFA and other surrounding areas, including the PPA and lower visual areas. In addition, by making use of a multivariate statistical technique, Partial Least Squares (PLS), we were able to look for changes in the global patterns of activity across the whole cortex. Despite behavioural evidence for substantial learning across the five days of training, we found that the degree of FFA activation by the trained houses did not change reliably across training. Moreover, activity in so-called house areas, including the PPA, TOS, and RSC, as well as activity in early visual areas, also were unaffected by training on the house task. The whole-brain PLS analyses confirmed that global patterns of cortical activity were unaffected by the intervening training sessions. Nonetheless, the small amount of training obtained on the scanning task derived during the pre-test session was associated with changes in cortical activation in two of the house areas (PPA and RSC) for one of the two house sets. Thus, like Op de Beeck et al. (2006), we were unable to replicate the original finding by Gauthier and Tarr (1997) of increased activity in the right FFA following in-lab training with novel stimuli. These results, in conjunction with those of Op de Beeck et al. (2006), suggest, at minimum, that there may be constraints on the type of stimulus or training that produces greater fusiform involvement after training. More surprisingly, we found that in-lab training on a visual discrimination task also had minimal, if any, effect on activation in early (i.e. retinotopic) visual areas or areas already involved in processing the houses.

The third study, presented in Chapters 4 and 5, adapted the design from the fMRI house training study to examine the effect of training on Event-Related Potentials (ERPs). Recording ERPs permitted us to examine the effect of training on the timing of processing and the amplitude of the N1, an ERP component that is particularly large for faces (e.g., Bentin et al., 1996; Eimer, 2000a; Itier and Taylor, 2004; Jeffreys and Tukmachi, 1992), and which previous studies have shown to be affected by experience (Tanaka and Curran, 2001; Gauthier et al., 2003; Scott et al., 2006, 2008). The two previous studies that directly compared N1 amplitude from pre-test to post-test (Scott et al., 2006, 2008), found training resulted in an increase in N1 amplitude across learning sessions. Unlike these previous studies, we found that training in our task significantly decreased the amplitude of the house ERP across sessions. This decrease was especially evident around 200-300 ms after stimulus onset, corresponding approximately to the onset of the P2 and P3 components. Additional analyses comparing the ERP responses associated

with these trained houses to those of other object classes indicated that training had minimal impact on the relative ERP responses across object classes.

In Chapter 1, I reviewed several focal issues that have arisen in the face processing literature: (1) whether face processing differs qualitatively from object processing (and if so, how); (2) whether face and object processing rely on different cortical structures; and (3) whether the observed behavioural differences between face and object processing can be attributed to a face-specific system or to the cumulative effect of experience on a more generalized object-processing system. The results of the studies in this thesis address these issues by suggesting that the face inversion effect is best understood as a more general inversion effect that can be observed in other classes of objects once sufficient experience has been attained. Inversion effects, therefore, are unlikely to be the result of qualitatively different processing, nor the result of a face-specific processing system. Nonetheless, training subjects to discriminate amongst these houses does not seem to dramatically affect the cortical organization of face or object processing, nor cause the ERP or fMRI signatures of these houses to more closely resemble those of faces.

One interesting, and surprising, result obtained in the fMRI experiment concerns how training in a 4AFC task affected performance in a one-back matching task. In that experiment, subjects performed a one-back task both before and after several days of practice in a 4AFC house identification task. The data from the fMRI and EEG experiments raise important questions about the transfer of learning across the behavioural tasks that we employed. The one-back matching task was selected because it engages a subject's attention without being overly demanding, and it often is used in fMRI experiments. Because the one-back task involves discrimination of objects at an individual exemplar level – i.e., subjects are required to press a button only when two consecutive items are identical – we assumed that the learning obtained in a 4AFC identification task would transfer to the one-back task. Yet, the data from the fMRI experiment did not fully support this assumption. Subjects demonstrated an increase in accuracy in the one-back task of approximately 10% on both experimental house sets from pre-test to post-test, with no increases in accuracy on any of the other object conditions. They also performed slightly, but significantly, better on the trained house set (a difference of 1% across house sets). However, similar changes in accuracy were found in a control group of subjects who performed the one-back task in two sessions (separated by a week) without any intervening 4AFC training sessions: Accuracy increased for the trained and untrained house sets relative to other objects, although, unlike subjects in

the fMRI experiment, there was no accuracy difference between trained and untrained houses. Together, these results suggest that most of the learning observed in the fMRI subjects was *not* due to the intervening training on the 4AFC task, as initially assumed. Results from the EEG experiment were more consistent with the idea that learning in a 4AFC task should transfer to a one-back matching task: Reaction time on the one-back task decreased significantly more for the trained than untrained houses, a result that was not observed in the control subjects from the fMRI study. Overall, however, training in a 4AFC task – which produced significant improvement in that task – had surprisingly small effects on performance in the one-back task.

6.2 Future Directions

Clearly, further work to clarify the degree to which learning transfers between one-back and *n*AFC tasks would be an important pre-condition of employing these tasks together in future experimental designs. Indeed, this question is interesting in and of itself, from the perspective of better understanding the nature of learning in these tasks. Learning often is highly specific to stimulus factors such as direction of motion (e.g., Ball and Sekuler, 1982), orientation (e.g., Schoups et al., 1995, 2001; Crist et al., 1997; Husk et al., 2007; Hussain et al., 2008) or retinotopic position (e.g., Schoups et al., 1995; Crist et al., 1997). In some cases, transfer is also quite specific to the trained task. For example, subjects tested on auditory frequency and intensity discrimination following auditory frequency training exhibit less evidence of learning for the untrained intensity judgments than for the trained frequency judgments (Hawkey et al., 2004). In another case, Ahissar and Hochstein (1993) trained subjects with rectangular arrays of oriented lines. Subjects were either instructed to discriminate the orientation of the entire rectangular array (global orientation) or to discriminate the orientation of the local line elements (local orientation). Local orientation training did not transfer to global judgments; and global orientation training transfered only partially to local orientation judgments, despite the use of identical stimuli for both tasks. However, in other cases, mere exposure to the relevant stimuli can lead to perceptual learning in the context of later discrimination tasks (Dinse et al., 2006). Additional studies are needed to identify the conditions under which learning transfers across tasks.

Because of the uncertainty about the degree of learning transfer across the tasks employed in the fMRI and EEG experimental designs, some follow-up studies with a simplified design would be useful, particularly for the fMRI experiment. The complexity

of the design was less of a concern for the EEG experiment, because EEG data was collected throughout the training sessions, enabling us to examine training-related changes during the training sessions, in addition to pre- and post-testing with the one-back task. On the other hand, the fMRI design was less flexible: the cortical data was collected only in the context of the one-back task, so that it was not possible to examine cortical changes directly in the context of the 4AFC task. It would be useful to conduct an experiment in which scanning data was collected in the context of the same task as employed during training (ideally, performing the training in the context of the fMRI scanning sessions, as was possible in the EEG design). Such a replication would more convincingly determine whether the null results observed in the fMRI experiment were due to a genuine lack of training-induced cortical reorganization, or whether these results were due, merely, to a mismatch between the training and scanning tasks. If the results of this replication were consistent with the data presented in this thesis, then some additional experiments would be useful to determine why we were unable to replicate the reported expertise effect in the FFA. A replication of the original greeble study using a training paradigm more closely matched to Gauthier et al. (1998) would be quite instructive. Gauthier et al. trained subjects with a highly varied training paradigm that involved learning the objects across a large battery of tests, whereas the house training in the current study involved training on a single visual discrimination task. Although Op de Beeck et al. (2006) note that their training task was longer than that employed by Gauthier et al., the length of the training period may not be as important as the variability of training tasks: like our own study, Op de Beeck et al. employed a single training task. It may be that the training of greebles across a wide range of task contexts resulted in more extensive cortical modification. Unfortunately, because no one has directly replicated Gauthier et al.'s original experiment, it is difficult to determine whether those results are a case of Type 1 error or indicate the importance of the training paradigm necessary to invoke fusiform involvement.

The learning task we employed in these experiments could be exploited to explore other questions that relate to object learning. For example, how does stimulus use change over time? The house stimuli employed in most of these studies were intentionally designed to be distinguishable on the basis of any one featural difference (subjects could rely upon the upper windows, lower window, or door), however these features were not equated in variability. Visual inspection of the stimulus sets suggest that there was minimal variability in the doors, and greatest variability in the lower window across exemplars. Reports from a subset of subjects in the fMRI experiment suggested at least

two separate strategies were being employed across subjects: subjects relied either on the upper windows or the lower window. It would be interesting to examine such feature reliance more carefully, and to determine whether reliance on stimulus regions altered over time with learning. For example, eye-tracking could be employed to more precisely determine the reliance on particular features. One aspect of the learning involved in this task might be the identification of features that best differentiate the task exemplars. In this case, one might expect to see the pattern of eye fixations narrow across training, with less random searching over time. Eye-tracking, although a useful tool, suffers from the problem that subjects are capable of using information outside the focus of fixation. Ideally, it would be useful to complement eye-tracking studies with classification images (Murray et al., 2002). Classification images identify the locations within a stimulus that are consistently used to differentiate between exemplars. When derived using the reverse-correlation method, where images are presented in visual noise that changes randomly from trial to trial, classification images represent those locations where the added noise reliably and consistently biases the decision of the observer. Currently, this technique is readily applied to 2AFC tasks but is more difficult to adapt to the more complex 4AFC and 10AFC designs employed in this thesis. The classification image technique has been previously applied to study learning in the context of 2AFC face and texture learning (Gold et al., 2004). Because classification images generally require a great number of trials, Gold et al. were only able to compare learning in the first half and last half of trials, averaging across the first and last 6 training sessions in the process. Nonetheless, their results demonstrated that subjects adopt consistent strategies by the end of training. For faces, these strategies were not only internally consistent, but also consistent across subjects, with strong reliance on the eye-region. For textures, the strategies adopted by subjects were more idiosyncratic, with different observers relying on different parts of the images. Since the publication of this work, new methods of reducing the number of trials necessary for clear classification images have been explored (e.g., Nagai et al., 2008), increasing the potential temporal resolution of classification images across learning. In light of these advances, it would be interesting to extend this earlier learning work to look more closely at the time course of learning across sessions.

6.3 Conclusions

The studies that compose this thesis share a common underlying logic. As noted in Chapter 1, this thesis addresses questions about the extent to which faces are special objects by examining the training of expertise for classes of non-face objects. This ap-

proach is based on the following assumption: If we can induce, through training alone, phenomena that normally occur only with faces, then it is not necessary to interpret these phenomena as being the result of face-specific perceptual processing. For example, if training of non-face objects produces inversion effects that bear the same qualities as face inversion effects, then it is not necessary to assume that a special face processor produces the face inversion effect. It is important to note that this approach cannot rule out the existence of a special face processor. Demonstrating that characteristics of face processing can be elicited by non-face objects implies only that a special face processor is not *necessary* to elicit these characteristics, but it remains possible that the same phenomena can be separately induced by a specialized face processor and by a more generalized object processor. Face inversion effects and trained non-face inversion effects may share many of the same qualities yet arise from different underlying mechanisms. How, then, has this work advanced our understanding of face processing? I have pointed out that the existence of parallel face and object processing cannot be ruled out by showing that training alone can induce face-like processing. Nonetheless, I would argue that it is more parsimonious to assume a single underlying process. The existence of a special face processor was first posited because faces and objects exhibit different behavioural phenomena: by showing that these behavioural differences disappear with training, the onus is placed on those that argue for a specialized face mechanism to demonstrate that such mechanisms are necessary to understand face processing.

References

Ahissar, M., Hochstein, S., 1993. Attentional control of early perceptual learning. Proc Natl Acad Sci U S A 90 (12), 5718–5722.

Ball, K., Sekuler, R., 1982. A specific and enduring improvement in visual motion discrimination. Science 218 (4573), 697–698.

Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience 8 (6), 551–565.

Crist, R. E., Kapadia, M. K., Westheimer, G., Gilbert, C. D., 1997. Perceptual learning of spatial localization: specificity for orientation, position, and context. J Neurophysiol 78 (6), 2889–2894.

Dinse, H. R., Kleibel, N., Kalisch, T., Ragert, P., Wilimzig, C., Tegenthoff, M., 2006.

Tactile coactivation resets age-related decline of human tactile discrimination. Ann Neurol 60 (1), 88–94.

Eimer, M., 2000a. Attentional modulations of event-related brain potentials sensitive to faces. Cognitive Neuropsychology 17 (1-3), 103–116.

Gauthier, I., Curran, T., Curby, K. M., Collins, D., 2003. Perceptual interference supports a non-modular account of face processing. Nat Neurosci 6 (4), 428–432.

Gauthier, I., Tarr, M. J., 1997. Becoming a "greeble" expert: exploring mechanisms for face recognition. Vision Res 37 (12), 1673–1682.

Gauthier, I., Williams, P., Tarr, M. J., Tanaka, J., 1998. Training 'greeble' experts: a framework for studying expert object recognition processes. Vision Res 38 (15-16), 2401–2428.

Gold, J., Sekuler, A., Bennett, P., 2004. Characterizing perceptual learning with external noise. Cognitive Science 28 (2), 167–207.

Hawkey, D. J. C., Amitay, S., Moore, D. R., 2004. Early and rapid perceptual learning. Nat Neurosci 7 (10), 1055–1056.

Husk, J. S., Bennett, P. J., Sekuler, A. B., 2007. Inverting houses and textures: investigating the characteristics of learned inversion effects. Vision Res 47 (27), 3350–3359.

Hussain, Z., Sekuler, A. B., Bennett, P. J., 2008. Robust perceptual learning of faces in the absence of sleep. Vision Res 48 (28), 2785–2792.

Itier, R. J., Taylor, M. J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cereb Cortex 14 (2), 132–142.

Jeffreys, D. A., Tukmachi, E. S., 1992. The vertex-positive scalp potential evoked by faces and by objects. Exp Brain Res 91 (2), 340–350.

Murray, R. F., Bennett, P. J., Sekuler, A. B., 2002. Optimal methods for calculating classification images: weighted sums. J Vis 2 (1), 79–104.

Nagai, M., Bennett, P. J., Sekuler, A. B., 2008. Exploration of vertical bias in perceptual completion of illusory contours: Threshold measures and response classification. J Vis 8 (7), 25.1–17.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., Kanwisher, N. G., 2006. Discrimination training alters object representations in human extrastriate cortex. J Neurosci 26 (50), 13025–13036.

Schoups, A., Vogels, R., Qian, N., Orban, G., 2001. Practising orientation identification improves orientation coding in v1 neurons. Nature 412 (6846), 549–553.

Schoups, A. A., Vogels, R., Orban, G. A., 1995. Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. J Physiol 483 (Pt 3), 797–810.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., Curran, T., 2006. A reevaluation of the electrophysiological correlates of expert object processing. J Cogn Neurosci 18 (9), 1453–1465.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., Curran, T., 2008. The role of category learning in the acquisition and retention of perceptual expertise: a behavioral and neurophysiological study. Brain Res 1210, 204–215.

Tanaka, J. W., Curran, T., 2001. A neural basis for expert object recognition. Psychol Sci 12 (1), 43–47.