Conjunctive Coding of Complex Object Features

by

Jonathan Erez

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Psychology
University of Toronto

© Copyright by Jonathan Erez 2015
Abstract

Critical to perceiving an object is the ability to bind its constituent features into a cohesive representation, yet the manner by which the visual system integrates object features to yield a unified percept remains unknown. Another important aspect of object perception is the ability to represent complex objects across different viewing angles, despite the drastic variability in appearance caused by shifting viewpoints. This dissertation describes a novel application of multivoxel pattern analysis (MVPA) of neuroimaging data that allows a direct investigation of whether neural representations integrate object features into a whole that is different from the sum of its parts and whether these representations are view invariant. Results showed that patterns of activity throughout the ventral visual stream (VVS), extending anteriorly into the perirhinal cortex (PRC), discriminated between the same sets of features when combined into different objects. Despite this sensitivity to the unique conjunctions of features, this activity was invariant to the viewpoints from which the conjunctions were presented. These results suggest that the manner in which our visual system processes complex objects depends on the explicit coding of the feature conjunctions comprising them, and that these representations are invariant to large changes in visual appearance caused by shifting viewpoints. To test whether experience with the objects changed the pattern of conjunctive representations across the VVS, participants
were scanned in a separate fMRI study following an extensive behavioural training period. Results showed that both view-dependent and view-invariant conjunctive representations shifted to more posterior regions of the VVS, suggesting that training affected how unique feature conjunctions are represented across the VVS. These results support the representational-hierarchical view of cortical organization and contribute to our understanding of object perception and memory.
Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Morgan Barense, for her support and guidance throughout my PhD. Thank you for your patience, generosity and for encouraging me to become a better researcher and writer. I would also like to thank the members of my thesis committee, Dr. Susanne Ferber and Dr. Brad Buchsbaum for their valued input and encouragement during this process. In addition, I wish to thank Dr. Rhodri Cusack for his theoretical and technical expertise.

Many thanks to my labmates, Rachel Newsome, Lok-Kin Yeung, Danielle Douglas, and Celia Fidalgo and the rest the members of the Barense Lab for their friendship and for fruitful discussions over the years. Special thanks to Felicia Zhang and Will Kendall for their assistance with the experiments in this thesis, and thanks to all of the participants that agreed to be scanned, and rescanned.

I also would like to thank NSERC, OGS, and the University of Toronto for their generous financial support of this research, as well as the hot dog van outside Sid Smith for staying open late most evenings to benefit from the aforementioned funding.

I am grateful for my wife, Noam, for her unwavering support during this process. Your love and patience made this possible. Lastly, I’d like to thank my parents who inspired me to pursue my passions and instilled in me a thirst for knowledge.
Contents

Acknowledgments........................................................................................................ iv

List of Tables .................................................................................................................. ix

List of Figures ................................................................................................................ x

List of Appendices ......................................................................................................... xii

General Introduction ..................................................................................................... 1

1.1 The binding problem ............................................................................................... 1

1.2 The view-invariance problem .................................................................................. 4

1.3 How might the brain solve these problems? ............................................................... 7

1.4 fMRI MVPA ............................................................................................................. 12

1.5 Thesis overview ....................................................................................................... 15

Chapter 2: Neural representations of feature conjunctions composing complex objects .......... 17

2.1 Introduction ............................................................................................................. 17

2.2 Materials and Methods .......................................................................................... 23

Participants ................................................................................................................... 23

Experimental Stimuli and Design ............................................................................... 23

Tasks .............................................................................................................................. 24

fMRI Data Acquisition ............................................................................................... 27

Multi-Voxel Pattern Analysis (MVPA) .......................................................................... 27

Control analysis to test for sub-additivity of the BOLD signal .................................. 33
2.3 Results and Discussion........................................................................................................35

Chapter 3: Neural representations of feature conjunctions composing complex objects: View

Invariance...................................................................................................................................42

3.1 Introduction..........................................................................................................................42

3.2 Materials and Methods.......................................................................................................44

Participants...............................................................................................................................44

Experimental Stimuli and Design.........................................................................................44

Image similarity analysis. .......................................................................................................44

Experimental task...................................................................................................................48

3.3 Results and Discussion.......................................................................................................48

Chapter 4: Experience modifies the conjunctive neural representations necessary for complex

object perception: behavioural training ................................................................................53

4.1 Introduction........................................................................................................................53

4.2 Materials and Methods.......................................................................................................57

Participants...............................................................................................................................57

Materials..................................................................................................................................57

Design and Behavioural Procedure.......................................................................................59

Data analysis............................................................................................................................61

4.3 Results and discussion........................................................................................................62

Visual search performance across training sessions ..........................................................62
Effect of viewpoint congruency between sample target and choice target ........................................ 65

4.4 General Discussion .................................................................................................................. 70

Chapter 5: Experience modifies the conjunctive neural representations necessary for complex
object perception .......................................................................................................................... 74

5.1. Introduction .......................................................................................................................... 74

5.2 Materials and Methods ........................................................................................................ 77

Participants .................................................................................................................................. 77

Experimental Stimuli and Design ............................................................................................... 77

Tasks ........................................................................................................................................... 80

fMRI Data Acquisition ................................................................................................................ 82

Multi-Voxel Pattern Analysis (MVPA) ......................................................................................... 82

5.3 Results .................................................................................................................................. 88

5.4 Discussion ............................................................................................................................... 92

Chapter 6: Experience modifies the conjunctive neural representations necessary for complex
object perception: Effect of viewpoint .......................................................................................... 97

6.1. Introduction .......................................................................................................................... 97

6.2 Materials and Methods ......................................................................................................... 99

Multi-Voxel Pattern Analysis (MVPA) ......................................................................................... 99

6.3 Results .................................................................................................................................. 101

6.4 Discussion ............................................................................................................................... 104
Chapter 7: General Discussion ........................................................................................................ 107

Conjunctive representations in the VVS are dynamic ................................................................. 108

Methodological considerations ................................................................................................. 110

Future studies ............................................................................................................................... 111

Conclusion ................................................................................................................................... 114

References .................................................................................................................................... 115

Appendices .................................................................................................................................... 131

Copyright Acknowledgements ........................................................................................................ 156
List of Tables

Table 2.1. Behavioural results of the 1-back task completed during scanning………………….24

Table 5.1. Behavioural results of the task completed in the scanner…………………………..80
List of Figures

Figure 1.1 The Representational-Hierarchical Model ............................................................. 7

Figure 1.2. MVPA versus a conventional univariate analysis ......................................................... 13

Figure 2.1 Stimuli and task ........................................................................................................ 19

Figure 2.2 Experimental question and similarity contrast matrix ................................................. 21

Figure 2.3 Post-scan memory task ................................................................................................ 27

Figure 2.4 The full data matrix and data analysis structure .......................................................... 30

Figure 2.5 Control analysis to test for sub-additivity of the BOLD signal ...................................... 34

Figure 2.6 Regions demonstrating explicit conjunctive coding .................................................... 36

Figure 2.7 Regions demonstrating single-feature coding ............................................................. 39

Figure 3.1 Experimental question and similarity contrast matrix ................................................ 43

Figure 3.2 An analysis of visual differences across objects using a root-mean-square difference (RMSD) measure on the RGB values of all images used in the experiment ................. 47

Figure 3.3 Regions demonstrating explicit conjunctive coding across viewpoints ....................... 50

Figure 4.1 The stimuli used by Czerwinski et al. (1992) ............................................................ 56

Figure 4.2 Sample stimuli use in the experiment ......................................................................... 59

Figure 4.3 Visual search task: examples of (A) a viewpoint-congruent target-present trial and (B) a viewpoint-incongruent target-present trial for a set size of 5 ........................................ 60

Figure 4.4 Mean reaction times by set size (A) across 5 training sessions and (B) for trained and untrained objects on the last training session ................................................................. 63

Figure 4.5 Mean search slopes by training session ........................................................................... 64

Figure 4.6. Mean accuracy across training sessions ..................................................................... 65
Figure 4.7 RT viewpoint incongruency costs by set size (A) for training sessions 1 through 5, and (B) for trained and novel objects on the last training session. ........................................... 67

Figure 4.8 Viewpoint incongruency accuracy costs by set size (A) for training sessions 1 through 5, and (B) for trained and novel objects on the last training session. ........................................... 68

Figure 4.9 Viewpoint incongruency analysis: speed-accuracy trade-off across training sessions. ........................................................................................................................................................................ 69

Figure 5.1 Stimuli used. ........................................................................................................................................................................................................................................................................... 79

Figure 5.2 Contrast matrices used to test four experimental questions. .................................................. 85

Figure 5.3 Regions demonstrating explicit conjunctive coding for novel and trained objects... 91

Figure 5.4 The proposed organization of visual object representations in the VVS according to the Representational-Hierarchical Model. ........................................................................................................................................................................ 95

Figure 6.1 Contrast matrices used to test three experimental questions. ............................................... 100

Figure 6.2 Regions demonstrating view-invariant conjunctive coding for novel and trained objects. ........................................................................................................................................................................ 103
List of Appendices

**Appendix 1:** Regions demonstrating explicit conjunctive coding: analyses computing correlations across runs only.

**Appendix 2:** Peak ROI coordinates for all participants, as defined by our functional localizer for the lateral occipital cortex (LOC), fusiform face area (FFA) and parahippocampal place area (PPA).

**Appendix 3:** (A) Regions that showed an effect of feature type in the univariate ANOVA (Figure 2.5) for feature set 1 (p < 0.001, uncorrected). (B) Regions that showed an effect of feature type in the univariate ANOVA (Figure 2.5) for feature set 2 (p < 0.001, uncorrected).

**Appendix 4:** *Conjunction Contrast:* Searchlight MVPA. Regions where the representation for the conjunction was different from the sum of its parts, shown at a more liberal threshold of p < 0.001 (uncorrected).

**Appendix 5:** (A) Regions identified by the *Conjunction Contrast* in the searchlight MVPA (whole-brain FWE-corrected, p < 0.05). (B) Regions identified by the Conjunction Contrast in the searchlight MVPA (p < 0.001, uncorrected).

**Appendix 6:** (A) Regions identified by *Viewpoint Contrast* in the searchlight MVPA (whole-brain FWE-corrected, p < 0.05). (B) Regions identified by Viewpoint Contrast in the searchlight MVPA (p < 0.001, uncorrected).

**Appendix 7:** Regions demonstrating view-invariant explicit conjunctive coding: analyses computing correlations across runs only.

**Appendix 8:** Regions identified by the *Conjunctive Coding for Novel Objects* contrast in the searchlight MVPA (p < 0.001, uncorrected).
Appendix 9. Regions identified by the *Conjunctive Coding for Trained Objects* contrast in the searchlight MVPA (p < 0.001, uncorrected).

Appendix 10. Regions identified by the *Conjunctive Coding for Trained > Novel Objects* contrast in the searchlight MVPA (p < 0.001, uncorrected).

Appendix 11. Regions identified by the *Conjunctive Coding for Novel > Trained Objects* contrast in the searchlight MVPA (p < 0.001, uncorrected).

Appendix 12. Regions identified by the *View-Invariant Conjunctive Coding for Novel Objects* contrast in the searchlight MVPA (p < 0.005, uncorrected).

Appendix 13. Regions identified by the *View-Invariant Conjunctive Coding for Trained Objects* contrast in the searchlight MVPA (p < 0.005, uncorrected).

Appendix 14. Regions identified by the *View-Invariant Conjunctive Coding for Trained > Novel Objects* contrast in the searchlight MVPA (p < 0.005, uncorrected).

Appendix 15. Regions identified by the *View-Invariant Conjunctive Coding for Novel > Trained Objects* contrast in the searchlight MVPA (p < 0.005, uncorrected).
General Introduction

How objects are represented in the brain is a core issue in neuroscience. In order to coherently perceive even a single object, the visual system must integrate its features (e.g., shape, color) into a unified percept (sometimes called the “binding problem”) and recognize this object across different viewing angles, despite the drastic variability in appearance caused by shifting viewpoints (the “invariance problem”). These are the most computationally demanding challenges faced by the visual system, yet humans can perceive complex objects across different viewpoints with exceptional ease and speed (Thorpe, Fize, & Marlot, 1996). The mechanism underlying this feat is one of the central unsolved puzzles in cognitive neuroscience and is the focus of this thesis.

1.1 The binding problem

The concept of binding is a central topic in cognitive neuroscience. Binding refers to the ability to combine different aspects of the environment into a coherent representation. Many instances of binding exist – we perform auditory binding when listening to music to combine several notes together in order to form one cohesive melody, or when trying to match a voice in the crowd with a specific person who is calling our name. We perform cognitive binding when we consciously or unconsciously match concepts and the percepts they might invoke (or vice versa). In the context of visual cognition, binding refers to the process by which we are able to comprehend our visual environment, such as perceiving several objects to perceive one unified scene. At a more elementary level, even perceiving just a single object requires binding – one needs to be able to combine basic features of an object (e.g., shape, color, size, location etc.) to
create a unified representation of that object (Engel, Fries, König, Brecht, & Singer, 1999; Roskies, 1999; Treisman, 1996). This ability is such a key part of human information processing that some researchers and philosophers postulate that it is an essential requirement to having a phenomenological experience, and that finding a solution of the binding problem might be an important part of a solution to the mystery of consciousness itself (Koch & Crick, 1994; Roskies, 1999). However, there are instances in which the need to recognize and combine these features proves to be an obstacle. For example, when attention is diverted or overloaded, features may be combined inaccurately, creating what is known as “illusory conjunctions” (Treisman & Schmidt, 1982). In addition, people with brain damage have difficulty binding object features (Barense, Gaffan, & Graham, 2007; Treisman & Kanwisher, 1998). From a neurophysiological view, evidence from many studies suggest that the different features that compose an object are processed in separate parts of the brain, and it is not well understood how all of this decoupled information is combined in the brain to create the perception of a cohesive object. The focus of this thesis is to understand how this binding process is accomplished at the single object level – how does the brain combine multiple features to create a cohesive representation of a single complex object?

Early suggestions that binding must exist in the brain stemmed from a reasoning that otherwise, there are just too few neurons in the brain to represent all possible percepts (Engel, Konig, Kreiter, Schillen, & Singer, 1992; Ghose & Maunsell, 1999). For example, if each neuron represents a specific manifestation of an object’s attribute (e.g. the color blue, horizontal orientation, a specific length, etc.) then many more neurons would be needed to represent each possible combination of attributes of just a single object. This problem, often referred to as combinatorial explosion (Ghose & Maunsell, 1999), quickly becomes computationally
intractable with each added dimension – for example, if 100 neurons are needed to represent all possible manifestations of just 6 different attributes, then $100^6$ neurons would be needed to represent each possible combination of these 6 attributes – more neurons in the entire brain by some estimates (Williams & Herrup, 1988). This problem can be avoided, however, if information from separate neurons is somehow combined.

Over the years, many theories have been proposed to explain the binding of object features. Two of the most influential classes distinguish between “hardwired” and “on-demand” feature conjunctions (Vanrullen, 2009). The hardwired models assume explicit conjunctive coding of bound features in which the neural representation for the conjunction is different from the sum of the individual parts. That is, these models suggest that there are neurons (perhaps in specific brain regions) that store representations for the conjunction of 2 or more basic features. By contrast, the on-demand models hypothesize that unitized objects are processed through a non-local binding mechanism, a process that does not necessitate explicit conjunctive coding of object features per se, but rather, the features are represented independently and bound by co-activation. Such a mechanism could include the synchronized activity of spatially-distributed neurons that represent the individual object features (Singer & Gray, 1995; Uhlhaas et al., 2009; Von Der Malsburg, 1981), or a separate brain region that temporarily reactivates and dynamically links of otherwise disparate feature representations (Damasio, 1989; Devlin & Price, 2007; Eckhorn, 1999; Meyer & Damasio, 2009). Thus, an explicit conjunctive coding mechanism predicts that the neural representation for a whole object should be different from the sum of its parts, whereas a non-local binding mechanism predicts that the representation for the whole object should not be different from the sum of the parts, because the unique conjunctive representations are never directly coded.
Regardless of the neural mechanism, being able to learn from one’s visual environment to form bound representations over time has several advantages. For example, it would confer a survival advantage to be able to discriminate all of the many yellow things and all of the many moving things in the environment to identify quickly that a particular yellow-moving object is a fast-approaching lion (Frank, Reavis, Tse, & Greenlee, 2014). Having instant access to already bound information can also be advantageous compared to invoking the individual features that comprise the object (lion in this example) on demand, because this process might be faster. Previous studies have shown that over time, participants do get faster and make less errors on tasks that require feature conjunctions (e.g., Carrasco, Ponte, Rechea, & Sampedro, 1998; Czerwinski, Lightfoot, & Shiffrin, 1992; Sireteanu & Rettenbach, 1995). Storing explicit feature conjunction and having immediate access to these can also allow the brain to be more efficient and allocate its limited resources elsewhere.

1.2 The view-invariance problem

The ability to coherently perceive even a single object requires not only integrating its features into a cohesive unit, but also recognizing it across multiple viewpoints. Recognizing objects from different viewpoints is a challenge for the visual system because a shift of viewpoint can often cause a dramatic change in the object’s appearance. This is one of the most computationally demanding challenges faced by the visual system, yet humans can perceive complex objects across different viewpoints with exceptional ease and speed (Thorpe et al., 1996).

One model suggests that this is accomplished through a view-invariant representation of objects in our visual system. This model assumes that objects are represented in our visual system relatively independently of the viewing angle from which they are presented. Marr and
Nishihara (1978) first formalized this mechanism with a computational model in which 3D objects are comprised of part-based structural representational volumes and the spatial relations between them. According to this model, the object-parts (usually generalized cones or cylinders) are represented in an object-centered manner – that is, the orientation of the object and the position of the viewer are decoupled. However, the model still faces the challenge of explaining how such objects remain stable in vision even when rotations in depth can cause their 2D retinal projection to change dramatically. Marr (1982) proposed a framework in which 3D perception is accomplished via a series of successive feed-forward steps that can be divided into three stages of increasing complexity: low-level, intermediate, and high-level. The most basic level, the “primal sketch” (V1), detects basic visual features such as edges, bars, lines, curves boundaries, and changes in intensity and illumination. The intermediate stage, termed “2.5D sketch”, is responsible for detecting local surface orientation, discontinuities in depth, luminance boundary and texture segmentation. Finally, a high-level stage includes a 3D model representation (IT cortex) which is hierarchically organized in terms of surface and volumetric primitives.

Biederman (1987) expanded Marr’s model and proposed a Recognition by Components (RBC) mechanism, according to which objects are constructed from a series of basic 3D individual building blocks (i.e., “geons”) and the relationship between them. View-invariant representations of complete objects are possible because the geons, which themselves are view-invariant, are connected through a series of basic conjunctions that remain stable with an object’s rotation. Biederman described such objects as having “non-accidental properties” – image properties that remain stable across changes in viewpoint (i.e., they are not an accident of view and can provide reliable information about 3D structure). One intriguing implication of these view-invariant models is that they assume that observers should have no difficulty recognizing familiar objects.
from novel viewpoints. That is, if there is no cost associated with recognizing an object from unfamiliar angles, response times and error rates should be roughly equivalent across different viewpoints.

An alternative model suggests that the visual system is in fact viewpoint-dependent. According to this model, invariant object recognition is possible through exposure to a series of 2D views of an object that together create a 3D representation of that object over time (e.g. Bülthoff & Edelman, 1992; Tarr, 1995), and invariant object recognition is eventually accomplished by interpolating across these multiple views (Logothetis et al., 1994; Poggio & Edelman, 1990). Because this process is not immediate, this model predicts that there will be a cost associated with viewing an object from novel viewpoints. For example, in one study, participants were trained to name objects at a specific orientation, and after extensive training were asked to name the same objects when these were shown from unfamiliar viewpoints (Tarr & Pinker, 1989). Participants were progressively slower (and made more errors) naming the objects at unfamiliar orientations that were further from the trained viewpoint. Similar effects were observed when testing participants with novel 3D objects rotated in depth (Tarr, 1995).

An important point worth emphasising is that these two competing models differ in how they account for the role of experience in creating view-invariant representations. The view-invariant model suggests that observers immediately exploit view-invariant information when encountering a never-before seen object (Biederman & Bar, 1999) and as a result, extensive experience with the object is not necessarily needed. In contrast, the view-dependent model assumes that view-invariant information is something that is learned over time, as an observer collects more evidence with multiple exposures to an object from several different viewpoints. As such, these two views make different predictions about the role of behavioural training in
shaping view-invariant representations – the former suggests that not much training is needed whereas the latter predicts that multiple training sessions are required to achieve view-invariance.

1.3 How might the brain solve these problems?

**Figure 1.1 The Representational-Hierarchical Model.** (A) Lateral view of the cerebral cortex showing the ventral visual stream (VVS) object processing pathway according to the representational-hierarchy model (Cowell, Bussey & Saksida, 2010). Basic visual information is processed in posterior regions of the stream (e.g. V1) and simple feature conjunctions in more anterior regions. The perirhinal cortex is proposed to reside at the apex of this processing pathway, containing representations for conjunction at the level of an object. (B) The proposed organization of visual object representations in the VVS. A, B, C, and D refer to relatively simple object features represented in posterior regions of the VVS. Simple feature conjunctions are represented in more anterior regions, and more complex conjunctions, approximately at the level of an object, are represented in the more anterior regions, including perirhinal cortex.

As visual information flows through successive stages of the ventral visual stream (VVS), a series of anatomically linked cortical fields originating in V1 and extending into the temporal lobe (Desimone & Ungerleider, 1989; Felleman & Van Essen, 1991; Gross, 1992; Hubel & Wiesel, 1965; Riesenhuber & Poggio, 1999), representations of stimulus features are organized hierarchically, with increasing levels of complexity (Desimone & Ungerleider, 1989;
Viewing a single object activates multiple representations of the object along the VVS, with basic object features presented in posterior regions of the VVS and the conjunctions of these features represented in more anterior regions. It was traditionally believed that the inferotemporal cortex (area IT) was the most anterior region in which objects are represented (Tanaka, 1996). According to this theory, the adjacent medial temporal lobe (MTL) was thought to be dedicated for a separate cognitive process, memory, with no role in perception (Squire & Wixted, 2011; Clark, Reinagel, Broadbent, Flister, & Squire, 2011; Kim et al., 2011). However, several studies have suggested that some structures in the MTL, in particular the perirhinal cortex (PRC), are not only important for memory but also essential for certain forms of perception (Barense, Ngo, Hung, & Peterson, 2012; Barense et al., 2005, 2007; Bartko, Winters, Cowell, Saksida, & Bussey, 2007; Baxter, 2009; Buckley, Booth, Rolls, & Gaffan, 2001; Lee, Buckley, et al., 2005; Lee, Bussey, et al., 2005), and that the MTL is likely important for categorization, long thought to be an implicit and non-MTL process (Nosofsky, Denton, Zaki, Murphy-Knudsen, & Unverzagt, 2012).

One theory that tries to consolidate perceptual and mnemonic functions is the representational-hierarchical model (Bussey, Saksida, & Murray, 2002; Saksida & Bussey, 2010). According to this model, the ventral visual stream (VVS) and MTL together comprise a processing stream devoted to the representation of objects, with different levels of representation at different stages in the pathway. While lower-level features of an object are processed by posterior regions of the VVS, the conjunctions of these features (approximately at the level of an object) are processed in more anterior regions that are classically associated with memory, such as the PRC (Figure 1.1). This model proposes that these representations are used for any cognitive function that requires them – be it perception or memory (Bussey & Saksida, 2002;
Cowell, Bussey, & Saksida, 2006), and that damage to anterior regions of the stream will compromise the integrity of complex object representations and this will impair both object perception and memory (Bussey & Saksida, 2007; Cowell, Bussey, & Saksida, 2006; 2010). Support for this reasoning comes from studies that have used a variety of paradigms to test perceptual discrimination ability in rats, monkeys, and humans with damage to the MTL (e.g. Barense, Ngo, Hung, & Peterson, 2012; Bartko, Winters, Cowell, Saksida, & Bussey, 2007; Buckley, Booth, Rolls, & Gaffan, 2001; Lee & Rudebeck, 2010; Newsome, Duarte, & Barense, 2012). The critical factor for PRC engagement on either perceptual (e.g., Bussey et al. 2002; Bartko et al. 2007; Lee and Rudebeck 2010; Barense, Ngo, et al. 2012; Barense et al. 2012) or memory tasks (e.g., Barense et al., 2005; Bartko, Cowell, Winters, Bussey, & Saksida, 2010; Bussey et al., 2002; McTighe, Cowell, Winters, Bussey, & Saksida, 2010; Winters, Forwood, Cowell, Saksida, & Bussey, 2004) is the visual complexity of the stimuli – more specifically, the requirement to represent complex conjunctions of features.

Yet it is not clear whether the complex feature conjunctions processed in the PRC are a result of explicit conjunctive coding or a non-local binding mechanism. Most empirical attention so far has focussed on area IT in monkeys and the object-selective lateral occipital complex (LOC) in humans – structures posterior to PRC and traditionally thought to be the anterior pinnacle of the VVS (Denys et al., 2004; Grill-Spector, Kourtzi, & Kanwisher, 2001; Kriegeskorte, Mur, Ruff, et al., 2008; Sawamura, Georgieva, Vogels, Vanduffel, & Orban, 2005; Ungerleider & Haxby, 1994). For example, single-cell recording in monkey IT cortex showed evidence for conjunctive processing whereby responses to the whole object could not be predicted from the sum of the parts (Baker, Behrmann, & Olson, 2002; Desimone, Albright, Gross, & Bruce, 1984; Gross, 2008), although these conjunctive responses might have arisen
from sensitivity to new features created by abutting features (Sripati & Olson, 2010). That said, there is also some support for a non-local binding mechanism interpretation. For example, a recent study found that in most neurons in a monkey’s IT cortex, information from a shape-color combination summed linearly, whereas only a few neurons showed conjunction selectivity to a shape-color combination (McMahon & Olson, 2009). Another recent study reported that a pattern-classification algorithm could be used to predict a target’s identity in the (left) anterior temporal lobe from the temporal convergence of information regarding its shape and color from earlier visual regions (LOC and right V4, respectively) (Coutanche & Thompson-Schill, 2014). However, no study so far has tested the notion that unique conjunctions of bound object features contribute differentially to the same overall configuration of a single object.

An important potential benefit of explicit conjunctive coding of whole objects is to provide stability of representation across changes in viewpoints, and invariance to the manifestation of individual object features (Biederman, 1987). However, the extent to which the human visual system supports view-invariant versus view-dependent representations of objects is unresolved (Biederman & Bar, 1999; Peissig & Tarr, 2007). Evidence from both single-cell recordings and fMRI studies suggest that posterior regions of the VVS are view-dependent whereas more anterior regions are view-invariant, although this evidence is far from conclusive. For example, Logothetis & Pauls (1995) measured single-unit monkey IT recordings in the superior temporal sulcus (STS) and anterior medial temporal sulcus (AMTS) while the monkeys viewed objects from familiar viewpoints. Approximately 10% of the recorded neurons were view-dependent and only about 1% of neurons were view-invariant. In another study, Booth & Rolls (1998) recorded from single STS neurons of monkeys after they trained with a set of objects for several weeks. They also found that the majority of neurons (49%) were selective for
specific viewpoints whereas only 14% were view-invariant. These two studies suggest that even though IT contains view-invariant neurons, these are not prevalent and that perhaps view-invariant perception is achieved through a pooling across multiple view-selective neurons (Perrett, Oram, & Ashbridge, 1998).

fMRI studies have reported mixed results regarding the ability of regions within the VVS to represent view-invariant objects. For example, results from several fMRI adaptation experiments suggest that object representations in the LOC are view-dependent (Fang & He, 2005; Isabel Gauthier et al., 2002; Grill-Spector et al., 1999), whereas others showed that the distributed LOC responses to a particular object remained stable across 60° rotations (Evelyn Eger, Ashburner, Haynes, Dolan, & Rees, 2008). Several neuroimaging studies have reported both view-selective and view-invariant representation of up to 60° for different object categories (Andresen, Vinberg, & Grill-Spector, 2009; Vuilleumier, Henson, Driver, & Dolan, 2002) leading some to conclude that there is no complete invariant object perception in the VVS (Andresen et al., 2009). However, other studies suggest that view-invariance might be achieved in more advanced regions in the representational hierarchy. For example, Quiroga et al. (2005) found that MTL neurons were invariant to different representations of the same concept, and nonhuman primates and humans with perirhinal cortex damage were impaired when discriminating complex objects and faces presented from different viewpoints (Buckley et al., 2001; Barense et al., 2005; 2007; Lee et al., 2006b). Consistent with this, an fMRI study involving neurologically intact participants reported that the PRC demonstrated increased activity when participants discriminated between faces and objects presented from different, relative to identical, viewpoints (Barense, Henson, Lee, & Graham, 2010).
1.4 fMRI MVPA

Traditionally, analysis of fMRI data employed a mass univariate approach – testing the relationship between cognitive states and individual brain voxels. However, the relatively recent neuroimaging method of multi-voxel pattern analysis (MVPA) offers promise for making more subtle distinctions between representational content than previously possible (Haxby et al., 2001; Kamitani & Tong, 2005). While a conventional univariate approach aims to identify whether a brain region is involved in a given cognitive process, MVPA aims to detect activity-pattern differences and inform about the kind of information is represented in a given brain region (Mur, Bandettini, & Kriegeskorte, 2009). MVPA is therefore well suited to address the questions posed in this thesis because it allows the differentiation of subtle differences in representational content – in our case, potential differences between representations of the same object that is composed of different unique conjunctions of the same features.

There are several benefits to using MVPA compared to individual-voxel-based methods. The main advantage is increased sensitivity (Haynes & Rees, 2006; Jimura & Poldrack, 2012; Norman, Polyn, Detre, & Haxby, 2006). In a traditional fMRI analysis (i.e., univariate), multiple voxels that respond to a particular condition are spatially averaged. This process reduces noise, but also reduces some of the fMRI signal as well – voxels with a weaker signal are often ignored even though they might still be informative in detecting the presence/absence of a condition. Moreover, spatial averaging removes the fine-grained patterns that might discriminate between conditions (Figure 1.2) (Norman et al., 2006). In comparison, MVPA involves extracting signal from multiple voxels and examining their pattern of activity even if none of the voxels respond to the conditions of interest individually.
Figure 1.2. MVPA versus a conventional univariate analysis. MVPA allows the detection of representational content that can be overlooked by a conventional fMRI analysis. Shown above are hypothetical activations for two complex objects represented in an ROI consisting of 13 voxels. The multi-voxel patterns for these two objects are different, indicating that the ROI represents information about these objects (reflecting differences in underlying neuronal population activity). A conventional univariate analysis, however, would not detect a difference between the two objects because the regional average is the same – there is no overall net change in activity. Figure adapted from (Mur et al., 2009; Raizada, Tsao, Liu, & Kuhl, 2010).

Another advantage of MVPA is the ability to relate brain activity to behavior on a trial-by-trial basis – it can measure the presence or absence of a cognitive state based on only a few seconds of brain activity. For example, in one study, MVPA was used to predict the time course in a free recall task and show that patterns of activity correlated with participants’ verbal responses (Polyn, Natu, Cohen, & Norman, 2005). MVPA also allows the possibility of designing neuroimaging studies that address hypotheses which cannot be addressed with a mass-univariate methods (Etzel, Zacks, & Braver, 2013), such as the studies described in this thesis – for instance, here we are looking for fine-grained differences between representations of the same object that is composed of different unique conjunctions of the same features. A univariate analysis would likely not be able to distinguish between these activity patterns because we would
not expect the various combinations of features to produce different overall signal levels. The differences are likely to be in how the information is coded (measured by MVPA), rather than how much activity is produced (measured by univariate approaches).

The increasing popularity of MVPA over the past several years led to the development of several different analytical approaches. Two of the most common ones are (1) pattern-classification algorithms, many of which are borrowed from the machine-learning literature (Kamitani & Tong, 2005; Pereira, Mitchell, & Botvinick, 2009), and (2) a representational similarity analysis (RSA) approach, which relies on correlations of activation patterns (Kriegeskorte, Mur, & Bandettini, 2008). Pattern classification analyses involve dividing the data into independent training and testing data sets, determining decision rules, and testing the validity of the classifier on an independent test data. The classifier’s accuracy is then determined by the percentage of test instances that are correctly classified. In contrast, an RSA assesses the similarity between multi-voxel patterns produced by the different stimuli used in an experiment (Kriegeskorte & Kievit, 2013). An inspection of the similarity structure can then reveal insights about the relationship between the different stimuli categories used (e.g., is there a higher correlation between activation patterns for stimuli from the same conceptual or perceptual category?). This is one useful advantage of RSA over a pattern classification approach – it can reveal if two brain regions have a different structure, whereas a pattern classification algorithm may find only that the classification accuracy is equivalent in those regions (Haxby et al., 2014).

A common method in both MVPA approaches is to utilize a searchlight analysis in which a multivariate analysis is performed on small sphere-shaped groups of voxels centered on each and every voxel in the brain. This method produces a statistical map indicating how well the multivariate pattern in each location represented/differentiated each experimental condition
or stimulus class (Kriegeskorte, Goebel, & Bandettini, 2006). The advantage of this whole-brain searchlight approach is that a priori hypotheses about the function of a specific brain structure are not needed. A complementary approach that can also be employed is to conduct an ROI-based MVPA, an analysis in which the activation patterns are compared only within specific structures that are determined a-priori.

In the current thesis, we used an RSA in which patterns of activation were compared with Pearson correlation coefficients as a similarity metric. This analysis was applied both across the whole brain using a searchlight and also in a series of four a priori defined ROIs. We chose RSA because it was shown that pattern-correlation linear classifiers can perform just as well as non-linear classifiers and are often simpler to implement and interpret (Misaki, Kim, Bandettini, & Kriegeskorte, 2010).

1.5 Thesis overview

In the following set of experiments, we used MVPA to investigate how the brain’s visual system integrates object features to yield a unified percept of a complete object. In the first fMRI experiment, we investigated whether (and where) the brain represents unique representations of bound object features (Chapter 2), and whether these representations remained stable across changes in viewpoint (Chapter 3). We hypothesized that any observation of explicit conjunctive coding would be found in anterior VVS, extending into medial temporal regions. Results from this study provide the first direct evidence that information coding in PRC simultaneously discriminated between the precise conjunctions of features comprising an object and was also invariant to changes in the visual appearance of those conjunctions caused by shifting viewpoints. These coding principles distinguished the PRC from other functionally-defined
regions in the VVS that are more classically associated with perception (e.g., LOC, FFA, and PPA).

In a second fMRI study, we investigated whether extensive training on a visual search task (Chapter 4), influenced the way these feature conjunctions were represented in the brain (Chapter 5), and whether the training influenced the way these objects were processed across different viewpoints (Chapter 6). We hypothesised that as a result of visual training, conjunctive representations will no longer be represented robustly in the PRC, but rather, will shift to more posterior regions of the VVS. Results showed that participants’ search slopes decreased over the course of training, indicating that visual search became more efficient and suggesting that object features became unitized. Results from the subsequent fMRI study revealed a different activation profile following training, with activity focused on more posterior VVS regions (middle fusiform gyrus). An analysis of view-invariant conjunctive coding strengthened these findings—whereas for novel objects, representations were concentrated in the PRC, conjunctive representations for trained objects shifted to more posterior regions. Together, these findings have implications for current theoretical models of visual object learning.

Note to readers:

Chapters 2 and 3 in this thesis were adapted from the following article:


© Oxford University Press; the use of this article is by permission of the copyright holder.
Chapter 2: Neural representations of feature conjunctions
composing complex objects

2.1 Introduction

As discussed in the Introduction, how we make sense of what we see is a central issue in neuroscience. Perceiving even a single object requires an ability to integrate its features (e.g., shape, color) into a unified percept (the “binding problem”) and recognize it across different viewing angles, despite the drastic variability in appearance caused by shifting viewpoints (the “invariance problem”). The mechanism underlying this feat is one of the central unsolved puzzles in cognitive neuroscience.

Two main classes of models have been proposed. The first are hierarchical models, in which representations of low-level features are transformed into more complex and invariant representations as information flows through successive stages of the ventral visual stream (VVS), a series of anatomically-linked cortical fields originating in V1 and extending into the temporal lobe (Desimone & Ungerleider, 1989; Gross, 1992; Hubel & Wiesel, 1965; Riesenhuber & Poggio, 1999; Tanaka, 1996). These models assume explicit conjunctive coding of bound features: posterior VVS regions represent low-level features and anterior regions represent increasingly complex and invariant conjunctions of these simpler features. By contrast, an alternative possibility is a non-local binding mechanism, in which the perception of a unitized object does not necessitate explicit conjunctive coding of object features per se, but rather, the features are represented independently and bound by co-activation. Such a mechanism could include the synchronized activity of spatially-distributed neurons that represent the individual object features (Singer & Gray, 1995; Uhlhaas et al., 2009), or a separate brain region that
temporarily reactivates and dynamically links otherwise disparate feature representations (Devlin & Price, 2007; Eckhorn, 1999). Thus, an explicit conjunctive coding mechanism predicts that the neural representation for a whole object should be different from the sum of its parts, whereas a non-local binding mechanism predicts that the whole should not be different from the sum of the parts, because the unique conjunctive representations are never directly coded.

The neuroimaging method of multi-voxel pattern analysis (MVPA) offers promise for making more subtle distinctions between representational content than previously possible (J V Haxby et al., 2001; Kamitani & Tong, 2005). Here, we used a novel variant of MVPA to adjudicate between these two mechanisms. Specifically, we measured whether the representation of a whole object differed from the combined representations of its constituent features (i.e., explicit conjunctive coding), and whether any such conjunctive representation was view-invariant (see Chapter 3). We examined the patterns of neural activity evoked by three features distributed across two individually-presented objects during a 1-back task (Figure 2.1).
Figure 2.1 Stimuli and task. (A) Objects had one, two, or three attached features, and were shown from one of two possible viewpoints (pertinent to a separate analysis discussed in Chapter 3). For illustrative purposes, we schematize objects with letters: “A” corresponds to a one-featured object and “AB” corresponds to a two-featured object consisting of features A and B. We included two feature sets to ensure that our analysis did not yield results idiosyncratic to a particular feature set. (B) Participants completed a 1-back task in which they responded to a sequentially repeating object, regardless of viewpoint (targets in red). Objects were always presented in isolation.

Our critical contrast measured the additivity of patterns evoked by different conjunctions of features across object pairs: A+BC vs. B+AC vs. C+AB, where A, B, and C each represent an object comprised of a single feature, and AB, BC, and AC each represent an object comprised of conjunctions of those features (Figure 2.2 A,B). Importantly, in this “Conjunction Contrast”, the object pairs were identical at the feature level (all contained A, B, and C), but differed in their conjunction (AB vs. BC vs. AC), allowing a clean assessment of the representation pertaining to
the conjunction, over and above any information regarding the component features. This balanced design also ensured that mnemonic demands were matched across comparisons. A finding of equivalent additivity (i.e., if $A+BC = B+AC = C+AB$) would indicate that information pertaining to the specific conjunctions is *not* represented in the patterns of activity – consistent with a non-local binding mechanism in which the features comprising an object are bound by their co-activation. By contrast, if the pattern sums are not equivalent (i.e., if $A+BC \neq B+AC \neq C+AB$), then the neural code must be conjunctive, representing information about the specific conjunctions of features over and above information pertaining to the individual features themselves – consistent with an explicit conjunctive coding mechanism. Importantly, our novel MVPA linearity design avoided making unbalanced comparisons (e.g., $A+B+C$ vs. $ABC$) where the number of object features was confounded with the number of objects.
Figure 2.2 Experimental question and similarity contrast matrix. (A) Our Conjunction Contrast investigated whether neural patterns of activity demonstrated explicit conjunctive coding (i.e., was the whole different from the sum of its parts?). To this end, we measured patterns of activity (schematized by a hypothetical ROI consisting of 13 voxels) to each of the objects that were presented individually during the 1-back task. We then computed linear summations of these patterns for three different pairs of objects (i.e., A+BC, B+AC, and C+AB), which were matched in terms of their individual features (A, B, C), but different in terms of their conjunction (i.e., AB, BC, AC). (B) MVPA correlations within and between these conjunctions were summarized in a matrix structure (the full 144 x 144 matrix is shown in Figure 2.4). This contrast tested whether correlations of repetitions of the same conjunctions (dark blue) were more similar in their activation pattern compared to correlations of different conjunctions (light blue). As shown by the zoomed-in cell, each cell in the Conjunction Contrast is in fact an average of a 12x12 correlation matrix that computed correlations across the 4 experimental runs and the 3 data subdivisions. Gray cells in the zoomed-in depiction reflect hypothetical data for illustrative purposes only.

Indeed, an aspect of our design that should be emphasized is that during the task, participants viewed objects displayed in isolation. This is important, because presenting two
objects simultaneously (e.g. Macevoy and Epstein 2009) could potentially introduce a bias, particularly when attention is divided between them (Agam et al., 2010; Reddy, Kanwisher, & Vanrullen, 2009). Whereas objects were presented in isolation during the task, responses to single objects were then combined during analysis. On each side of every comparison, we combined across an equal number of objects (two), as there will be activity evoked by an object that does not scale with its number of features. So, for example, we rejected a simpler design in which A+B=AB was tested, as there are an unequal number of objects combined on the two sides of the comparison (2 vs. 1).

We hypothesized that any observation of explicit conjunctive coding would be found in anterior VVS, extending into anterior temporal regions. In particular, one candidate structure that has received intensified interest is the perirhinal cortex (PRC) – a medial temporal lobe (MTL) structure whose function is traditionally considered exclusive to long-term memory (Squire & Wixted, 2011), but has recently been proposed to sit at the apex of the VVS (Barense, Groen, et al., 2012; Murray, Bussey, & Saksida, 2007). Yet to our knowledge, there have been no direct investigations of explicit conjunctive coding in the PRC. Instead, most empirical attention has focussed on area TE in monkeys and the object-selective lateral occipital complex (LOC) in humans – structures posterior to PRC and traditionally thought to be the anterior pinnacle of the VVS (Denys et al., 2004; Grill-Spector et al., 2001; Kriegeskorte, Mur, Ruff, et al., 2008; Sawamura et al., 2005; Ungerleider & Haxby, 1994). For example, single-cell recording in monkey area TE showed evidence for conjunctive processing whereby responses to the whole object could not be predicted from the sum of the parts (Baker et al., 2002; Desimone et al., 1984; Gross, 2008), although these conjunctive responses might have arisen from sensitivity to new features created by abutting features (Sripati & Olson, 2010). Here, with our novel
experimental design we were able to directly measure explicit conjunctive coding of complex objects for the first time in humans. We used both a whole-brain approach and an ROI-based analysis that focused specifically on the PRC and functionally-defined anterior structures in the VVS continuum. Our results revealed that regions of the VVS, extending into the PRC, contained unique representations of bound object features, consistent with an explicit conjunctive coding mechanism predicted by hierarchical models of object recognition.

2.2 Materials and Methods

Participants
Twenty neurologically-normal right-handed participants gave written informed consent approved by the Baycrest Hospital Research Ethics Board and were paid $50 for their participation. Data from one participant was excluded due to excessive head movement (>10° rotation), leaving 19 participants (18–26 years old, mean=23.6 years, 12 female).

Experimental Stimuli and Design
Participants viewed novel 3-dimensional objects, created using Strata Design 3D CX 6. Each object was assembled from one of two feature sets and was composed of a main body with one, two, or three attached features (depicted as “A”, “B” and “C” in Figure 2.1A). There were seven possible combinations of features within a feature set (A, B, C, AB, AC, BC, ABC). Features were not mixed between sets. Each object was presented from one of two possible angles separated by a 70° rotation along a single axis: 25° to the right, and 45° to the left from central fixation. We ensured that all features were always visible between angle changes. There were 28 images in total, created from every unique combination of the experimental factors: 2 (feature
sets) x 2 (viewpoints) x 7 (possible combinations of features within a set). Figure 2.1A depicts the complete stimulus set.

**Tasks**

*Experimental task.* We administered four experimental scanning runs during which participants completed a 1-back task to encourage attention to each image. Participants were instructed to press a button with their right index finger whenever the same object appeared twice in succession, regardless of its viewpoint (behavioural results in Table 2.1). Feedback was presented following each button press (correct or incorrect) and at the end of each run (proportion of correct responses during that run). Trials on which a response was made were not included in the analysis.

**Table 2.1.** Behavioural results of the 1-back task completed during scanning.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected recognition (%)</td>
<td>89.85 (9.11)</td>
</tr>
<tr>
<td>Hits (%)</td>
<td>90.55</td>
</tr>
<tr>
<td>False alarms (%)</td>
<td>0.70</td>
</tr>
<tr>
<td>Correct rejections (%)</td>
<td>98.94</td>
</tr>
<tr>
<td>Misses (%)</td>
<td>9.45</td>
</tr>
<tr>
<td>Reaction times (ms)</td>
<td>723.94 (184.07)</td>
</tr>
<tr>
<td>Sensitivity (d’)</td>
<td>$d_L = 7.52 (1.40)$</td>
</tr>
</tbody>
</table>

Mean corrected recognition (proportion of hits – proportion of false alarms), and percentage of hits, false alarms, correct rejections and misses. Mean reaction times (ms) were calculated for correct trials only. Sensitivity (d’) was calculated by applying signal detection theory to logistic distributions: $dL=\ln([H(1-FA)] / [(1-H)FA])$, where ln = natural log (Snodgrass & Corwin, 1988). Standard deviations shown in parentheses.

Objects were presented centrally on the screen and had a visual angle of $5.1^\circ \times 5.3^\circ$, which would likely encompass the receptive fields of PRC ($\sim 12^\circ$) (Nakamura, Matsumoto, Mikami, & Kubota, 1994), V4 (4-6° at an eccentricity of 5.5°) (Kastner et al., 2001), LOC (4-8°) (Dumoulin & Wandell, 2008), FFA and PPA (likely greater than > 6°) (Desimone & Duncan,
1995; Kastner et al., 2001; Kornblith, Cheng, Ohayon, & Tsao, 2013). The visual angle of the individual object features was approximately $2.1^\circ \times 2.2^\circ$, which would likely encompass the receptive fields of more posterior regions in the VVS (2-4$^\circ$ in V2) (Kastner et al., 2001). Each image was displayed for 1s with a 2s inter-stimulus interval. Each run lasted 11min 30s, and for every 42s of task time there was an 8s break (to allow BOLD signal to reach baseline) during which a fixation cross appeared on the screen. Each run comprised 6 blocks of 28 trials, which were presented in a different order to each participant. The 14 images composing each feature set were randomly presented twice within each block. Across consecutive blocks, the feature sets alternated (3 blocks per feature set per run). Each block contained between one and four target objects (i.e., sequential repeats), such that the overall chance that an object was a target was 10%. In total, each image was presented 24 times (6 times within each run). Prior to scanning, each participant performed a 5-minute practice of 60 trials.

**Localizer Task.** After the four experimental runs, an independent functional localizer was administered to define participant-specific ROIs (LOC, FFA and PPA, described below). Participants viewed scenes, faces, objects, and scrambled objects in separate 15s blocks (there was no overlap between the images in the experimental task above and the localizer task). Within each block, 20 images were presented for 300ms each with a 450ms ISI. There were four groups of 12 blocks, with each group separated by a 15s fixation-only block. Within each group, three scene, face, object, and scrambled-object blocks were presented (order of block type was counterbalanced across groups). To encourage attention to each image, participants were instructed to press a button with their right index finger whenever the same image appeared twice in succession. Presentation of images within blocks was pseudo-random: immediate repeats occurred between 0-2 times per block.
**Memory Task.** After participants completed the scanning session, they were given a surprise memory task in which they were shown a sequence of individually-presented objects and indicated whether they had seen an object before in the scanner (Figure 2.3). Twenty-six objects were shown in total, and to prevent source confusion, each object was presented only once from only one viewpoint. Of the 26 objects, 14 were “old” objects (all of the objects that were used in the experiment – half shown from viewpoint 1 and half shown from viewpoint 2, counterbalanced across participants), and 12 “new” objects, which were recombinations of features from the two feature sets. For example, as shown below, a new object could comprise two features from feature set 1 and one feature from feature set 2. The images were shown in a random order and the task was self-paced. Discrimination between familiar and novel configurations was significantly greater than chance (mean proportion correct = 0.94, t(18) = 21.27, p < 0.001). Participants’ excellent memory for the objects suggests that they had bound the features of a given object in a way that allowed them to subsequently distinguish it from novel objects that were created by recombining these features – indicating that the features were bound not only for the 1-back task in the scanner, but that this binding also transferred into longer-term memory.
Figure 2.3 Post-scan memory task (A) Following scanning, participants were administered a memory task in which they determined whether a series of objects were shown during scanning. Objects were either “old” (all of the objects that were used in the experiment) or “new” (recombinations of features from the two feature sets). (B) Proportion of hits, false alarms, correct rejections and misses. Error bars represent s.e.m.

fMRI Data Acquisition

Scanning was performed using a 3.0-T Siemens MAGNETOM Trio MRI scanner at the Rotman Research Institute at Baycrest Hospital using a 32-channel receiver head coil. Each scanning session began with the acquisition of a whole-brain high-resolution MP-RAGE T1-weighted structural image (repetition time=2s, echo time=2.63ms, flip angle=9°, field of view=25.6cm², 160 oblique axial slices, 192×256 matrix, slice thickness=1mm). During each of four functional scanning runs, a total of 389 T2*-weighted echo-planar images were acquired using a two-shot gradient echo sequence (200×200mm field of view with a 64×64 matrix size), resulting in an in-plane resolution of 3.1×3.1mm for each of 40 2-mm axial slices that were acquired along the axis of the hippocampus. The interslice gap was 0.5-mm; repetition time=2s; echo time=30msec; flip angle=78°).

Multi-Voxel Pattern Analysis (MVPA)

Functional images were preprocessed and analyzed using SPM8 (www.fil.ion.ucl.ac.uk/spm) and a custom-made, modular toolbox implemented in an automatic analysis pipeline system (https://github.com/rhodricusack/automaticanalysis/wiki). Prior to MVPA the data were preprocessed, which included realignment of the data to the first functional scan of each run (after five dummy scans were discarded to allow for signal equilibrium), slice-timing correction, coregistration of functional and structural images, nonlinear normalization to the Montreal Neurological Institute (MNI) template brain, and segmentation of gray and white matter. Data were high-pass filtered with a 128s cut-off. The data were then “denoised” by deriving regressors
from voxels unrelated to the experimental paradigm and entering these regressors in a general linear model (GLM) analysis of the data, using the GLMdenoise toolbox for Matlab (Kay, Rokem, Winawer, Dougherty, & Wandell, 2013). Briefly, this procedure includes taking as input a design matrix (specified by the onsets for each stimulus regardless of its condition) and an fMRI time-series, and returns as output an estimate of the hemodynamic response function (HRF) and BOLD response amplitudes (beta weights). (It is important to emphasize that the design matrix did not include the experimental conditions upon which our contrasts relied; these conditions were specified only after denoising the data.) Next, a fitting procedure selected voxels that are unrelated to the experiment (cross-validated $R^2$ less than 0%), and a principal components analysis (PCA) was performed on the time-series of these voxels to derive noise regressors. A cross-validation procedure then determined the number of regressors that were entered into the model (Kay et al., 2013).

We specified the onsets for each individual object (i.e., A, B, C, AB, BC, AC) for each of the two feature sets and two viewpoints. Our model then created a single regressor for each of the three different pairs of objects (i.e., A+BC, B+AC, and C+AB). This was done separately for each of the two feature sets and two viewpoints. For example, events corresponding to the singly-presented ‘A’ object from feature set 1 and viewpoint 2 and events corresponding to the singly-presented ‘BC’ object from feature set 1 and viewpoint 2 were combined to create the single regressor for ‘A+BC’ from feature set 1 and viewpoint 2. More specifically, within each run, the voxel-wise data of each object pair were split into 3 subdivisions that were each comprised of every third trial of a given image (following Zeithamova et al. 2012) (Figure 2.2B, zoomed-in cell). The pattern similarity of each condition in each subdivision was compared to that of each condition in every other subdivision. We designed the subdivisions so that our
comparisons were relatively equidistant in time. For example, the 1\textsuperscript{st} subdivision for the A + BC regressor included $A_{1\text{st\ presentation}} + BC_{1\text{st\ presentation}} + A_{4\text{th\ presentation}} + BC_{4\text{th\ presentation}}$; the 2\textsuperscript{nd} subdivision included $A_{2\text{nd\ presentation}} + BC_{2\text{nd\ presentation}} + A_{5\text{th\ presentation}} + BC_{5\text{th\ presentation}}$, etc. This resulted in 36 regressors of interest per run [2(feature sets) x 2(viewpoints) x 3(conjunctions) x 3(subdivisions)]. We also modeled 8 regressors of no interest for each run: trials of three-featured objects (ABC), trials in which participants responded with a button press on the 1-back task, and six realignment parameters to correct for motion. Events were modeled with a delta (stick) function corresponding to the stimulus presentation onset convolved with the canonical hemodynamic response function as defined by SPM8. This resulted in parameter estimates ($\beta$) indexing the magnitude of response for each regressor. Multi-voxel patterns associated with each regressor were then Pearson-correlated. Thus, each cell in our planned contrast matrices was comprised of a 12x12 correlation matrix that computed correlations within and across all runs and data subdivisions (Figure 2.2B, zoomed-in cell; see also Figure 2.4 for the full data matrix). This process was repeated for each cell in the contrast matrix, and these correlation values were then averaged and condensed to yield the 12x12 contrast matrix (similar to Linke et al. 2011). We then subjected these condensed correlation matrices to our planned contrasts (Figure 2.2B). In addition to an analysis that computed correlations both across and within runs, we also conducted an analysis in which we ignored within-run correlations and computed correlations across runs only. Results from this across-run only analysis are shown in Appendix 1. In brief, this analysis revealed the same pattern of results as the analysis that computed correlations both across and within runs.
Figure 2.4 The full data matrix and data analysis structure. (A) The entire data matrix for the PRC ROI, averaged across all participants (all ROIs had the same structure). This 144 x 144 matrix summarizes 2(feature sets) x 2(viewpoints) x 3(conjunctions) x 3(subdivisions) x 4(Runs). The identity line (i.e., main diagonal) included a series of ones (perfect correlation) and was removed from the analysis to avoid a bias that would yield a positive conjunction contrast. The black squares highlight a series of 12x12 matrices (e.g., M1, M2…M144), each of which includes the correlations of our 2(feature sets) x 2(viewpoints) x 3(conjunctions) for a given run and subdivision. (B) The matrix depicted in the Conjunction Contrasts (Figure 2.2B) was constructed by condensing the full 144x144 data matrix into a 12 x12 matrix that was the average of all correlations across all experimental runs and data subdivisions. That is, we calculated the average of all 144 “M-labelled” matrices (e.g., M1, M2…M144) comprising the full data matrix shown in Figure 2.4A. (C) The Conjunction Contrast was an average of all cells from the same feature set and same viewpoint. That is, we calculated the average of the four “C-labelled” matrices (i.e., C1…C4). The main diagonal contained correlations of the same type of conjunction, whereas the off-diagonal cells contained correlations of different types of conjunctions. We also conducted all analyses using correlations computed across runs only (i.e., removing within-run correlations), and found the same pattern of results (Appendix 1).
**Searchlight analysis.** A spherical ROI (10 mm radius, restricted to gray matter voxels and including at least 30 voxels) was moved across the entire acquisition volume, and for each ROI, voxel-wise, unsmoothed β-Values were extracted separately for each regressor (Kriegeskorte et al., 2006). The voxel-wise data (i.e., regressors of interest) were then Pearson-correlated within and across runs, and condensed into a 12×12 correlation matrix (see Figure 2.2B). Predefined similarity contrasts containing our predictions regarding the relative magnitude of pattern correlations within and between conjunction types specified which matrix elements were then subjected to a two-sample t-test. This analysis was performed on a single-subject level, and a group statistic was then calculated from the average results, indicating whether the ROI under investigation coded information according to the similarity matrix. Information maps were created for each subject by mapping the t-statistic back to the central voxel of each corresponding ROI. These single-subject t-maps were then smoothed with a 12mm FWHM Gaussian kernel to compensate for anatomical variability across participants. The resulting contrast images were then subjected to a group analysis that compared the mean parameter-estimate difference across participants to zero (i.e., a one-sample t-test relative to zero). Results shown in Figure 2.6A are superimposed on the single-subject MNI brain template.

**ROI analysis.** We investigated four ROIs defined a priori. Three were functionally-defined regions well-established as part of the VVS: lateral occipital complex (LOC), fusiform face area (FFA), and the parahippocampal place area (PPA). The fourth ROI was the PRC, which was defined by an anatomical probability map created by Devlin and Price (Devlin & Price, 2007). We included areas which had at least a 30% or more probability of being the PRC, as done previously (Barense, Henson, & Graham, 2011). For our functional localizer, we used identical stimuli to those employed in Watson et al. (2012). We defined the LOC as the region that was
located along the lateral extent of the occipital lobe and responded more strongly to objects compared with scrambled objects (p < 0.001, uncorrected) (Malach et al., 1995). We defined the FFA as the set of contiguous voxels in the mid-fusiform gyrus that showed significantly higher responses to faces compared to objects (p < 0.001, uncorrected) (Liu, Harris, & Kanwisher, 2010), and the PPA as the set of contiguous voxels in the parahippocampal gyrus that responded significantly more to scenes than to objects (p < 0.001, uncorrected) (Reddy & Kanwisher, 2007). These regions were defined separately for each participant by a 10mm radius sphere centered around the peak voxel in each hemisphere from each contrast, using the MarsBar toolbox for SPM8 (http://marsbar.sourceforge.net/). All ROIs were bilateral, except for three participants in whom the left FFA could not be localized, and another participant in whom the right LOC could not be localized (Appendix 2 displays the peak ROI coordinates for each ROI for each participant). The ROI MVPA was conducted in an identical manner to the searchlight analysis; voxel-wise data were Pearson-correlated and condensed into a 12×12 correlation matrix, except that here each ROI was treated as a single region (i.e., no searchlights were moved within an ROI). Before applying our contrasts of interest, we ensured that these correlation values were normally distributed (Jarque-Bera test; P > 0.05). We then applied our Conjunction and Viewpoint contrasts within each of the four ROIs and obtained, for each participant and each contrast, a t-value reflecting the strength of the difference between our correlations of interest (Figure 2.2 B). From these t-values, we calculated standard r-effect sizes that allowed us to compare the magnitude of effects across the ROIs (Rosenthal, 1994) (Figure 2.6B). Specifically, we transformed the r-effect sizes to Fisher’s z-scores (as they have better distribution characteristics than correlations; e.g. Mullen, 1989). We then conducted t-tests on the z-scores effect sizes obtained for each region, which provided a measure of the statistical significance.
between our cells of interest in each of our two contrast matrices (i.e., dark and light colored cells). We then compared the z-scores in each ROI to zero using Bonferroni-corrected one-sample t-tests, and conducted for each of the two contrasts paired-samples t-tests to compare the effect sizes observed in the PRC to the three more posterior ROIs (Figure 2.6B).

**Control analysis to test for sub-additivity of the BOLD signal**

To test the possibility that our results (Figure 2.6) were driven by signal saturation due to non-linearities in neuro-vascular or vascular-MR coupling, we conducted a univariate ANOVA to test whether there were differences in the overall signal evoked by different single features (Figure 2.5). A standard univariate processing pipeline was then followed, comprising the preprocessing steps described for the MVPA, but also smoothing of the imaging data with a 12 mm FWHM Gaussian kernel. We then conducted first-level statistical analyses. Within each run, there were 12 regressors of interest [2 (feature sets) x 2 (viewpoints) x 3 objects (A, B, and C)] and 8 regressors of no interest corresponding to trials of three-featured objects (ABC), trials in which participants responded with a button press on the 1-back task, and six realignment parameters to correct for motion. Within each regressor, events were modeled by convolving a delta (stick) function corresponding to the stimulus presentation onset with the canonical hemodynamic response function as defined by SPM8. Second-level group analyses were conducted separately for each of the two feature sets by entering the parameter estimates for the 6 single-featured objects (i.e., A, B, and C from each viewpoint) of each subject into a single General Linear Model (GLM), which treated participants as a random effect. This analysis was conducted using a factorial repeated-measures ANOVA, in which a model was constructed for the main effect of condition (i.e., the 6 single features). Within this model, an F-contrast was computed to test for
areas that showed a significant effect of feature type for each feature set. Statistical parametric maps (SPMs) of the resulting F-statistic were thresholded at $p < 0.001$, uncorrected.

**Figure 2.5** Control analysis to test for sub-additivity of the BOLD signal. (A) Saturation of the BOLD signal could produce non-linearities that could be misinterpreted as a conjunction – when in fact the coding was linear – if there were stronger activation for some features more than others (e.g., illustrated in the above hypothetical scenario for feature B). In this instance, adding another feature (e.g., A, C) to the feature that already produces strong activity (B) could cause the BOLD signal to saturate (AB, BC – dotted line). The BOLD response to the feature pairs would be a non-linear addition of that to the component features – not because of a conjunctive neural representation, but rather due to sub-additivity of the BOLD signal. (B) Regions identified by our univariate ANOVA as showing a difference in terms of overall signal evoked by the different single features in either feature set 1 (warm colors) or feature set 2 (cool colors). This analysis revealed activity in posterior visual regions, but not in PRC (see Appendix 3). Searchlight MVPA results from the (C) Conjunction Contrast (all regions from Figure 2.6A shown here in blue) and (D) Viewpoint Contrast (see Chapter 3; all regions from Figure 3.3A shown in blue), displayed with regions from the combined results of the univariate ANOVA for feature sets 1 and 2 (shown in purple). Overlapping regions between the two analyses are shown in orange.
2.3 Results and Discussion

Our primary analysis involved a planned comparison, a *Conjunction Contrast*, which determined whether the neural patterns of activity demonstrated explicit conjunctive coding (i.e., whether activity patterns represented information specific to the conjunction of features comprising an object, over and above information regarding the features themselves) (Figure 2.2). For this planned comparison, we performed two independent planned analyses – a *searchlight analysis* to investigate the activity across the whole brain and an *ROI analysis* to investigate activation in specific VVS ROIs. The Conjunction Contrast was applied to a correlation matrix that included all possible correlations within and across the different conjunctions (Figure 2.2 B). The novelty of this design ensured that our comparisons were matched in terms of the number of features that needed to be bound. That is, our comparison terms (e.g., A+BC vs B+AC) included both a combination of a single-featured object and a two-featured object, and thus, binding and memory requirements were matched – what differed was the underlying representation for the conjunctions themselves.
Figure 2.6 Regions demonstrating explicit conjunctive coding. (A) Regions where the representation for the conjunction was different from the sum of its parts (p < 0.05, whole-brain FWE-corrected; results at p < 0.001, uncorrected shown in Appendix 4). Broadly speaking, these regions included the PRC and earlier VVS regions (with the local maxima in V4). (B) The strength of conjunctive coding in the LOC, FFA, PPA and PRC ROIs (shown at bottom are ROIs from a representative participant superimposed on that individual’s structural scan). Error bars indicate s.e.m.; *p < 0.05 for comparisons across ROIs (paired-samples t-tests indicated by brackets) and comparisons relative to zero within an ROI (one-sample t-tests indicated by an asterisk above a bar).

For the Conjunction Contrast, a whole-brain searchlight MVPA (Kriegeskorte et al., 2006) revealed conjunctive coding in the VVS, with a global maxima in V4 (Rottschy et al.,
and activity that extended laterally into the LOC and posteriorly into V3 and V1 (peak x,y,z = 32,-80,-4, Z-value = 5.93), as well as conjunctive coding that extended anteriorly to the right PRC (Figure 2.6A; peak x,y,z = 30,-6,-30, Z-value = 5.47) (all results reported are whole-brain FWE-corrected at p < 0.05; Appendix 5 summarizes all regions). We next performed an ROI-based MVPA that applied the same contrast matrices and methods used for the searchlight analysis, except was focused only on the PRC and three functionally-defined regions (LOC, FFA, and PPA) that are posterior to the PRC and well-established as part of the VVS. This analysis allowed direct comparison of conjunctive coding strength across regions (Figure 2.6B). The Conjunction Contrast in this ROI MVPA revealed conjunctive coding in PRC (t(18) = 3.89, p < 0.01, r_{effect size} = 0.24) and the LOC (t(18) = 3.85, p < 0.01, r_{effect size} = 0.10), but not in FFA (t(18) = 1.80, p = 0.35 , r_{effect size} = 0.06) or PPA (t(18) = 2.66, p = 0.06, r_{effect size} = 0.10) (all one-sample t-tests Bonferroni-corrected). Comparisons across ROIs demonstrated stronger conjunctive coding in the PRC relative to each of the three more posterior VVS ROIs (p’s < 0.05). Thus, consistent with recent hierarchical models proposing explicit conjunctive coding in regions not traditionally associated with the VVS (Barense, Groen, et al., 2012; Murray et al., 2007), we found that PRC representations explicitly coded information regarding the object’s conjunction, over and above its individual features.

In addition to the PRC, we also observed conjunctive coding in V4 as well as in LOC, indicating that the conjunctive coding mechanism is not selective to PRC. Indeed, there is evidence to suggest that V4 and LOC are important for representing feature conjunctions. For example, recent studies showed that learning-induced performance changes on a conjunctive visual search were correlated with increasing activity in V4 and LOC (Frank et al., 2014), and that these regions are important for conjoined processing of color and spatial frequency
(Pollmann, Zinke, Baumgartner, Geringswald, & Hanke, 2014). In support of a causal role for both LOC and PRC in representing feature conjunctions, patients with selective LOC damage (Behrmann & Williams, 2007; Konen, Behrmann, Nishimura, & Kastner, 2011) and those with PRC damage (Barense et al., 2005; Barense, Groen, et al., 2012) were impaired on tasks that tax integrating object features into a cohesive unit.

We also investigated representations for the simpler one-featured objects. To this end, we conducted an MVPA that tested whether correlations of repetitions of the same single features (A, B, C) were more similar in their activation pattern compared to correlations of different single features (Figure 2.7A). Whole-brain searchlight results showed that predominantly posterior VVS regions were implicated, and an ROI MVPA indicated that LOC demonstrated stronger single-feature coding than PRC, FFA, or PPA (Figure 2.7D). However, we still observed above baseline activity in PRC. How can we reconcile this with the claim of conjunctive coding in PRC? Our null hypothesis for the primary comparisons of interest (i.e., A+BC vs. B+AC vs. C+AB) was that each feature will activate a pattern regardless of its context (e.g., in this case, the context could be the other features that are present on the object). Our evidence that the patterns associated with A+BC ≠ B+AC ≠ C+AB allows us to reject the null hypothesis and show that this “feature context” does matter to certain parts of the brain, including to the PRC. This coding of feature context will likely operate even for the single feature objects (i.e., A, B, C) that composed the single feature contrast. For example, for stimulus A, PRC may be encoding the relationship between Feature A and the object’s blue body. That is, just as PRC represents the AB conjunction, it may also represent the A-body conjunction. Nonetheless, we did observe clear differences in conjunctive vs. single-feature coding across our ROIs, with PRC demonstrating stronger conjunctive coding than LOC, FFA, and PPA (Figure
2.6B), and LOC demonstrating stronger single-feature coding than PRC, FFA, or PPA (Figure 2.7D).

**Figure 2.7** Regions demonstrating single-feature coding. (A) MVPA contrast matrix investigating representations for single-features. We investigated whether correlations of repetitions of the same single features (A, B, C; dark blue) were more similar in their activation pattern compared to correlations of different single features (light blue). (B) Searchlight MVPA results demonstrating regions that coded the single-feature objects, superimposed on the single subject MNI brain template (p < 0.05, whole-brain FWE-corrected). (C) Table displaying the
searchlight MVPA results revealing regions that coded the single-feature objects (whole-brain FWE-corrected, p < 0.05). (D) Results of the ROI MVPA for the LOC, FFA, PPA and PRC. Although all ROIs showed significant coding of the single features (all t’s(18) > 2.97; all p’s < 0.05) (all tests Bonferroni-corrected), this coding was significantly stronger in the LOC relative to each of the three anterior ROIs (all p’s < 0.05).

Although the current study lacks the temporal resolution to address this directly, one possibility is that this more posterior activity may also reflect higher-level feedback, such as from PRC. Indeed, bidirectional interactions exist throughout the VVS (Coutanche & Thompson-Schill, 2014; Hochstein & Ahissar, 2002; Seymour, Clifford, Logothetis, & Bartels, 2009), and previous work has suggested that feedback from the PRC modulates familiarity responses to object parts in V2 (Barense et al. 2011; Peterson et al. 2012). An alternative possibility is that the PRC activity reflects a feedforward cascade from structures such as the LOC (Bullier, 2001; Lamme & Roelfsema, 2000).

Finally, it is important to rule out the possibility that the results we obtained were driven by signal saturation due to non-linearities in neurovascular or vascular-MR coupling (Figure 2.5). Such signal saturation could produce non-linearities that could be misinterpreted as a conjunction, when in fact the coding was linear, if there were stronger activation for some features more than others (Figure 2.5A). To evaluate whether our data fell within this regime, we conducted a univariate ANOVA to test whether there were differences in the overall signal evoked by different features when these features were presented alone. These analyses revealed activity predominately in early visual areas that were largely non-overlapping with the results from our critical Conjunction and Viewpoint Contrasts (Figure 2.5C,D; Appendix 3). Importantly, there were no significant differences between the basic features in terms of overall activity in the PRC, or in the LOC typically observed in our participants (even at a liberal
uncorrected threshold of \( p < 0.001 \)). Thus, there was no evidence to suggest our observation of conjunctive coding is driven by spurious BOLD signal saturation.

In conclusion, this chapter investigated the neural mechanism that supports the integration of object features into a whole that is different from the sum of its parts. Results showed that patterns of activity throughout the ventral visual stream (VVS), extending anteriorly into the perirhinal cortex (PRC), discriminated between the precise conjunctions of features comprising a complex object. The next chapter investigated whether these activity patterns remained stable across large changes in viewpoint.
Chapter 3: Neural representations of feature conjunctions composing complex objects: View Invariance

3.1 Introduction

As stated in the General Introduction, an important potential benefit of explicit conjunctive coding demonstrated in Chapter 2 is to provide stability of representation across changes in viewpoints, and invariance to the manifestation of individual object features (Biederman, 1987). The extent to which the human visual system supports view-invariant versus view-dependent representations of objects is unresolved (Biederman & Bar, 1999; Peissig & Tarr, 2007). Much research in this area has focused on VVS regions posterior to the MTL (Andresen et al., 2009; Vuilleumier et al., 2002) but some work has indicated that for very complex stimuli, structures in the MTL may be central to view-invariance (Barense et al., 2010; Quiroga et al., 2005). Despite this, to our knowledge, no study has directly probed how the specific conjunctions comprising complex objects are neurally represented across viewpoints to support object recognition in different viewing conditions. That is, as the representations for objects become increasingly precise and dependent on the specific combinations of features comprising them, can they also become more invariant to the large changes in visual appearance caused by shifting viewpoints?

To investigate this important question, we adapted the analyses and methods described in Chapter 2 to determine if the conjunctions were view-invariant (Figure 3.1). The same objects described in Chapter 2 were employed in the present chapter. These objects were presented at 70° rotations (Figure 2.1), a manipulation that caused a more drastic visual change than changing the identity of the object itself (Figure 3.2). In a “Viewpoint Contrast”, we measured whether the conjunctive representations of the object features changed when they were presented from these
different viewpoints. To assess view-invariance, our Viewpoint Contrast compared the additivity of patterns as we did in Chapter 2 (i.e., A+BC vs. B+AC vs. C+AB), except that the comparisons were made across the two viewpoints (Figure 3.1).

**Figure 3.1.** Experimental question and similarity contrast matrix. (A) Our “Viewpoint Contrast” investigated whether the conjunctive representations were view-invariant. As with the Conjunction Contrast (Figure 2.2), we measured the patterns of activity evoked by individually-presented objects and computed linear summations of these patterns for each object pair. We then tested whether these patterns of activity were sensitive to the viewpoint at which the conjunction was presented. (B) MVPA correlations within and across the same conjunctions shown from different viewpoints were summarized in a matrix structure. This contrast tested whether correlations of repetitions of the same conjunctions presented from different viewpoints (dark orange), were more similar in their activation pattern compared to correlations between different conjunctions presented from different viewpoints (light orange). Note that the gray cells in the zoomed-in depiction reflect hypothetical data for illustrative purposes only.
3.2 Materials and Methods

Participants

Participants were the same participants as in Chapter 2.

Experimental Stimuli and Design

The experimental stimuli and design were identical to that of Chapter 2. To recapitulate, each object was presented from one of two possible angles separated by a 70° rotation along a single axis: 25° to the right, and 45° to the left from central fixation. We ensured that all features were always visible between angle changes. There were 28 images in total, created from every unique combination of the experimental factors: 2 (feature sets) x 2 (viewpoints) x 7 (possible combinations of features within a set). Figure 2.1A depicts the complete stimulus set.

Image similarity analysis. To ensure that our viewpoint manipulation in fact caused a substantial change in the visual appearance of the objects, we conducted an analysis to determine how similar two images are on the most basic of levels – that is, how different the images would appear to the retina. Specifically, we performed an analysis in which we measured the basic visual similarity between the objects in our stimulus set — we calculated the root-mean-square difference (RMSD) between each of the 24 objects (all one-featured and two-features objects) and compared this value to every other object using the following function:

$$\sqrt{\frac{\sum_{i=1}^{n} (\text{Image}_{1i} - \text{Image}_{2i})^2}{n}}$$

where i is a pixel position in the image and n is the total number of pixels in the image. Thus, this function compares all of the pixels in one image with the corresponding pixels in a second
image and yields a value that indicates the similarity between two images, ranging from 0 (if the two images are identical) to 1 (if the two images are completely different) (Figure 3.2A). If our viewpoint shifts were insignificant (e.g., 1° difference), any observation of view-invariance for this small visual change would not be particularly meaningful. However, if we could show that our shift in viewpoint caused a visual change that was as significant as a change in the identity of the object itself, a demonstration of view-invariance would be much more compelling. To this end, we calculated a contrast matrix (Figure 3.2B) that compared the RMSD values of the same objects from different viewpoints (dark orange) versus RMSD values of different object features from the same viewpoint (light orange). Different objects from the same viewpoint were compared only if they were matched for the number of features. A t-test revealed that there was a significant difference between RMSD values of the same features shown from different viewpoints ($M = 0.16$, $SD = 0.03$) compared to different features shown from the same viewpoint ($M = 0.11$, $SD = 0.02$; $t(34) = 6.11$, $p < 0.001$) revealing that our viewpoint manipulation caused a change in visual appearance that was more drastic than maintaining the same viewpoint but changing the identity of the features altogether. Next, we conducted an RMSD analysis that was very similar to the Viewpoint Contrast (Figure 3.1B) used in our MVPA. Here, we compared the RMSD values of a change in viewpoint (dark orange) to a change in both viewpoint and feature type (light orange) (Figure 3.2C). A t-test revealed that there was no significant difference between RMSD values of the same features shown from different viewpoints ($M = 0.16$, $SD = 0.03$) compared to different features shown from different viewpoints ($M = 0.17$, $SD = 0.02$; $t(70) = 1.28$, $p = 0.20$). This indicates that the shift in viewpoint within a feature set caused a change in visual appearance that was as drastic as
changing the features altogether. That is, our viewpoint manipulation was not trivial and caused a substantial visual change of the appearance of the objects.
Figure 3.2 An analysis of visual differences across objects using a root-mean-square difference (RMSD) measure on the RGB values of all images used in the experiment indicated that the viewpoint manipulation produced a substantial change in the visual appearance of the object that was as drastic as switching the identity of the object altogether. (A) Raw RMSD values for all images used in the experiment. RMSD values range from 0 (if the two images were identical) to 1 (if the two images were completely different, see bar on the right). (B) A contrast matrix that compared the RMSD values of the same objects from different viewpoints (dark orange) versus RMSD values of different objects from the same viewpoint (light orange). Objects from the same viewpoint were compared only if they were matched for the number of features. (C) Similar to the Viewpoint Contrast (Figure 3.1B) used in our MVPA, we constructed a contrast matrix that compared the RMSD values of a change in viewpoint (dark orange) to a change in both viewpoint and feature type (light orange).
It is worth noting that because RMSD values constitute a difference score that reflects the visual differences between one object with respect to another object, we could not calculate RMSD difference scores to the pairs of objects as we did in the MVPA (e.g., A + BC vs. B + AC). Put differently, in the MVPA we could measure the patterns of activity evoked by the presentation of a single object (e.g., A), add this pattern of activity to that evoked by a different single object (e.g., BC), and then compare across different pattern sums (e.g., A + BC vs. B + AC). By contrast, because an RMSD value reflects the visual difference between two objects (rather than to an object on its own), we could not measure an RMSD value to object “A” and add that value to the RMSD value of object “BC”. As such, it was not possible to calculate RMSD values for the pattern sums.

**Experimental task.**

The experimental task was identical to that of Chapter 2. Participants completed a 1-back task to encourage attention to each image, and were instructed to press a button with their right index finger whenever the same object appeared twice in succession, regardless of its viewpoint. Feedback was presented following each button press and at the end of each run. Trials on which a response was made were not included in the analysis.

**3.3 Results and Discussion**

Our analysis involved a planned comparison, a Viewpoint Contrast, which determined whether the neural patterns of activity demonstrated view-invariant conjunctive coding (Figure 3.1). As in Chapter 2, here we performed two independent planned analyses – a searchlight analysis and an ROI analysis to investigate activation in specific VVS ROIs. The Viewpoint Contrast was
applied to a correlation matrix that included all possible correlations within and across the different conjunctions (Figure 3.1B).

At a stringent whole-brain FWE-corrected threshold of $p < 0.05$ our searchlight MVPA revealed limited activity throughout the brain (2 voxels, likely in the orbitofrontal cortex, Appendix 6A). However, at a more liberal threshold ($p < 0.001$ uncorrected), we observed view-invariant conjunctive coding in the VVS, with maxima in V4 (Rottschy et al., 2007); peak $x,y,z = 32,-70,-4$, $Z$-value = 4.07), lateral IT cortex (peak $x,y,z = -52,-28,-20$, $Z$-value = 4.15), as well as activity that extended anteriorly to the left PRC (peak $x,y,z = -36,-4,-26$, $Z$-value = 3.25) (Figure 3.3A, Appendix 6B). The ROI MVPA of the Viewpoint Contrast revealed view-invariance in the PRC ($t(18) = 3.04$, $p < 0.05$, $r_{\text{effect.size}} = 0.21$), but not the LOC, FFA or PPA ($t$’s $< 0.75$, p’s $> .99$, all $r_{\text{effect.size}} < 0.03$; Figure 3.3B) (all tests Bonferroni-corrected). A direct comparison across regions confirmed that this view-invariant conjunctive coding was stronger in the PRC compared to the three more posterior VVS ROIs (p’s $< 0.05$). Analyses computing correlations across runs only revealed the same pattern of results as our main findings (see Appendix 7), indicating that our results were not driven by within-run correlations alone.
Figure 3.3 Regions demonstrating explicit conjunctive coding across viewpoints. (A) Regions demonstrating view-invariant conjunctive representations (p < 0.001, uncorrected; no VVS regions survived whole-brain FWE correction). Broadly speaking, these regions included the PRC, as well as V4 and lateral IT cortex. (B) The strength of view-invariant conjunctive coding within and across the LOC, FFA, PPA and PRC ROIs (shown at bottom are ROIs from a representative participant superimposed on that individual’s structural scan). Error bars indicate s.e.m.; *p < 0.05 for comparisons across ROIs (paired-samples t-tests indicated by brackets) and comparisons relative to zero within an ROI (one-sample t-tests indicated by an asterisk above a bar).

Taken together, results from the Conjunction (Chapter 2) and Viewpoint (Chapter 3) contrasts provide the first direct evidence that information coding in PRC simultaneously discriminated between the precise conjunctions of features comprising an object, yet was also
invariant to the large changes in the visual appearance of those conjunctions caused by shifting viewpoints. These coding principles distinguished the PRC from other functionally-defined regions in the VVS that are more classically associated with perception (e.g., LOC, FFA, and PPA). It is important to note that our results were obtained in the context of a 1-back memory task, and thus, one might be cautious to interpret our results in the context of a perceptual role for the PRC. Indeed, we prefer to consider our results in terms of the representational role for any given brain region, rather than in terms of the cognitive process it supports – be it either memory or perception. To this end, our experimental design ensured that memory demands were matched in our comparisons of interest – all that differed was how the same three features were arranged to create an object. With this design, explicit memory demands were equated, allowing a clean assessment of the underlying object representation. That said, one might still argue that these object representations were called upon in the service of a 1-back memory task. However, a wealth of evidence suggests that PRC damage impairs performance on both perceptual (e.g., Bussey et al. 2002; Bartko et al. 2007; Lee and Rudebeck 2010; Barense, Ngo, et al. 2012; Barense et al. 2012) and explicit memory tasks (Barense et al., 2005; Bartko et al., 2010; Bussey et al., 2002; McTighe et al., 2010; Winters et al., 2004). The critical factor in eliciting these impairments was whether the task required the objects to be processed in terms of their conjunctions of features, rather than on the basis of single features alone. We argue that these seemingly disparate mnemonic and perceptual deficits can be accounted for by the fact that both the mnemonic and perceptual tasks recruited the conjunctive-level representations we have measured in the current study.

In conclusion, this study provides new evidence regarding the functional architecture of object perception, demonstrating neural representations that both integrated object features into a
whole that was different from the sum of its parts and were also invariant to large changes in viewpoint. To our knowledge, this constitutes the first direct functional evidence for explicit coding of complex feature conjunctions in the human PRC, suggesting that visual-object processing does not culminate in IT cortex as long believed (e.g., Lafer-Sousa and Conway 2013), but instead continues into MTL regions traditionally associated with memory. This is consistent with recent proposals that memory (MTL) and perceptual (VVS) systems are more integrated than previously appreciated (Barense, Groen, et al., 2012; Murray et al., 2007). Rather than postulating anatomically separate systems for memory and perception, these brain regions may be best understood in terms of the representation they support – and any given representation may be useful for many different aspects of cognition. In the case of the PRC, damage to these complex conjunctive representations would impair not only object perception, but would also cause disturbances in the recognition of objects and people that are critical components of amnesia.

The subsequent chapters investigated whether extensive training with the stimuli used in the current study changed the pattern of conjunctive representations across the VVS. Participants were scanned after completing a series of visual search training sessions — the next chapter describes this behavioural training programme.
Chapter 4: Experience modifies the conjunctive neural representations necessary for complex object perception: behavioural training

4.1 Introduction

Integrating complex object features into a unified percept often occurs in the real world not as a result of any conscious effort to combine several object features, but rather as a result of extensive exposure to the object. For example, a typical basketball has a specific size, color, and texture, but one can immediately recognize a basketball without having to combine deliberately its different features each time. In other words, experience plays an important role in forming feature conjunctions in our daily lives. In this study we were interested in investigating how the representations of the feature conjunctions (reported in the previous two chapters) would change following extensive training with the stimuli. We trained participants with the same set of objects used before and scanned them immediately following these sets of training sessions. The behavioural training programme is the focus of this chapter, and the subsequent fMRI study will be described in Chapters 5 and 6.

The main goal of the training sessions was to familiarize participants with our stimuli and to promote the unitization of object features through repeated exposure. Unitization refers to a perceptual learning mechanism that involves the construction of a cohesive unit from a complex configuration (Goldstone, 1998). By integrating parts into single wholes, unitization allows for a task, which originally would have required the detection of several parts, to be accomplished by detecting a single unit. Evidence for unitization originated from word perception; Laberge (1973) found that participants were faster at responding to actual letters compared to letter-like controls,
arguing that shape components of frequently encountered stimuli (i.e., the line segments that form actual letters) become processed as a single cohesive unit with practice. Smith and Haviland (1972) reported a similar finding when letters were treated as the individual features and words as the unitary objects — actual words were perceived more accurately than non-words. Evidence for perceptual unitization comes from configural perception — for instance, the finding that recognition for inverted faces is slower and more error-prone compared to upright faces (Tanaka & Farah, 1993; Valentine, 1988), and that inversion costs are associated with well-trained compared to untrained 3D objects (Gauthier & Tarr, 1997). These findings suggested that experience with an object promotes the development of a single, cohesive representation of its multiple parts.

One task that has been shown to be useful in promoting perceptual unitization is visual search (e.g. Shiffrin & Schneider, 1977; Czerwinski et al., 1992; Cousineau, Donkin, & Dumesnil, 2015). In a typical visual search task, participants search a display of items for the presence of a predefined target. Reaction times (RTs) can then be plotted as a function of the number of items in the search array (i.e., the display set size). These “RT x set size” plots are then used to calculate the search slope, which provides an index of search efficiency (Wolfe, 1998). In an efficient search, RT slopes are relatively flat – RTs remain at about the same level with increasing items in the search array; this is also referred to “parallel search” (as the visual system can examine all items at once) or “pop-out” (due to the target’s apparent leap from the display without effort from the observer) (Wolfe, Cave, & Franzel, 1989). On the other hand, in an inefficient search, RTs typically increase linearly with set size; this is commonly referred to as “serial search” because participants must inspect each item in order to locate the target (Wolfe, Cave, & Franzel, 1989).
Ann Treisman’s Feature Integration Theory (FIT) predicts that when performing visual search, basic features can be processed in ‘parallel’ through preattentive processing (efficient search), while a conjunction of two or more features will require attention and thus will be attended to in a serial manner (inefficient search) (Treisman & Gelade, 1980). Other models have built upon Treisman’s work, some showing that visual search efficiency depends also on the similarity between targets and non-targets (Duncan & Humphreys, 1989), and others arguing that efficient search involves combining information from both top-down and bottom-up processes to create a ranking of items in order of their attentional priority (Guided Search Model; Wolfe, Cave, & Franzel, 1989; Wolfe, 1994).

A second goal of the training sessions was to familiarize participants with the two viewpoints from which objects could be presented. As discussed in the General Introduction, there are two competing models which relate to the role experience plays in shaping view-invariant processing of complex objects. One theory suggests that view-invariant properties are extracted immediately, and therefore an extensive training period is not necessarily needed (Biederman & Bar, 1999; Biederman & Gerhardstein, 1993). In contrast, according to a second theory, view-invariance is gradually achieved through exposure to multiple views of the object, and hence, training sessions are beneficial (Hayward & Tarr, 2000; Tarr & Pinker, 1989; Tarr, 1995). Part of the data analysis in this experiment (as will be described below) examined trials in which target objects were presented from either the same or a different viewpoint, to examine whether training was beneficial for achieving view-invariance.

The design employed in the present study was inspired by Czerwinski et al. (1992). In this study, participants were trained for 50 days on a visual search task in which participants made target present/absent judgments on a set of stimuli (rectangles with different spokes
pointing inward (see Figure 4.1) in which no single feature defined a target. The logic behind this was that by using stimuli with a high degree of feature overlap, search efficiency will depend on the level of unitization that has occurred during the course of training. Results from their study showed that participants’ performance improved remarkably during the course of training, as evident by shallower search slopes with more days of training. These results provided evidence for the notion that the stimulus features became unitized with training through a process of perceptual learning.

**Figure 4.1** The stimuli used by Czerwinski et al. (1992). All stimuli shared a common rectangular frame and included three internal line segments. No line segment was shared across feature sets, and every stimulus within a set shared one line segment with another stimulus – forcing visual search to rely on a conjunction of at least two line segments.

Although we were reluctant to subject our participants to the two-month intensive training programme used by Czerwinski et al., we hoped to see a similar trend for improvement following 6 training sessions, using our more complex set of stimuli. Notably, even following ~50 days of training, the performance of participants in the Czerwinski et al. study on the conjunctive search was not indicative of parallel processing. That is, search slopes were still
relatively steep following unitization of features (i.e., unitization does not necessitate “pop-out”). Similarly, our hope and expectation in conducting these training sessions was not to see flat search slopes, but to see a more efficient search across training sessions (i.e., improvement in terms of reaction times and/or accuracy). We then planned to investigate the neural underpinnings of these training effects in a subsequent fMRI study – the topics of Chapters 5 and 6.

4.2 Materials and Methods

Participants
Participants were 18 young adults (19-26 years; M = 23.4, SD = 2.33; 11 females). All participants received monetary compensation for their participation. All participants had normal or corrected-to-normal vision. This study was approved by the Ethics Review Board at the University of Toronto.

Materials
Participants viewed novel 3D objects described previously (Chapter 2). Each object was assembled from one of two feature sets and was composed of a main body with 1, 2, or 3 attached features (depicted as “A,” “B,” and “C” in Figure 2.1A). There were 7 possible combinations of features within a feature set (A, B, C, AB, AC, BC, ABC), and each object was presented from one of two possible angles. 28 images served as the target set from which potential targets were selected. These images were created from every unique combination of the experimental factors: 2 (feature sets) × 2 (viewpoints) × 7 (possible combinations of features within a set). In addition, we created an recombin ed set in which the features from the two feature sets (feature set 1, hereafter termed A_1 B_1 C_1 and feature set 2, hereafter termed A_2 B_2 C_2)
were recombined: there were 6 possible two_featured mixtures (A_1B_2, A_1C_2, A_2B_1, A_2C_1, B_1C_2, B_2C_1) and 6 three_featured mixtures (A_1B_1C_2, A_1B_2C_1, A_2B_1C_1, A_2B_2C_1, A_2B_1C_2, A_1B_2C_2) from two possible viewpoints, resulting in (6 + 6) × 2 = 24 additional objects in the recombined set. Thus, in total there were 52 objects on which participants were trained (the target set plus the recombined set). All trained objects had a blue_coloured body. In order to examine whether any potential training effects were only task_related and independent of the stimuli used, we additionally created a set of novel objects, which was made from a yellow body and 2 novel feature sets. This novel stimulus set was shown only in the last half of the last testing session. As for the trained objects, we repeated the same permutations and recombination of features described above to create the novel set. Thus, overall this resulted in 52 additional objects in this novel stimulus set (see Figure 4.2 for sample stimuli).
Figure 4.2 Sample stimuli use in the experiment. (A) Sample objects from the training stimulus set. The target set consisted of the original 28 objects described in Chapter 2. The recombined set included in addition objects created from a recombination of features from the two feature sets. Together, the target set and recombined set composed the trained object set. (B) Sample objects from the novel stimulus set. These objects were only used on day 6 to examine whether any potential training effects were only task-related and independent of the stimuli used. Both training and novel stimulus sets included 28 objects as targets and an additional 24 recombined objects.

Design and Behavioural Procedure

Participants completed a series of six 1-hour visual search training sessions. All training sessions for all but one participant were completed within 10 days. Each trial began with a display of a black fixation cross for 1 second at the centre of the screen, followed by a display of a target object for 1 second, a blank screen for 1 second, and a search array comprised of 1-8 objects.
These objects remained on the screen until a response was made (Figure 4.3). On each trial, participants were asked to indicate whether the target object was present or absent by pressing one of two green coloured buttons on a laptop keyboard (“L” if the target was present and “A” if the target was absent). Participants were instructed to respond as quickly as possible on each trial, without sacrificing accuracy. One-second auditory feedback was provided for each trial in the form of a high/low tone for correct/incorrect response, respectively.

**Figure 4.3** Visual search task: examples of (A) a viewpoint-congruent target-present trial and (B) a viewpoint-incongruent target-present trial for a set size of 5.

On each trial, a target object was randomly selected from the training target set. Half the trials were ‘target present’ trials. A target was considered present if the same object appeared in the search array from either the same or a different viewpoint. That is, for target present trials, targets were presented from either the same viewpoint as the sample object (Figure 4.3A), or
from a different viewpoint (Figure 4.3B). This difference in trial types was later utilized to conduct an analysis that examined the effects of viewpoint-congruency on visual search performance. The remaining half of trials were ‘target absent’ trials, in which only distractor objects were presented in the search array. Each target object was associated with a set of possible distractors for that object. For one-featured targets, possible distractors were all other one-featured objects or objects with a conjunction of the target feature with one or two additional features. For two- and three-featured objects, possible distractors were all objects that included at least one of the features of the target object. Objects in the search array were presented at random locations on an imaginary circle that included 10 possible slots around the (now absent) target object (see Figure 4.3).

In each training session, participants completed 4 blocks of the visual search task, each lasting approximately 15 minutes. Each block comprised 20 trials from each of 1-8 possible stimuli set sizes, resulting in (20 x 8) 160 trials per block, or (160 x 4) 640 trials per training session. Participants completed 5 visual search training sessions on 5 separate days. On the 6th and last training session, participants completed 2 blocks with the trained objects, followed by 2 training blocks with the novel objects set they had not seen before.

**Data analysis**

Two types of data analysis were conducted: (a) An analysis of visual search slopes across training sessions (i.e., search efficiency analysis), and (b) A “viewpoint-incongruency cost” analysis. For the search efficiency analyses, hypothesis tests were performed using repeated-measures analyses of variance (ANOVAs) of participants’ RTs across training sessions for days 1 through 5. To examine whether potential training effects were task-related only and independent of the stimuli used, search efficiency was also compared between trained objects
The degrees of freedom were corrected when Mauchly’s test revealed significant deviations from sphericity by using the Greenhouse–Geisser epsilon (G-GE); for simplicity, the original values are presented along with the correction factor.

and novel objects on day 6, the last training session. For the viewpoint congruency cost analyses, we assessed whether there was a cost associated with comparing targets across different viewpoints (in terms of RTs and/or accuracy) (see Figure 4.3). To this end, we calculated RTs separately for viewpoint-congruent and viewpoint-incongruent trials, and calculated a viewpoint incongruency cost measure that captured the difference between these two types of trials.

4.3 Results and discussion

Visual search performance across training sessions

Search Slopes. A 5 (training sessions) x 8 (set size) repeated-measures analysis of variance (ANOVA) of reaction times (RTs) revealed main effects of set size \([F(7,119) = 373.05, p < .001, \text{G-GE*} = .17, \eta^2 = .96]\) and training session \([F(4,68) = 26.40, p < .001, \text{G-GE} = .67, \eta^2 = .61]\), indicating that RTs were higher for larger set sizes, and that RTs decreased across the days of training (Figure 4.4A). There was also a significant interaction between the two factors \([F(28, 476) = 3.26, p < .05, \text{G-GE} = .16, \eta^2 = .16]\), driven primarily by larger improvement in RT on the larger set sizes (6,7,8) on the second day of training. Next, we investigated performance on the last day of training (Day 6) to compare RTs for trained and untrained objects. A 2 (trained/untrained objects) x 8 (set size) repeated-measures analysis of variance (ANOVA) of reaction times (RTs) revealed main effects of object category \([F(1,17) = 8.08, p < .05, \eta^2 = .32]\) and set size \([F(7,119) = 208.76, p < .001, \text{G-GE*} = .26, \eta^2 = .93]\), indicating that participants responded faster to trained objects (Figure 4.4B). The interaction between the two factors was significant \([F(7,199) = 2.18, p < .05, \eta^2 = .11]\). Paired samples t-tests to investigate this interaction further revealed that there were significant differences in RT between trained and

* The degrees of freedom were corrected when Mauchly’s test revealed significant deviations from sphericity by using the Greenhouse–Geisser epsilon (G-GE); for simplicity, the original values are presented along with the correction factor.
untrained objects on set size 3 ($t(17) = 2.15, p < .05$), set size 5 ($t(17) = 2.29, p < .05$), set size 6 ($t(17) = 2.41, p < .05$), and set size 7 ($t(17) = 2.97, p < .01$).

**Figure 4.4** Mean reaction times by set size (A) across 5 training sessions and (B) for trained and untrained objects on the last training session. Vertical bars represent standard error of the mean (SEM). See Appendix 16 for separate reaction times for target present/absent trials only, and one/two-featured objects only.
Next, we calculated the mean search slopes from the RTs shown in Figure 4.4. A one-way repeated measures ANOVA with days of training as a factor revealed a significant main effect of days of training \([F(4,68) = 4.75, p < .01, G-GE = .55, \eta^2 = .22]\), indicating that the search slopes became shallower across the 5 training sessions (Figure 4.5). Post-hoc pairwise comparisons (Bonferroni corrected) across consecutive training sessions showed a significant difference between the search slopes of day 1 (M = 144.67) and day 2 (M = 128.08, \(t(1,17) = 3.29, p < 0.05\)). All other comparisons were non-significant.

A paired t-test between the search slopes for trained and untrained novel objects on the last day of training (day 6) revealed a significant difference between them: search slopes of trained objects were shallower (M = 119.46, SD = 31.06) compared to those for untrained objects (M = 131.36, SD = 33.12; \(t(17) = 2.78, p < 0.05\)).

![Average Visual Search Slope](image)

**Figure 4.5** Mean search slopes by training session. Vertical bars represent standard error of the mean (SEM). See Appendix 17 for separate search slopes for target present/absent trials only, and one/two-featured objects only.
Accuracy. An ANOVA with accuracy scores across days of training (Figure 4.6) was not significant (p = .218), although there was a significant linear trend (F(1,17)=4.83, p< 0.05).

![Mean Accuracy](image)

**Figure 4.6.** Mean accuracy across training sessions. Vertical bars represent standard error of the mean (SEM). See Appendix 18 for accuracy displayed separately for target present/absent trials, and one/two-featured objects.

**Effect of viewpoint congruency between sample target and choice target**

To assess whether comparing targets across different viewpoints (see Figure 4.3) affected search performance and whether the ability to do so improved during the course of training, we calculated RTs and accuracy separately for viewpoint-congruent and viewpoint-incongruent trials. We then computed measures that captured the “viewpoint-incongruency cost” of comparing a target from different vs. the same viewpoint for each of our dependent variables (for RT: RTs on viewpoint incongruent trials – RTs on viewpoint congruent trials; for accuracy: accuracy on viewpoint incongruent trials – accuracy on viewpoint congruent trials; for speed-accuracy trade-off: RTs/accuracy on viewpoint incongruent trials – RTs/accuracy on viewpoint congruent trials).
**Reaction time.** A 5 (training sessions) x 8 (set size) repeated measures ANOVA of RTs did not reveal main effects of training session, set size, or an interaction between these two factors (all Fs < 1.87, all ps > 0.16) (Figure 4.7A). A viewpoint incongruency cost analysis between trained and untrained objects on the last day of training was non-significant as well, both for these two factors and their interaction (all Fs < 1.46, all ps > 0.19) (Figure 4.7B).
Figure 4.7 RT viewpoint incongruency costs by set size (A) for training sessions 1 through 5, and (B) for trained and novel objects on the last training session. Vertical bars represent standard error of the mean (SEM).
Figure 4.8 Viewpoint incongruency accuracy costs by set size (A) for training sessions 1 through 5, and (B) for trained and novel objects on the last training session. Vertical bars represent standard error of the mean (SEM).
**Accuracy.** A 5 (training sessions) x 8 (set size) repeated measures ANOVA of accuracy scores did not reveal main effects of training session, set size, or an interaction between these two factors (all Fs < 1.32, all ps > 0.27) (Figure 4.8A). A viewpoint incongruency cost analysis between trained and untrained objects on the last day of training was non-significant as well, both for set size, trained/untrained objects and their interaction (all Fs < 1.55, all ps > 0.16) (Figure 4.8B).

![Viewpoint Incongruency Cost: Speed-Accuracy Trade-off](image)

**Figure 4.9** Viewpoint incongruency analysis: speed-accuracy trade-off across training sessions. Vertical bars represent standard error of the mean (SEM).

**Speed-Accuracy trade-off.** Taking into consideration the notion that participants might have utilized different strategies when performing the practice sessions (some might have emphasised speed while others accuracy), we also analyzed our data using an inverse efficiency measure, which combines both reaction time and accuracy (Townsend & Ashby, 1978, 1983). We first
calculated the mean reaction times divided by the proportion of trials correct, for viewpoint-congruent and view-incongruent trials separately and then calculated the “viewpoint incongruency cost” measure.

A repeated measures ANOVA with these speed-accuracy trade-off scores revealed a significant main effect of days of training (F(4,68) = 2.75, p < .05, \( \eta^2 = .14 \)), indicating that viewpoint incongruency cost decreased with more days of training (Figure 4.9). A paired t-test between speed-accuracy trade-off scores for trained vs. novel objects on the last day of training was not significant (t(17) = 0.66, p = 0.52), suggesting that processing objects from different viewpoints was just as challenging for trained objects as it was for untrained objects.

4.4 General Discussion

The goals of this experiment were to investigate whether visual search training would promote unitization of object features and induce view invariance, before measuring the neural correlates of these potential effects with fMRI. Results from this study showed that participants’ performance on the task improved during the course of training: RTs and search slopes decreased across training sessions, indicating that search became more efficient. When novel objects were used on the last training session, search slopes increased compared to the trained objects, indicating that the improvement in performance was related to participants’ experience with the object features rather than purely practice-related benefits that were independent of the stimuli used. In addition, as indexed by decreases in speed-accuracy trade-off, identifying object features from different viewpoints became easier across training sessions, although complete view-invariance was not achieved.

One question of interest regarding these results is whether this experiment achieved its goal – that is – were we successful in promoting perceptual unitization of object features? The
fact that search slopes became shallower while accuracy was not sacrificed is evidence for perceptual learning (Goldstone, 1998) and unitization of features (Czerwinski et al., 1992).

Although by the classic account of FIT (Treisman & Gelade, 1980) an efficient search is thought to be characterized by flat search slopes (parallel search), given the complexity of our stimuli we did not expect flat search slopes. Likewise, in a study that used a similar paradigm with simpler stimuli (see Figure 4.1), search slopes were still relatively steep (70ms per item) following 50 days of training (Czerwinski et al., 1992). In comparison, in our experiment the mean search slope was 119ms per item on the last day of training. Given that in the Czerwinski et al. study there were still improvements after 35 days of training, it is likely that search slopes would have continued to decrease with additional training sessions (although like in their study, unlikely to have become flat). However, the risk of attaining a high drop-out rate by using such a demanding training programme (e.g. Cousineau et al., 2015) and our hope that all of the participants would take part in the follow-up fMRI study influenced our decision to limit training to 6 sessions.

Although we would like to argue that at least some unitization occurred, as evidenced by the fact that search became more efficient with training, it is hard to assess whether the strict definition of feature binding was satisfied. That is, did two or more features necessarily combine to create a cohesive unit when solving the search trials? This is possible, but we cannot rule out other potential explanations for our participants’ improved performance. One possibility, for example, is that the detection of one feature facilitated (or lowered the threshold for) the detection of a second feature with which it reliably co-occurs (Goldstone, 1998; Laberge, 1973; Mordkoff & Yantis, 1991). In that sense, features were not necessarily unitized, but were processed faster as a result of the visual system’s ability to exploit statistical regularities in the visual input. A future eye-tracking experiment might shed light on this issue (i.e., do participants
fixate more on one of the features, alternate fixations between features, or fixate on the
intersection between features?) One point worth emphasising is that this task was challenging
and whenever the target had more than one feature, the trial could not be solved by relying on a
single feature alone. Detecting target objects in these cases required the ability to process
multiple features in order to compare the object with the distractors in the search array, many of
which shared the same features as the target.

Results from the viewpoint congruency analysis suggest that identifying object features
from different viewpoints became easier across training sessions. However, training gains were
early and then plateaued, and complete view-invariance was not achieved. These results are
consistent with other studies reporting viewpoint-dependency for objects throughout expertise
training (e.g. Gauthier & Tarr, 2002). Likewise, the results do not support a model in which
view-invariant information is immediately exploited without prior engagement with the objects
(Biederman & Bar, 1999; Biederman & Gerhardstein, 1993), as is evident from the difference in
processing speeds between congruent and incongruent viewpoint trials on the very first training
session. Although there was a clear cost associated with comparing objects from different
viewpoints, participants did improve across training sessions (smaller difference between
viewpoint congruent and incongruent comparisons), and it is conceivable that with more training
sessions this difference would have decreased further. On the other hand, our results do not
completely support a strong interpretation of the view-independent model either (e.g. Bülthoff &
Edelman, 1992; Tarr, 1995). On the last training session (day 6) there was no difference in
viewpoint incongruency costs between trained and novel objects – i.e., the slowing of reaction
times due to the viewpoint shift was not different across trained and untrained objects. As with
any null result, however, it is conceivable that with more training sessions we would have observed significantly smaller incongruency costs for trained objects.

It is also important to keep in mind the context under which these results were obtained. Visual search with highly confusable distractors, as used in this experiment, is quite different from many of the tasks used in the viewpoint debate literature (e.g., comparing only two objects separated by a mask) (Biederman & Bar, 1999; Biederman & Gerhardstein, 1993). Perhaps complete view-invariance would have been achieved had our distractors been less complex, or had the object features not mixed between feature sets.

The next chapter used fMRI to investigate whether the series of training sessions reported here influenced the way the feature conjunctions were represented in the brain.
Chapter 5: Experience modifies the conjunctive neural representations necessary for complex object perception

5.1. Introduction

The goal of the study reported in the previous chapter was to familiarize participants with a set of objects and to promote perceptual unitization of their features through an extensive visual search training program. We next wanted to investigate the neural underpinnings of these effects. The results reported in Chapter 2 revealed conjunctive coding for novel (i.e., untrained) objects throughout the VVS but more prominently so in the PRC, compared to more posterior VVS regions. In this study, we wanted to investigate whether following extensive exposure to the objects, these conjunctive representations would remain stable or shift in the VVS continuum. To this end, participants took part in an fMRI study in which they viewed two sets of stimuli – objects on which they have previously trained (Chapter 4), and novel objects they had not seen before.

Evidence suggests that prolonged visual experience changes the tuning curves of neurons in the VVS (e.g. sharpening of tuning, increasing/reducing responses to a preferred/non-preferred stimulus; Op de Beeck & Baker, 2010). For example, in one study, monkeys were trained to discriminate between multipart baton-shaped stimuli that consisted of two distinct elements joined by a vertical stem. Single-neuron recordings showed that neurons in IT cortex responded more selectively to trained batons, and had a marked enhancement of selectivity for the conjunctions of object parts, compared to the individual parts (Baker et al., 2002). Given these and other results from previous studies indicating that extensive visual training changes the spatial distribution of activity in the brain (e.g. Gauthier & Tarr, 2002; Hoffman & Logothetis,
we predicted that the representations of the conjunctions would change as a result of training, and we hypothesized a few potential (and non-mutually exclusive) outcomes, described below.

Hierarchical models of visual processing hold that basic visual features are processed in more posterior regions of the visual stream, whereas feature conjunctions are processed in more anterior regions (Desimone & Ungerleider, 1989; Riesenhuber & Poggio, 1999); the Representational-Hierarchical model (Bussey & Saksida, 2002; Cowell, Bussey, & Saksida, 2010) suggests, as we have found in Chapters 2 and 3, that this hierarchy extends into the PRC. One possibility is that the behavioural training effects reported in Chapter 4 (unitization of object features) led to neural changes that would affect how these conjunctions are represented in the brain. That is, if the unitization process has led to binding of the features, it is conceivable that, over time, they were converted to a fused/simpler/single feature. If this were the case, we could expect conjunctive representations to be shifted to more posterior regions of the VVS, because according to this model, single features are represented more posteriorly. On the other hand, a unitization process could implicate the PRC as well – there is support for its role in conceptual unitization for example, although most of these studies have relied on verbal stimuli (Haskins, Yonelinas, Quamme, & Ranganath, 2008).

Evidence from the expertise literature also support the notion that following training, conjunctive representations will no longer be represented robustly in the PRC, but rather, will shift to more posterior regions of the VVS. There are several possible reasons why this may be the case. For example, numerous studies reported that the LOC is implicated in object expertise (e.g. Grill-Spector et al., 2001; Malach et al., 1995). For example, Op de Beeck et al. (2006) trained participants to discriminated between exemplars within computer-generated 3D novel
object classes. Importantly, objects were manipulated such that they could not be discriminated on the basis of a single portion of an object (e.g. the size/thickness of several shape protrusions were manipulated so that several object segments had to be processed in order to obtain good discrimination performance). They found that training changed the spatial distribution of neural responses across the VVS, with the LOC exhibiting some of the strongest training-related effects. A nearby VVS region, the middle fusiform gyrus (FFA) has also been shown to increase in activity with expertise for novel objects (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999), as well as cars and birds (Gauthier, Skudlarski, Gore, & Anderson, 2000; Xu, 2005). Several studies have suggested that the FFA is engaged in any task that relies on experience to discriminate between individual objects that share a common configuration of features (Harel, Kravitz, & Baker, 2013; McGugin, Gatenby, Gore, & Gauthier, 2012; Gauthier & Tarr, 2002). Given that the structure of our objects is not unlike that of their Greeble-cousins (Gauthier, Williams, Tarr, & Tanaka, 1998) it is possible that we would see an increase in conjunctive representations in the FFA following training.

Another possibility is that, although this was not part of the task demands or intended outcomes of the experiment, participants will have attempted to extract broad-based prototypes as a basis for categorization (Nosofsky et al., 2012), or form explicit categorization rules to guide their behaviour (Nosofsky & Palmeri, 1998). For example, if during the course of training, participants labeled objects from feature set 1 as “builders” and objects from feature set 2 as “diggers”, we might see effects related to conceptualization /semantization of the stimuli. There is increasing evidence that the PRC is important for such high-level conceptual processing (Moss, Rodd, Stamatakis, Bright, & Tyler, 2005; Taylor, Moss, Stamatakis, & Tyler, 2006; Tyler et al., 2003). For instance, a recent study found that activation patterns within the PRC reflected
semantic similarities between highly-confusable objects (Clarke & Tyler, 2014). Along a similar vein, the PRC has been implicated in discriminations involving familiar (e.g., household everyday items), compared to novel (e.g., greebles), objects (Barense, Henson, & Graham, 2011) and it is possible that the increased familiarity with the trained objects may implicate the PRC.

5.2 Materials and Methods

Several aspects of the methods (experimental design and tasks, fMRI data acquisition, MVPA method) were identical to those described in Chapter 2. For ease of reading, I repeat them in the sections below.

Participants

Eighteen neurologically-normal right-handed participants took part in the experiment (19-26 years; M = 23.4, SD = 2.33; 11 females). Functional localizer scans were collected from all but one participant, therefore the ROI MVPA described in this chapter included 17 participants. All participants provided written informed consent approved by the Baycrest Hospital Research Ethics Board and received monetary compensation for their participation. All participants had normal or corrected-to-normal vision.

Experimental Stimuli and Design

Participants viewed novel 3-dimensional objects, created using Strata Design 3D CX 6. Objects belonged to one of two feature-set categories: *Trained Objects* or *Novel Objects*. The *Trained Objects Set* comprised the 28 target objects from the training study described in Chapter 4 (the same objects were also used in the experiments described in Chapters 2-3). These were created from every unique combination of the experimental factors: 2 (feature sets) x 2 (viewpoints) x 7 (possible combinations of features within a set). The *Novel Objects set*, a newly-created stimulus
set that was not previously used (i.e., different from the stimuli used on the last day of visual search training described in Chapter 4), comprised 28 objects with a novel (red) main body. These were composed from the same number of permutations and recombination of features that were used to create the Trained Objects: the 28 objects were created from every unique combination of the experimental factors: 2 (feature sets) x 2 (viewpoints) x 7 (possible combinations of features within a set). Figure 5.1 depicts the complete stimulus set, which comprised 4 feature sets (2 Trained Object feature sets + 2 Novel Object feature sets). During scanning, each participant viewed only one feature set from each of the Trained and Novel stimuli (two feature sets viewed in total). Half of the participants (n=9) viewed feature sets 1&3, and the other half viewed feature sets 2&4.
Figure 5.1 Stimuli used. (A) Trained object set. These are the same objects used in chapters 2-4. Objects had one, two, or three attached features, and were shown from one of two possible viewpoints (pertinent to a separate analysis discussed in Chapter 6). For illustrative purposes, we schematize objects with letters: “A” corresponds to a one-featured object and “AB” corresponds to a two-featured object consisting of features A and B. The trained object set comprised two unique feature sets. (B) Novel object set. This set included objects participants have not seen before. The novel object set comprised two unique features sets, and all had a red main body.
**Tasks**

**Experimental task.** The task was the same as that described in Chapter 2. We administered four experimental scanning runs during which participants completed a 1-back task to encourage attention to each image. Participants were instructed to press a button using their right index finger whenever the same object appeared twice in succession, regardless of its viewpoint (behavioural results in Table 5.1). Feedback was presented following each button press (correct or incorrect) and at the end of each run (proportion of correct responses during that run). Trials in which a response was made were not included in the analysis.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct recognition (%)</td>
<td>89.01 (6.90)</td>
</tr>
<tr>
<td>Hits (%)</td>
<td>89.71</td>
</tr>
<tr>
<td>False alarms (%)</td>
<td>0.69</td>
</tr>
<tr>
<td>Correct rejections (%)</td>
<td>99.30</td>
</tr>
<tr>
<td>Misses (%)</td>
<td>10.29</td>
</tr>
<tr>
<td>Reaction times (ms)</td>
<td>605.17 (61.29)</td>
</tr>
<tr>
<td>Sensitivity (d’)</td>
<td>dL = 7.40 (1.13)</td>
</tr>
</tbody>
</table>

Table 5.1. Behavioural results of the task completed in the scanner.

Mean corrected recognition (proportion of hits – proportion of false alarms), and percentage of hits, false alarms, correct rejections and misses. Mean reaction times (ms) were calculated for correct trials only. Sensitivity (d’) was calculated by applying signal detection theory to logistic distributions: \[dL = \ln\left(\frac{H(1-FA)}{(1-H)FA}\right)\], where \(\ln\) = natural log (Snodgrass & Corwin, 1988). Standard deviations shown in parentheses.

Objects were presented centrally on the screen and had a visual angle of 5.1° x 5.3°, which would likely encompass the receptive fields of PRC (~12°) (Nakamura et al., 1994), V4 (4-6° at an eccentricity of 5.5°) (Kastner et al., 2001), LOC (4-8°) (Dumoulin & Wandell, 2008), FFA and PPA (likely greater than > 6°) (Desimone & Duncan, 1995; Kastner et al., 2001; Kornblith, Cheng, Ohayon, & Tsao, 2013). The visual angle of the individual object features was approximately 2.1° x 2.2°, which would likely encompass the receptive fields of more posterior regions in the VVS (2-4° in V2) (Kastner et al., 2001). Each image was displayed for 1s with a
2s inter-stimulus interval. Each run lasted 11min 30s, and for every 42s of task time, there was an 8s break (to allow BOLD signal to reach baseline) during which a fixation cross appeared on the screen. Each run comprised 6 blocks of 28 trials, which were presented in a different order to each participant. Within each block, the 14 images composing each of the Trained and Novel object sets were randomly presented twice. Trained and Novel object sets were not intermixed within a block. Across consecutive blocks, the object sets alternated (3 blocks per object set per run). Each block contained between one and four target objects (i.e., sequential repeats), such that the overall chance that an object was a target was 10%. In total, each image was presented 24 times (6 times within each run). Prior to scanning, each participant performed a 5-minute practice of 60 trials with the Trained Objects Set only.

**Localizer Task.** After four experimental runs, an independent functional localizer was administered to define participant-specific ROIs (LOC, FFA and PPA, described below). Participants viewed scenes, faces, objects, and scrambled objects in separate 15s blocks (there was no overlap between the images in the experimental task above and the localizer task). Within each block, 20 images were presented for 300ms each with a 450ms ISI. There were four groups of 12 blocks, with each group separated by a 15s fixation-only block. Within each group, three scene, face, object, and scrambled-object blocks were presented (order of block type was counterbalanced across groups). To encourage attention to each image, participants were instructed to press a button with their right index finger whenever the same image appeared twice in succession. Presentation of images within blocks was pseudo-random: immediate repeats occurred between 0-2 times per block. Localizer scans were collected from all but one participant.
fMRI Data Acquisition

Scanning was performed using a 3.0-T Siemens MAGNETOM Trio MRI scanner at the Rotman Research Institute at Baycrest Hospital using a 32-channel receiver head coil. Each scanning session began with the acquisition of a whole-brain high-resolution MP-RAGE T1-weighted structural image (repetition time=2s, echo time=2.63ms, flip angle=9°, field of view=25.6cm², 160 oblique axial slices, 192×256 matrix, slice thickness=1mm). During each of four functional scanning runs, a total of 389 T2*-weighted echo-planar images were acquired using a two-shot gradient echo sequence (200×200mm field of view with a 64×64 matrix size), resulting in an in-plane resolution of 3.1×3.1mm for each of 40 2-mm axial slices that were acquired along the axis of the hippocampus. The interslice gap was 0.5-mm; repetition time=2s; echo time=30msec; flip angle=78°).

Multi-Voxel Pattern Analysis (MVPA)

Functional images were preprocessed and analyzed using SPM8 (www.fil.ion.ucl.ac.uk/spm) and a custom-made, modular toolbox implemented in an automatic analysis pipeline system (https://github.com/rhodricusack/automaticanalysis/wiki). Prior to MVPA, the data were preprocessed, which included realignment of the data to the first functional scan of each run (after five dummy scans were discarded to allow for signal equilibrium), slice-timing correction, coregistration of functional and structural images, nonlinear normalization to the Montreal Neurological Institute (MNI) template brain, and segmentation of gray and white matter. Data were high-pass filtered with a 128s cut-off. The data were then “denoised” by deriving regressors from voxels unrelated to the experimental paradigm and entering these regressors in a general linear model (GLM) analysis of the data, using the GLMdenoise toolbox for Matlab (Kay et al., 2013). Briefly, this procedure includes taking as input a design matrix (specified by the onsets
for each stimulus regardless of its condition) and an fMRI time-series, and returns as output an estimate of the hemodynamic response function (HRF) and BOLD response amplitudes (beta weights). The design matrix did not include the experimental conditions upon which our contrasts relied; these conditions were specified only after denoising the data. Next, a fitting procedure selected voxels that are unrelated to the experiment (cross-validated R² less than 0%), and a principal components analysis (PCA) was performed on the time-series of these voxels to derive noise regressors. A cross-validation procedure then determined the number of regressors that were entered into the model (Kay et al., 2013).

We specified the onsets for each individual object (i.e., A, B, C, AB, BC, AC) for each of the two object sets (Trained and Novel) and two viewpoints. Our model then created a single regressor for each of the three different pairs of objects (i.e., A+BC, B+AC, and C+AB). This was done separately for each of the two object sets and two viewpoints. For example, events corresponding to the singly-presented ‘A’ object from the Novel object set and viewpoint 2 and events corresponding to the singly-presented ‘BC’ object from the Novel object set and viewpoint 2 were combined to create the single regressor for ‘A+BC’ from the Novel object set and viewpoint 2. More specifically, within each run, the voxel-wise data of each object pair were split into 3 subdivisions that were each comprised of every third trial of a given image (following Zeithamova et al., 2012). The pattern similarity of each condition in each subdivision was compared to that of each condition in every other subdivision. We designed the subdivisions so that our comparisons were relatively equidistant in time. For example, the 1<sup>st</sup> subdivision for the A + BC regressor included A<sub>1st presentation</sub> + BC<sub>1st presentation</sub> + A<sub>4th presentation</sub> + BC<sub>4th presentation</sub>; the 2<sup>nd</sup> subdivision included A<sub>2nd presentation</sub> + BC<sub>2nd presentation</sub> + A<sub>5th presentation</sub> + BC<sub>5th presentation</sub>, etc. This resulted in 36 regressors of interest per run [2(object sets) x 2(viewpoints) x 3(conjunctions) x
We also modeled 8 regressors of no interest for each run: trials of three-featured objects (ABC), trials in which participants responded with a button press on the 1-back task, and six realignment parameters to correct for motion. Events were modeled with a delta (stick) function corresponding to the stimulus presentation onset convolved with the canonical hemodynamic response function as defined by SPM8. This resulted in parameter estimates (β) indexing the magnitude of response for each regressor. Multi-voxel patterns associated with each regressor were then Pearson-correlated. Thus, each cell in our planned contrast matrices was comprised of a 12x12 correlation matrix that computed correlations within and across all runs and data subdivisions (Figure 5.2). This process was repeated for each cell in the contrast matrix, and these correlation values were then averaged and condensed to yield the 12x12 contrast matrix (similar to Linke et al., 2011). We then subjected these condensed correlation matrices to our planned contrasts (Figure 5.2).
Figure 5.2 Contrast matrices used to test four experimental questions. MVPA correlations within and between feature conjunctions were summarized in a (12 x 12) matrix structure. (A) Our Conjunctive Coding for Novel Objects contrast investigated whether neural patterns of activity demonstrated explicit conjunctive coding (i.e., was the whole different from the sum of its parts?) only for untrained objects. This contrast tested whether correlations of repetitions of the same conjunctions (dark red) were more similar in their activation pattern compared to correlations of different conjunctions (light red). (B) The Conjunctive Coding for Trained Objects contrast investigated whether explicit conjunctive coding existed for the trained objects. This contrast tested whether correlations of repetitions of the same conjunctions (dark blue) were more similar in their activation pattern compared to correlations of different conjunctions (light blue). (C) Our third contrast tested whether the conjunctive representations for trained objects were greater than those for novel objects (contrast B – contrast A). (D) Our last contrast tested
whether the conjunctive representations for novel objects were greater than those for trained objects (contrast A – contrast B). Note that some of the matrix cells flipped (+/-) symbol to reflect these last two comparisons.

**Searchlight analysis.** A spherical ROI (10 mm radius, restricted to gray matter voxels and including at least 30 voxels) was moved across the entire acquisition volume, and for each ROI, voxel-wise, unsmoothed β-Values were extracted separately for each regressor (Kriegeskorte et al., 2006). The voxel-wise data (i.e., regressors of interest) were then Pearson-correlated within and across runs, and condensed into a 12×12 correlation matrix (see Figure 5.2). Predefined similarity contrasts containing our predictions regarding the relative magnitude of pattern correlations within and between conjunction types specified which matrix elements were then subjected to a two-sample t-test. This analysis was performed on a single-subject level, and a group statistic was then calculated from the average results, indicating whether the ROI under investigation coded information according to the similarity matrix. Information maps were created for each subject by mapping the t-statistic back to the central voxel of each corresponding ROI. These single-subject t-maps were then smoothed with a 12mm FWHM Gaussian kernel to compensate for anatomical variability across participants. The resulting contrast images were then subjected to a group analysis that compared the mean parameter-estimate difference across participants to zero (i.e., a one-sample t-test relative to zero). Results shown in Figure 5.3 are superimposed on the single-subject MNI brain template. Experimental effects exceeding a threshold of p < 0.001 (uncorrected) were considered significant.

**ROI analysis.** We investigated four ROIs defined a priori. Three were functionally-defined regions well-established as part of the VVS: lateral occipital complex (LOC), fusiform face area (FFA), and the parahippocampal place area (PPA). The fourth ROI was the PRC, which was defined by an anatomical probability map created by Devlin and Price (Devlin & Price,
2007). We included areas which had at least a 30% or more probability of being the PRC, as done previously (Barense et al., 2011). For our functional localizer, we used identical stimuli to those employed in Watson et al. (2012). We defined the LOC as the region that was located along the lateral extent of the occipital lobe and responded more strongly to objects compared with scrambled objects (p < 0.001, uncorrected) (Malach et al., 1995). We defined the FFA as the set of contiguous voxels in the mid-fusiform gyrus that showed significantly higher responses to faces compared to scrambled objects (p < 0.001, uncorrected), and the PPA as the set of contiguous voxels in the parahippocampal gyrus that responded significantly more to scenes than to objects (p < 0.001, uncorrected) (Reddy & Kanwisher, 2007). These regions were defined separately for each participant by a 10mm radius sphere centered around the peak voxel in each hemisphere from each contrast, using the MarsBar toolbox for SPM8 (http://marsbar.sourceforge.net/). All LOC and PPA ROIs were bilateral, except for one participant in whom the right LOC could not be localized. FFA ROIs were bilateral in 10 participants; for 6 participants the left FFA could not be localized and for one participant the right FFA was not localized. The ROI MVPA was conducted in an identical manner to the searchlight analysis; voxel-wise data were Pearson-correlated and condensed into a 12×12 correlation matrix, except that here, each ROI was treated as a single region (i.e., no searchlights were moved within an ROI). Before applying our contrasts of interest, we ensured that these correlation values were normally distributed (Jarque-Bera test; P > 0.05). We then applied our two contrasts within each of the four ROIs and obtained, for each participant and each contrast, a t-value reflecting the strength of the difference between our correlations of interest (Figure 5.3). From these t-values, we calculated standard r-effect sizes that allowed us to compare the magnitude of effects across the ROIs (Rosenthal, 1994) (Figure 5.3). Specifically, we
transformed the r-effect sizes to Fisher’s z-scores (as they have better distribution characteristics than correlations; e.g. Mullen, 1989). We then conducted t-tests on the z-scores effect sizes obtained for each region, which provided a measure of the statistical significance between our cells of interest in each of our two contrast matrices (i.e., dark and light colored cells). We then compared the z-scores in each ROI to zero using Bonferroni-corrected one-sample t-tests.

### 5.3 Results

We applied four contrasts to investigate activity patterns reflecting explicit conjunctive coding for (a) Novel Objects (b) Trained Objects (c) Trained > Novel Objects, and (d) Novel > Trained Objects (Fig 5.2). As in Chapter 2, we performed two independent planned analyses using each contrast – a whole-brain *searchlight analysis*, and an *ROI analysis* to investigate activation in specific VVS ROIs.

*Conjunctive coding of novel objects.* The first contrast tested for conjunctive coding of novel objects only (Figure 5.2A). This contrast revealed two large bilateral clusters in the middle occipital gyrus on the right (2061 voxels; peak x,y,z = 34,-72,16, Z-value = 4.38), and on the left (2053 voxels; peak x,y,z = -16,-104,8, Z-value = 4.24), that included V1 and extended into V3v and V4 (Rottschy et al., 2007) (all results reported are p < 0.001, uncorrected). This contrast also revealed a cluster in the left inferior temporal lobe (363 voxels; peak x,y,z = -46,6,-36, Z-value = 4.25) that included the left PRC (Figure 5.3A; peak x,y,z = -36,0,-36, Z-value = 3.24). In addition this contrast revealed a large cluster in the middle cingulate cortex (2875 voxels; peak x,y,z = -0,-34,38, Z-value = 4.24) (Appendix 8 summarizes all regions). These results are consistent with the results obtained in Chapter 2 (*Conjunction Contrast*, Figure 2.2B) in which we also observed PRC and ventral visual cortex activity. In contrast to the results reported in
Chapter 2, the ROI MVPA results did not reveal any significant coding for any of the ROIs (all Ps > 0.41).

*Conjunctive coding of trained objects.* Our second contrast tested for conjunctive coding of trained objects only (Figure 5.2B). This contrast revealed a large cluster in the left fusiform gyrus (10595 voxels, peak x,y,z = -32,-46,-14, Z-value = 5.04), that extended on the left to V3/V4 (Figure 5.3B). In addition, there was a cluster in the left inferior temporal lobe (283 voxels, peak x,y,z = -42,-4,-36, z-value = 4.14) that extended into the PRC (-38,-8,-36, z = 3.34; Appendix 9 summarizes all regions). The ROI MVPA using this contrast revealed conjunctive coding only in PPA (t(16) = 4.00, p < 0.01, r effect size = 0.19), but not in PRC, FFA or LOC (all t(16) < 2.10, ps > 0.2, all r effect size < 0.14) (all one-sample t-tests Bonferroni-corrected). The fact that the ROI MVPA for the FFA was not significant using this contrast, despite the robust left fusiform gyrus activity, could be related to the fact that the FFA could not be localized on the left hemisphere in 6 of the participants, the same hemisphere in which most of the activity was observed with this contrast.

*Conjunctive coding of trained > novel objects.* Our third MVPA contrast tested for conjunctive coding of trained objects vs. novel objects (Figure 5.2C). This contrast determined which regions demonstrated explicit conjunctive coding more strongly for trained compared to novel objects. This contrast revealed only two clusters— One large cluster on the left fusiform gyrus (2931 voxels, peak x,y,z = -40, -50,-20, Z-value = 4.83), that extended into the PRC (x,y,z = -40,-18,-30, Z-value = 3.10), and a smaller cluster on the right fusiform gyrus (531 voxels, peak x,y,z = 38,-42,-6, Z-value = 3.69; Figure 5.3C; Appendix 10 summarizes all regions). These cluster peaks corresponded with the mean FFA ROI coordinates from all participants, defined via an independent localizer scan (mean left FFA ROI x,y,z = -42, -53, -22, very close to the global
maxima of this contrast, reported above; mean right FFA ROI: x,y,z = 43, -52, -21). The ROI MVPA using this contrast revealed conjunctive coding both in the FFA (t(16) = 4.10, p < 0.01, r_{effect size} = 0.21) and the PPA (t(16) = 4.01, p < 0.01, r_{effect size} = 0.23), but not in the LOC (t(16) = 0.29, p = 3.11, r_{effect size} = 0.02) or PRC (t(16) = 1.44, p = 0.68, r_{effect size} = 0.07) (all one-sample t-tests Bonferroni-corrected).

 Conjunctive coding of novel > trained objects. Finally, our fourth MVPA contrast tested for conjunctive coding of novel objects vs. trained objects (Figure 5.2D). This contrast determined which regions demonstrated explicit conjunctive coding more strongly for novel compared to trained objects. This contrast revealed only one small cluster on the left middle occipital gyrus spanning V1 – V3 (272 voxels, peak x,y,z = -20, -102, 8, Z-value = 3.79; see Appendix 11). The ROI MVPA using this contrast did not reveal conjunctive coding in any of the ROIs (all one-sample t-tests Bonferroni-corrected).
Figure 5.3 Regions demonstrating explicit conjunctive coding for novel and trained objects. (all results shown are \( p < 0.001 \), uncorrected). (A) Regions where the conjunctive representations for novel objects were different from the sum of their parts. These regions included the PRC and earlier VVS regions (V1 through V4). ROI MVPA results (shown on the right) did not reveal any significant coding for any of the ROIs. (B) Regions demonstrating conjunctive representations for trained objects (\( p < 0.001 \), uncorrected). Broadly speaking, these regions
included the left V4/fusiform gyrus, extending into the inferior temporal lobe. ROI MVPA results revealed significant conjunctive coding only in the PPA. (C) Regions where the conjunctive coding was stronger for trained objects compared to novel objects. These included mainly the left fusiform cortex. ROI MVPA results indicated that both the FFA and PPA contained conjunctive representations for trained > novel objects. (D) Regions where the conjunctive coding was stronger for novel objects compared to trained objects. Only one cluster in the left middle occipital gyrus was significantly activated. ROI MVPA results did not reveal any significant coding for any of the ROIs. ROI error bars indicate s.e.m.; *p < 0.05, comparisons are relative to zero within an ROI (significant one-sample t-tests indicated by an asterisk above a bar).

5.4 Discussion

The goal of this study was to investigate whether the conjunctive representations of object features would change following extensive training. We scanned participants while they viewed both well-trained objects and novel objects, which they saw in the scanner for the first time. For novel objects, we found similar results to those previously obtained (Chapter 2) – unique feature conjunctions were represented in the PRC and in primary visual cortex, extending to V4. For trained objects, the analysis revealed a different activation profile with activity focused on more posterior regions (middle fusiform gyrus). Further, a contrast that directly compared trained vs. novel feature conjunctions found that conjunctive representations were concentrated in the FFA, extending to the PPA. Our central finding was this different response profile: instead of activity centralized in the PRC and the visual cortex (V1-V4), which was observed for novel objects, we found that the peak activity for conjunctive representations following training was focused in the fusiform gyrus. Finally, these findings also support the view that cortical representations of objects along the VVS are dynamic, and continuously modulated by experience (Op de Beeck et al., 2006).

Our central finding, stronger conjunctive coding in the FFA following training, could have been a result of several processes. First, these could have resulted from expertise effects. It was suggested that the FFA is part of a network tuned by experience to discriminate visually
similar objects (Harel et al., 2013; McGugin et al., 2012). Studies have reported object expertise effects in the FFA for proficiency developed during the course of many years in the domains of birds (Gauthier, Skudlarski, Gore, & Anderson, 2000), cars (Xu, 2005) and chess (Bilalić, Langner, Ulrich, & Grodd, 2011), but also for expertise for novel objects developed in the lab over the course of approximately 7 hours of training (Gauthier et al., 1999). Alternatively, as its name suggests, the fusiform face area was originally proposed to be exclusively devoted to face recognition (Kanwisher, McDermott, & Chun, 1997), and some continue to argue that this is indeed the case (Kanwisher & Yovel, 2006). It is possible that because our objects comprised three distinct features that could potentially be labeled as ‘hat’, ‘hands’ and ‘nose’, some participants viewed them as “face-like”, or ‘little people’, and in turn this contributed to an FFA activity, as this could potentially be the case following intensive training (Tarr & Gauthier, 2000). In addition, it is interesting to note that whereas here conjunctive coding was stronger in the FFA than in the LOC, the opposite was true in Chapter 2 for Novel objects. This is consistent with evidence that the FFA becomes more engaged and the LOC less so with expertise training (e.g., training to interpret radiographic images; Harley et al., 2009; but see Op de Beeck et al., 2006). We also found unexpected conjunctive representations in the PPA following training, although there is some precedence for PPA activity for object perception (Harel et al., 2013; Janzen, Wagensveld, & Van Turennout, 2007) for example, when objects are highly contextualized (Bar, Aminoff, & Aminoff, 2003).

Another possible interpretation of these results is that objects were semanticized to some degree during the course of training. As anecdotal evidence – many of the participants in the study reported that they resorted to labeling the objects (e.g. “The little guy with the funny hands and round hat is Joe”) to assist their performance both during the visual search training sessions
and during the subsequent scanning session, despite not being explicitly instructed to do so. The observed FFA activity is consistent with similar effects reported by Gauthier et al. (1999). In their study, participants trained with novel objects (i.e., Greebles) until they were as fast at categorizing them at the individual level as they were at categorizing them at the “family” level, with performance assessed via name-verification trials during which participants indicated whether a label matched the object that was presented shortly after. This raises the question of whether semantic labeling is a by-product of visual training, and whether alternative training tasks (i.e., not visual search) would have resulted in similar findings. Other studies found that there was increased phase-locking between the left anterior temporal lobe and left fusiform cortex that was associated with basic-level naming (Clarke, Taylor, & Tyler, 2011), indicative of semantic integration with visual object processing, and left middle fusiform gyrus activity for a semantic (vs. non-semantic) task assessing verbal memory (Wagner et al., 1998).

Within the FFA/fusiform region, the increase in conjunctive coding following training was predominantly in the left hemisphere (Figure 5.3). This was not expected, as most FFA-related training effects have been reported to be on the right hemisphere (Gauthier et al., 1999; Gauthier & Tarr, 2002; Harley et al., 2009), although this is not always the case (e.g. Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). A few studies have reported training-related activities on the left hemisphere – for example, an electrophysiological study found that the N170, a bilateral occipitotemporal cortex event-related potential (ERP) component that is typically thought to reflect the neural processing of faces, was modulated by expertise training with Greebles, an effect that was primarily observed in the left hemisphere (Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002). These laterality effects could underlie the inconsistency between the whole-brain searchlight and ROI MVPA results for the Trained Objects contrast.
(Figure 5.3B). More specifically, the fact that the ROI MVPA did not reveal significant conjunctive coding for Trained Objects despite robust fusiform activity as evident from the whole-brain analysis, could be related to the fact that for 6 participants the FFA was not localized in the left hemisphere. In comparison, the fact that the FFA ROI showed stronger conjunctive coding for Trained relative to Novel Objects (Figure 5.3C) is likely related to the fact that for this contrast, the searchlight results were more bilateral.

The mid-fusiform gyrus activity following training is also consistent with the Representational-Hierarchical Model of object processing (Bussey & Saksida, 2002; Cowell, Bussey, & Saksida, 2010). This model holds that basic visual features will be processed in more posterior regions of the visual stream whereas feature conjunctions will be processed in more anterior regions, including the PRC. It is possible that the extensive training period with the objects encouraged the unitization of its constituent features, which in turn has led to the creation of a simpler, fused feature that is represented in earlier regions (i.e., the fusiform gyrus) (see Figure 5.4).

**Figure 5.4** The proposed organization of visual object representations in the VVS according to the Representational-Hierarchical Model. Relatively simple object features are represented in posterior regions of the VVS, simple feature conjunctions are represented in more anterior
regions, and more complex conjunctions, approximately at the level of the object, are represented in the more anterior regions, including the PRC. One interpretation of the results reported in this chapter is that an extensive training period with complex objects encourages the unitization of its constituent features. As a result, feature conjunctions that were represented in the PRC before training, shift to more posterior regions of the hierarchy following training. This process likely occurs via a retuning process of neurons in these earlier regions of the VVS (i.e., fusiform gyrus), that represent more unitized features as (newly-created) simplified features.

That said, we did observe that conjunctive coding for the Trained vs. Novel objects extended into PRC, which might have been predicted based on some evidence in the literature — specifically, there is evidence that when complex stimuli are unitized, they can be represented by the PRC and support familiarity-based memory (Haskins et al., 2008). The unitization interpretation is also supported by the finding that when participants experienced two disjointed visual forms as being fused (vs. non-fused), both the fusiform and PRC cortices were recruited (Rubin, Chesney, Cohen, & Gonsalves, 2013).

It is worth mentioning that our results for the Conjunctive Coding of Novel Objects contrast did not replicate exactly our results from Chapter 2 (Conjunction Contrast, Figure 2.6). Although the PRC was implicated in the current chapter, results were less robust, with results from the MVPA ROI analysis not reaching significance. This could be due to the fact that here we have less power (there were half as many contrast matrix cells because, to accommodate the Trained Object trials, we could show only one feature set). Alternatively, it is possible that although these were novel objects that participants had not seen before, there was some general familiarity with these objects that transferred across stimulus sets, because they all share the same structure. The data from Chapter 4 support this idea (i.e., Day 6 performance on the newly-introduced novel objects (those with the yellow body) was better than Day 1 performance for the objects prior to extensive training). The next chapter investigated training effects on view-invariant conjunctive coding.
Chapter 6: Experience modifies the conjunctive neural representations necessary for complex object perception: Effect of viewpoint

6.1. Introduction

Results from the previous chapter indicated that following training with complex objects, conjunctive representations were concentrated in the fusiform gyrus. Those results focused on analyses that considered feature conjunctions from the same view. The current chapter examines the effects of training on view-invariant conjunctive representations. When examining view-invariant representations in naïve participants (Chapter 3) we found conjunctive coding in the VVS—with maxima in V4, as well as activity in lateral IT cortex that extended anteriorly to the left PRC. A direct comparison across brain regions indicated that this view-invariant conjunctive coding was stronger in the PRC compared to the three more posterior VVS ROIs. In the current chapter, we examined the effects of the behavioural training (previously discussed in Chapter 4) on the view-invariant representation of these objects.

Previous research indicates that training affects view-invariant representations throughout the VVS. For example, single-unit recordings in nonhuman primates following training with complex objects found view-invariant neurons in the IT cortex, particularly, the anterior medial temporal sulcus (AMTS) and superior temporal sulcus (STS) (Booth & Rolls, 1998; Logothetis & Pauls, 1995). After monkeys were allowed to explore a set of objects in their living environment, stronger view-invariance was observed for the studied objects compared to novel objects (Booth & Rolls, 1998). Furthermore, the PRC has been found to be implicated in view-invariant perception as well —studies show that both nonhuman primates and humans with PRC
damage were impaired at discriminations involving complex objects and faces presented from different viewpoints (Buckley et al., 2001; Barense et al., 2005, 2007; Lee et al., 2006b). In addition, Barense et al. (2010) found increased PRC activity in neurologically intact participants who discriminated between faces and objects presented from different (rather than identical) viewpoints. Others have also found that neurons in the MTL were invariant not only to viewpoints of the same person or object, but also to different representations of its overall concept (e.g., a single neuron was selectively activated by pictures of the actress Halle Berry, but also to a drawing of her, and the letter string “Halle Berry”; Quiroga et al., 2005).

As discussed in the introduction, there is debate surrounding the nature of view-invariant object processing. There is behavioural evidence to support both a view-invariant model (e.g. Biederman & Bar, 1999; Biederman & Gerhardstein, 1995) and a view-dependent model (e.g. Tarr & Pinker, 1989; Tarr, 1995). Neuroimaging studies are inconclusive as well. For example, fMRI repetition suppression effects were reported to be greatest when objects are repeatedly presented from the same viewpoint, and decreased with incremental changes in viewpoint (Andresen et al., 2009; Gauthier et al., 2002), supporting a view-specific model. However, other studies have reported at least some degree of view-invariance in the fusiform cortex, for example (Eger, Henson, Driver, & Dolan, 2004; Simons, Koutstaal, Prince, Wagner, & Schacter, 2003; Vuilleumier et al., 2002). An important distinction between these two models is the way that they account for the role of experience in creating view-invariant representations. The view-invariant model suggests that view-invariant information is extracted immediately, even from objects never encountered before (Biederman & Bar, 1999). In contrast, according to the view-dependent model, view-invariance is developed over time as a result of multiple encounters with a stimulus from several different viewpoints. As such, these two views make different
predictions about the nature of view-invariant conjunctive representations following training –
the former suggests that one should not expect to see much difference between novel and trained
objects, because view-invariance has already been achieved from the outset, while the latter view
predicts that conjunctive representations will change as a result of training, because view-
invariant representations are not yet fully formed and are moulded over time.

6.2 Materials and Methods

All Participants, Experimental Stimuli and Design, and Tasks were identical to those described
in Chapter 5. For a review see 5.2 Materials and Methods.

Multi-Voxel Pattern Analysis (MVPA)

The data analysis conducted the same as the one described in Chapter 5, with the exception that
here we defined four new contrast matrices in order to investigate view-invariant conjunctive
coding for novel and trained objects, and compare trained vs. novel objects (Figure 6.1).
Figure 6.1 Contrast matrices used to test three experimental questions. MVPA correlations within and between feature conjunctions were summarized in a (12 x 12) matrix structure. (A) Our first contrast tested whether neural patterns of activity demonstrated view-invariant conjunctive coding (i.e., was the whole different from the sum of its parts?) only for untrained objects. This contrast tested whether correlations of repetitions of the same conjunctions (dark red) were more similar in their activation pattern compared to correlations of different conjunctions (light red). (B) The second contrast investigated whether view-invariant conjunctive coding existed for the trained objects. This contrast tested whether correlations of repetitions of the same conjunctions (dark blue) were more similar in their activation pattern compared to correlations of different conjunctions (light blue). (C) Our third contrast tested whether view-invariant conjunctive representations for trained objects were greater than those
for novel objects (contrast B – contrast A). (D) Our last contrast tested whether view-invariant conjunctive representations for novel objects were greater than those for trained objects (contrast A – contrast B). Note that some of the matrix cells flipped (+/-) symbol to reflect these last two comparisons.

6.3 Results

Four contrasts determined which regions demonstrated view-invariant conjunctive coding for (a) novel objects (b) trained objects (c) trained > novel objects, and (d) Novel > Trained Objects (Fig 6.1). We repeated the same analysis procedure by performing two independent planned analyses – a whole-brain searchlight analysis, and an ROI analysis to investigate activation in specific VVS ROIs.

View-invariant conjunctive coding of novel objects. The first MVPA contrast tested for view-invariant conjunctive coding of novel objects (Figure 6.1A). This contrast revealed one large cluster in the left middle occipital gyrus/ V4 (1013 voxels; peak x,y,z = -26,-84,0, Z-value = 3.54), and a right MTL cluster that included the PRC (112 voxels; peak x,y,z = 36,8,-36, Z-value = 2.93) (Figure 6.2A; all results reported are p < 0.005, uncorrected). Appendix 12 summarizes all observed regions. The ROI MVPA results did not reveal significant coding for any of the ROIs (all Ps = 0.99, all one-sample t-tests, Bonferroni-corrected).

View-invariant conjunctive coding of trained objects. The second MVPA contrast tested for view-invariant conjunctive coding of trained objects (Figure 6.1B). This contrast revealed a large cluster in the left fusiform gyrus (1263 voxels, peak x,y,z = -38,-32,-30, z-value = 3.31) that extended anteriorly to the PRC (peak x,y,z = -36,-20,-32, Z-value = 2.79), and a smaller cluster in the right visual cortex (V1/V2) (495 voxels, peak x,y,z = 20,-96,-2, z-value = 3.85; Figure 6.2B). Appendix 13 summarizes all observed regions. The ROI MVPA results did not reveal significant coding for any of the ROIs (all Ps > 0.13).
View-invariant conjunctive coding of trained > novel objects. The third MVPA contrast tested for view-invariant conjunctive coding of trained objects vs. novel objects (Figure 6.1C). This contrast tested which regions demonstrated view-invariant conjunctive coding more strongly for trained compared to novel objects. This contrast revealed a larger cluster on the left fusiform gyrus (5044 voxels, peak x,y,z = -42,-50,-36, Z-value = 4.33), that extended into the PRC (x,y,z = -36,-20,-32, Z-value = 3.44), and a cluster on the right fusiform gyrus (4162 voxels, peak x,y,z = 16,-46,-30, Z-value = 4.52; Figure 6.2C). Appendix 14 summarizes all observed regions. The ROI MVPA revealed conjunctive coding in the FFA only (t(16) = 3.64, p < 0.01, r_{effect size} = 0.20), but not for any of the other three ROIs (all Ps > 0.23, all one-sample t-tests, Bonferroni-corrected).

View-invariant conjunctive coding of novel > trained objects. Finally, the last MVPA contrast tested for view-invariant conjunctive coding of novel objects vs. trained objects. This contrast determined which regions demonstrated explicit view-invariant conjunctive coding more strongly for novel compared to trained objects. This contrast revealed only one cluster on the left middle occipital gyrus, spanning V1 – V3 (834 voxels, peak x,y,z = -18, -104, 4, Z-value = 3.99), and a smaller cluster in the right superior occipital gyrus (121 voxels, peak x,y,z = 24, -94, 22, Z-value = 3.04; Figure 6.2D; see Appendix 15). The ROI MVPA using this contrast did not reveal conjunctive coding in any of the ROIs (all one-sample t-tests, Bonferroni-corrected).
Figure 6.2 Regions demonstrating view-invariant conjunctive coding for novel and trained objects. (A) Regions where view-invariant conjunctive representations for novel objects were different from the sum of their parts (p < 0.005, uncorrected). These regions included the right PRC and left V4/fusiform gyrus. ROI MVPA results (shown on the right) did not reveal significant coding for any of the ROIs. (B) Regions demonstrating view-invariant conjunctive representations for trained objects (p < 0.005, uncorrected). These regions included the left
V4/fusiform gyrus, extending into the inferior temporal lobe, as well as the right visual cortex (see Appendix 13 for all regions). ROI MVPA results did not reveal significant coding for any of the ROIs. (C) Regions where view-invariant conjunctive coding was stronger for trained objects compared to novel objects. These regions included bilateral fusiform gyrus, extending to the PRC on the left. ROI MVPA results indicated that the FFA contained view-invariant conjunctive representations for trained vs. novel objects. (D) Regions where view-invariant conjunctive coding was stronger for novel objects compared to trained objects. Only two clusters in the occipital gyrus (V1/V2) were activated. ROI MVPA results did not reveal any significant coding for any of the ROIs. ROI error bars indicate SEM; *p < 0.05, comparisons are relative to zero within an ROI (significant one-sample t-tests indicated by an asterisk above the bar).

### 6.4 Discussion

The focus of this chapter was to investigate view-invariance in conjunctive coding of object features following extensive exposure to those objects. For novel objects, view-invariant conjunctive representations concentrated in a cluster in the visual cortex that included V4, and a cluster in MTL that included the PRC — these results replicate the findings previously reported in naïve participants (Viewpoint Contrast; Chapter 3). For trained objects, results showed that view-invariant representations were concentrated in the left fusiform gyrus and extended anteriorly to include the PRC. A direct comparison between trained and novel objects revealed a large cluster on the left fusiform gyrus that extended to the PRC, and an ROI MVPA with the same comparison revealed that the FFA was the only predefined region that significantly coded view-invariant representations.

The main findings from this chapter indicate that following training, view-invariant representations became stronger, and shifted to more posterior regions of the VVS, in particular the FFA. This is consistent with the literature on expertise, which suggests that the FFA plays a role in view-invariant processing of well-practiced stimuli, such as faces (Axelrod & Yovel, 2012; Ewbank & Andrews, 2008) and objects (Pourtois, Schwartz, Spiridon, Martuzzi, & Vuilleumier, 2009; Vuilleumier et al., 2002). The FFA activity we observed is also consistent
with an interpretation of the representational-hierarchical model in which extensive training with complex objects encourages the unitization of its constituent features. According to this interpretation, such a unitization process will cause feature conjunctions that were once represented in anterior VVS regions (before training), to be shifted to more posterior regions of the VVS hierarchy (following training), where a newly-created, unified feature is now represented (see Figure 5.4).

These results are also consistent with recent studies reporting sensitivity to viewpoint mirror-symmetry for faces in monkeys (Freiwald & Tsao, 2010) and humans (Kietzmann, Swisher, König, & Tong, 2012). Kietzmann et al. (2012) found that cortical activity patterns for mirror-symmetric views of faces were spatially distributed across a large band of higher order visual areas that included the FFA and PPA. Although in this study, rotations were not entirely view-symmetrical (objects were separated by a 70° rotation along a single axis: 25° to the right, and 45° to the left from central fixation), it is possible that more complete view-invariance is achieved in more anterior regions; for example, if the visual stimuli required a higher degree of feature conjunctions (for example, a wider selection of possible viewing angles) more anterior regions would have been recruited more robustly.

It is worth noting that the PRC also represented view-invariant conjunctions for both novel and trained objects, and a direct comparison between trained and novel objects revealed that the PRC was involved (at least according to the searchlight analysis). This is consistent with previous studies indicating that the PRC is important for processing conjunctions across different viewpoints (Barense et al., 2010). The fact that the PRC ROI MVPA was not significant for View-invariant conjunctive coding of novel objects (whereas it was significant for view-invariant conjunctions in Chapter 3) could be related to the fact that in this study only half of the data was
used to calculate view-invariant novel/trained conjunctions (because half of the objects were novel and half trained, while in Chapter 3 all were novel). Alternatively, as was suggested in the previous chapter, it is possible that because the overall configuration of the novel objects was similar to that of the trained objects, some of the training effects transferred to the novel objects as well.

One thing to note is that similar to Chapter 3, the overall activity of the view-invariance analysis was generally weaker than conjunctive representations from the same view (Chapter 2 and 5) — judging by the whole-brain searchlight analyses. This is supported by single-neuron recordings in monkeys that show that in IT cortex, for example, there are proportionally more view-specific neurons compared to view-invariant neurons (49% compared to 14%; Booth & Rolls, 1998). View-invariant perception is hard, and accomplished by fewer neurons.

Finally, results from this chapter solidify the findings reported in Chapter 5 — together, these results indicate that training contributes to the creation of view-invariant neuronal representations that are hosted in more posterior regions of the VVS. These findings support an interpretation of the view-dependent model of object perception (Tarr & Pinker, 1989; Tarr, 1995) according to which, view-invariance is developed over time as a result of multiple exposures to a stimulus from several different viewpoints. Furthermore, these findings suggest that view-invariant conjunctive representations change as a result of training, and that cortical representations along the VVS are dynamic and modulated by experience (Op de Beeck et al., 2006).
Chapter 7: General Discussion

The purpose of this research project was to investigate the manner in which our visual system processes complex objects. This dissertation described a novel application of MVPA of neuroimaging data that allows a direct investigation of whether neural representations integrate object features into a whole that is different from the sum of its parts, and whether these representations are view invariant. Results from Chapter 2 and 3 showed that patterns of activity throughout the ventral visual stream (VVS), extending anteriorly into the perirhinal cortex (PRC), discriminated between the precise conjunctions of features comprising a complex object. Results from this study provide the first direct evidence that information coding in PRC simultaneously discriminated between the precise conjunctions of features comprising an object, and was also invariant to changes in the visual appearance of those conjunctions caused by shifting viewpoints. These coding principles distinguished the PRC from other functionally-defined regions in the VVS that are more classically associated with perception (e.g., LOC, FFA, and PPA).

In a subsequent study, we investigated whether extensive training on a visual search task promoted the unitization of object features (Chapter 4), tested whether training influenced the way these feature conjunctions were represented in the brain (Chapter 5), and whether it influenced the way these objects were represented across different viewpoints (Chapter 6). Results showed that visual search became more efficient over the course of training, suggesting that a unitization process took place. The subsequent fMRI study revealed a different activation profile for trained objects compared to novel objects. Whereas for novel objects, conjunctive representations were focused on the PRC and in primary visual cortex (extending to V4),
representations for trained objects were concentrated in more posterior VVS regions — mostly in the middle fusiform gyrus. An analysis of view-invariant conjunctive coding strengthened these findings — whereas for novel objects, representations were concentrated mostly in the PRC, conjunctive representations for trained objects shifted to more posterior regions, concentrating in the FFA. Together, these findings have implications for current theoretical models of visual object learning.

**Conjunctive representations in the VVS are dynamic**

One question worth considering is how the results from the studies described in this thesis relate to the theoretical framework of the representational-hierarchical model. This model posits that basic visual information is processed in posterior regions of the VVS stream, and more complex feature conjunctions in more anterior regions, with the PRC representing conjunctive representation at approximately the level of an object. The results obtained in Chapter 5 indicated that an extensive training period with complex objects led to a change in how they were represented. More specifically, conjunctive representations that were represented in the PRC before training, shifted to more posterior VVS regions (fusiform gyrus) after training. We speculate that this was because feature conjunctions were unitized and a new unified single feature was created in their stead (suggested by the posterior VVS activity), this would raise an intriguing implication regarding the nature of the visual representations in the VVS. More specifically, this suggests that object representations are malleable and are dependent on prior experience. Moreover, this suggests that what constitutes a “feature” or an “object” is a relative concept. That is, what one observer may perceive as an object comprised of two separate features, another observer may perceive as a unified single feature. This also raises the possibility that should unified conjunctive representations shift to more posterior VVS regions,
anterior regions (PRC) would be free to process representations that are yet more complex (for example, if a new feature is attached to a complex object whose features have been unitized through training).

Future experiments could investigate these possibilities. For instance, participants could be trained to perceive a visual structure as being composed of a conjunction of individual features, whereas other participants could perceive the same structure, but never experience the features that constitute it as separate. At test (and during scanning), additional features could then be attached on top of this unified structure. For example, considering the objects used in this thesis, this is akin to never separating features A and B; or on the other hand, dividing the basic blue body (Figure 2.1) into two separate features that together compose the main body (in the experiments in this dissertation, all participants learned to treat the blue body as a single cohesive structure, but this need not necessarily be the case).

The training effects on view-invariant conjunctive representations also have intriguing implications for the representational-hierarchical model. Results from Chapter 6 indicate that following training, view-invariant conjunctive representations shifted from anterior VVS regions (i.e., PRC) to more posterior regions of the VVS, in particular the FFA. This is consistent with an interpretation of the representational-hierarchical model in which extensive training with complex objects encourages the unitization of its constituent features into a view-invariant, abstract representation, and that perceiving complex objects across different viewpoints requires the ability to combine its different constituent features into such a unified, and more abstract representation. However, these results were obtained using trained objects that were presented only from two possible viewpoints. This raises a few questions. First, is it possible that more complete view-invariance is achieved in more anterior regions? That is, if after training, view-
invariant representations shifted to the FFA, would newly-presented viewpoints of the same objects be represented in these same posterior regions, or would they be represented in more anterior regions (i.e., PRC)? A related question is whether a longer training period would have been necessary to achieve the same results had a wider selection of possible viewpoints been used. In other words, what is the relationship between (a) the selection of viewpoints from which objects are presented, (b) the amount of training sessions necessary to achieve view-invariant perception, and (c) the influence that these two factors exert on the neural representations of complex objects along the VVS hierarchy.

**Methodological considerations**

Another point worth emphasising is that the findings in this thesis need to be considered within the limits of the employed methodology. The measurement of neuronal activity as measured by fMRI is indirect, involves the magnetization levels of blood as a marker for neural activity, and is a slow and nonlinear hemodynamic response (Haynes, 2015; Logothetis & Wandell, 2004). Moreover, our MVPA approach is only one of many – there is no single best method for investigating representational content with fMRI (Davis & Poldrack, 2013). Our chosen method is not free from some of the limitations other pattern analysis methods face. For example, results from the searchlight analyses need to be interpreted with care – each voxel in a searchlight map depicts the ability of the local voxel-sphere to represent the information content under investigation. Hence, the resulting maps depict the centres of the informative voxel spheres, but not the informative voxels themselves (Haynes, 2015). Moreover, two spheres might be equally informative, but represent neural information in different ways. For example, in one case only a single voxel in a sphere may be informative, whereas in a different sphere, the majority of voxels could be informative (Viswanathan, Cieslak, & Grafton, 2012).
In addition, the relation between our MVPA results and the computations performed by individual neurons is not straightforward; fMRI has a limited spatial resolution, in which each voxel can contain up to several million neurons (Logothetis, 2008). For example, we found view-invariant conjunctive representations in the FFA following training, but what can we say about the computations performed by individual neurons in that region? Single-unit recordings in monkeys suggest that only a small percentage of IT neurons are view-invariant (Booth & Rolls, 1998; Logothetis, Pauls, & Poggio, 1995). Is it the case then, that following training, the proportion of view-invariant neurons in the fusiform gyrus increased? Or is it the case that training resulted in a redistribution of view-invariant neurons — for example, if before training neurons computing view-invariant conjunctive representations were less selective and distributed across a wider region of the VVS, and after training, were more sparse and clustered in a smaller region (Reddy & Kanwisher, 2006). Another possibility is that view-invariant conjunctive representations were computed mostly by view-dependent neurons. That is, if the FFA in fact contained the same proportion of view-dependent and view-invariant neurons, before and after training, and training simply resulted in a more efficient process of pooling across multiple view-selective neurons to achieve view-invariance (Perrett et al., 1998).

Future studies

The role of categorization. Following training, conjunctive representations were concentrated in the FFA, and extended anteriorly to the PRC. These results might be related to a semantization process in which visual objects are conceptualized over time. For example, this activation pattern is consistent with a study in which participants were asked to name objects at either a domain level (e.g. indicate whether an item is living or nonliving), or at the more basic level (where they were asked to produce the unique name of the object, e.g. cow, hammer, etc.)
MEG recordings showed increased phase-locking between the left anterior temporal lobe and left fusiform cortex during the basic-level naming condition, indicating that there were recurrent interactions between the two regions, and that this process was modulated by greater semantic integration demands. Another study found that the perirhinal cortex represented object-specific semantic information and that PRC representations were modulated by how semantically confusable the objects were (Clarke & Tyler, 2014). Future research could investigate how the conjunctive representations reported in this thesis relate to categorization/semantic information. For example, a similar paradigm could be used, in which semantic labels are assigned both at the feature set (i.e., category) level and the item-specific (i.e., individual object) level. It would be interesting to know if view-invariant conjunctive representations in posterior/anterior VVS regions are modulated by a requirement to process objects at the category/individual object level.

In the wild, category membership is often bound up with visual similarity. For example, golden retriever and labrador are visually similar and also from the same category. But there are exceptions which we have no trouble dealing with: Chihuahua and Bullmastiff; hummingbirds vs ostrichs; pear vs light-bulb. Another future study could investigate how conjunctive representations are influenced by both semantic categorization and visual similarity. For example, such a study could involve the creation of several object prototypes, each serving as a “head of a family”. Each prototype would comprise a different constellation of features within a given feature set (perhaps with no overlapping features across prototypes). Additional objects could then be treated as “family members” and belong to one of the families if they share a certain number of features with a specific prototype / “head of a family”. To be able to orthogonally manipulate visually similarity, non-informative features would also be included,
such that we could create across-family objects that share the same number of features as within-family objects. Such a paradigm would enable the dissociation of visual similarity from categorical similarity the investigation of how conjunctive representations are differentially influenced by these two factors – with perhaps more anterior VVS regions recruited for visual similarity / categorical decisions that rely on multiple features (Nosofsky et al., 2012; Zaki & Nosofsky, 2001).

*Functional connectivity.* The findings from Chapter 5 and 6 suggest that it might be worthwhile to investigate patterns of functional connectivity between the PRC and earlier visual stream regions. In particular, it is possible that connectivity patterns will change before and after training – with perhaps the PRC having stronger connectivity with occipital cortex regions, including V4 before training, and stronger connectivity with the fusiform gyrus, including the FFA following training. Previous research indicated that there is both structural and functional connectivity between these regions. For example, a diffusion tensor (DT)-tractography study showed that the inferior longitudinal fasciculus, a major white-matter tract traversing the visual ventral stream, directly connects occipital and anterior temporal regions, including V4 and the fusiform gyrus (Catani, Jones, Donato, & Ffytche, 2003). In addition, fMRI studies have shown that the PRC has both intrinsic connectivity with fusiform regions (Libby, Ekstrom, Ragland, & Ranganath, 2012), and also task-based connectivity— for example, in a study in which participants performed discriminations between similar faces, that could not be based on simple perceptual features, the PRC showed strong functional connectivity to the fusiform gyrus (O’Neil et al., 2011). This suggests that for well-trained stimuli such as faces, the PRC exchanges information with the fusiform gyrus.
Conclusion

Collectively, the studies presented in this dissertation provide new evidence regarding the functional architecture of object perception, showing neural representations that both integrated object features into a whole that was different from the sum of its parts and were also invariant to large changes in viewpoint. These computation principles distinguished the PRC from other functionally-defined regions in the VVS that are more classically associated with perception (e.g., LOC, FFA, and PPA). These results suggest that that visual-object processing does not culminate in IT cortex, but instead continues into MTL regions traditionally associated with memory. Visual training resulted in a different activation profile— with view invariant conjunctive representations concentrated in the fusiform gyrus. These findings support the representational-hierarchical view of object perception, and suggest that cortical representations along this hierarchy are dynamic and modulated by experience.
References


Appendix 1. Regions demonstrating explicit conjunctive coding: analyses computing correlations across runs only (i.e., correlations within runs did not contribute to the analysis). These analyses revealed the same pattern of results as our main findings (Figure 2.6), indicating that our results were not driven by within-run correlations alone. (A) Regions where the representation for the conjunction was different from the sum of its parts (p < 0.05, whole-brain FWE-corrected). These regions were very similar to the regions reported in our primary Conjunction Contrast (Figure 2.6A) that computed both within- and across-run correlations. (B) The strength of conjunctive coding in the LOC, FFA, PPA and PRC ROIs (shown at bottom are
ROIs from a representative participant superimposed on that individual’s structural scan. These results are very similar to those shown in Figure 2.6B.

Appendix 2. Peak ROI coordinates for all participants, as defined by our functional localizer for the lateral occipital cortex (LOC), fusiform face area (FFA) and parahippocampal place area (PPA).

<table>
<thead>
<tr>
<th>Participant</th>
<th>LOC Left</th>
<th>LOC Right</th>
<th>FFA Left</th>
<th>FFA Right</th>
<th>PPA Left</th>
<th>PPA Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>S03</td>
<td>-48, -76, -4</td>
<td>46, -72, -6</td>
<td>-40, -70, -8</td>
<td>52, -66, -20</td>
<td>-28, -52, -6</td>
<td>30, -44, -8</td>
</tr>
<tr>
<td>S04</td>
<td>-50, -80, 0</td>
<td>48, -86, -2</td>
<td>-46, -58, -16</td>
<td>40, -52, -16</td>
<td>-30, -42, -10</td>
<td>26, -40, -12</td>
</tr>
<tr>
<td>S05</td>
<td>-40, -88, -8</td>
<td>50, -74, -2</td>
<td>-38, -50, -20</td>
<td>46, -56, -24</td>
<td>-32, -50, -8</td>
<td>32, -36, -18</td>
</tr>
<tr>
<td>S06</td>
<td>-48, -82, -8</td>
<td>52, -74, -14</td>
<td>-44, -62, -16</td>
<td>52, -78, -6</td>
<td>-30, -46, -10</td>
<td>32, -42, -12</td>
</tr>
<tr>
<td>S07</td>
<td>-40, -80, -8</td>
<td>48, -82, -4</td>
<td>-40, -66, -16</td>
<td>40, -48, -26</td>
<td>-26, -44, -12</td>
<td>24, -44, -10</td>
</tr>
<tr>
<td>S08</td>
<td>-46, -78, -10</td>
<td>48, -74, 0</td>
<td>-56, -66, -6</td>
<td>58, -66, 4</td>
<td>-24, -54, -6</td>
<td>28, -46, -6</td>
</tr>
<tr>
<td>S09</td>
<td>-54, -72, -4</td>
<td>50, -72, -4</td>
<td>----</td>
<td>44, -46, -20</td>
<td>-28, -46, -10</td>
<td>24, -48, -10</td>
</tr>
<tr>
<td>S10</td>
<td>-40, -80, -8</td>
<td>46, -80, -10</td>
<td>-40, -58, -20</td>
<td>40, -52, -20</td>
<td>-28, -48, -8</td>
<td>28, -42, -12</td>
</tr>
<tr>
<td>S11</td>
<td>-44, -76, -2</td>
<td>56, -66, 10</td>
<td>-40, -50, -12</td>
<td>42, -32, -20</td>
<td>-24, -44, -12</td>
<td>24, -32, -18</td>
</tr>
<tr>
<td>S12</td>
<td>-46, -80, -6</td>
<td>50, -72, -14</td>
<td>-38, -34, -14</td>
<td>42, -40, -16</td>
<td>-26, -46, -10</td>
<td>28, -42, -12</td>
</tr>
<tr>
<td>S14</td>
<td>-48, -84, -6</td>
<td>46, -78, 0</td>
<td>-36, -50, -24</td>
<td>42, -50, -22</td>
<td>-24, -46, -6</td>
<td>24, -42, -10</td>
</tr>
<tr>
<td>S15</td>
<td>-50, -82, -6</td>
<td>42, -82, -8</td>
<td>-42, -48, -26</td>
<td>40, -46, -18</td>
<td>-26, -46, -8</td>
<td>26, -52, -8</td>
</tr>
<tr>
<td>S16</td>
<td>-40, -84, -8</td>
<td>50, -68, -8</td>
<td>----</td>
<td>40, -44, -18</td>
<td>-32, -42, -8</td>
<td>34, -42, -8</td>
</tr>
<tr>
<td>S17</td>
<td>-40, -90, -14</td>
<td>46, -80, -8</td>
<td>46, -54, -14</td>
<td>46, -52, -18</td>
<td>-28, -50, -6</td>
<td>28, -44, -8</td>
</tr>
<tr>
<td>S20</td>
<td>-46, -76, -8</td>
<td>44, -80, -4</td>
<td>----</td>
<td>14, -48, -4</td>
<td>-26, -46, -12</td>
<td>24, -40, -14</td>
</tr>
<tr>
<td>S21</td>
<td>-44, -82, -2</td>
<td>48, -70, 0</td>
<td>-40, -60, -14</td>
<td>46, -60, -16</td>
<td>-26, -46, -8</td>
<td>32, -40, -10</td>
</tr>
<tr>
<td>S22</td>
<td>-46, -66, -16</td>
<td>52, -66, -6</td>
<td>-56, -68, 8</td>
<td>64, -56, 8</td>
<td>-26, -44, -6</td>
<td>32, -38, -12</td>
</tr>
</tbody>
</table>
For all tables, when more than one suprathreshold voxel appeared for a given cluster within a given brain region, the coordinate with the highest Z-score is reported. Multiple coordinates are reported if they are separated by a Euclidean distance of more than 12 mm within a cluster. If a cluster included both the left and right hemispheres, it was assigned to the hemisphere containing the peak voxel. Clusters are ordered anterior to posterior, according to the peak voxels of each cluster.

Appendix 3A. Regions that showed an effect of feature type in the univariate ANOVA (Figure 2.5B) for feature set 1 (p < 0.001, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>P-value (uncorrected)</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>20</td>
<td>7</td>
<td>-48, 0, -32</td>
<td>&lt;0.0005</td>
<td>3.44</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>20</td>
<td>6</td>
<td>-40, -32, -18</td>
<td>0.001</td>
<td>3.29</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>1</td>
<td>-32, -36, -22</td>
<td>0.001</td>
<td>3.14</td>
</tr>
<tr>
<td>Posterior cingulate cortex</td>
<td>26</td>
<td>5</td>
<td>-8, -40, 28</td>
<td>&lt;0.0005</td>
<td>3.4</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>101</td>
<td>-34, -54, -18</td>
<td>0.001</td>
<td>5.18</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>2</td>
<td>-34, -60, -8</td>
<td>0.001</td>
<td>3.27</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>3</td>
<td>-34, -64, -8</td>
<td>&lt;0.0005</td>
<td>3.31</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>18</td>
<td>1409</td>
<td>-14, -86, -14</td>
<td>&lt;0.0005</td>
<td>7.22</td>
</tr>
<tr>
<td>Middle occipital gyrus</td>
<td>18</td>
<td></td>
<td>-24, -92, 0</td>
<td>&lt;0.0005</td>
<td>6</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>19</td>
<td></td>
<td>26, -66, -4</td>
<td>&lt;0.0005</td>
<td>4.37</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>17</td>
<td></td>
<td>0, -80, 2</td>
<td>&lt;0.0005</td>
<td>4.03</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post central gyrus</td>
<td>3</td>
<td>1</td>
<td>46, -16, 36</td>
<td>&lt;0.0005</td>
<td>3.32</td>
</tr>
<tr>
<td>Thalamus</td>
<td>-</td>
<td>7</td>
<td>22, -18, 0</td>
<td>&lt;0.0005</td>
<td>3.54</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>27</td>
<td>11</td>
<td>8, -30, -4</td>
<td>&lt;0.0005</td>
<td>3.81</td>
</tr>
<tr>
<td>Parahippocampal gyrus</td>
<td>27</td>
<td>12</td>
<td>22, -34, -4</td>
<td>&lt;0.0005</td>
<td>3.92</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>2</td>
<td>36, -44, -22</td>
<td>0.001</td>
<td>3.11</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>20</td>
<td>4</td>
<td>48, -46, -16</td>
<td>&lt;0.0005</td>
<td>3.52</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>17</td>
<td>34, -50, -20</td>
<td>&lt;0.0005</td>
<td>3.94</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>1</td>
<td>16, -62, -24</td>
<td>0.001</td>
<td>3.17</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>19</td>
<td>2</td>
<td>28, -74, -10</td>
<td>0.001</td>
<td>3.14</td>
</tr>
<tr>
<td>Superior occipital gyrus</td>
<td>19</td>
<td>25</td>
<td>28, -78, 18</td>
<td>&lt;0.0005</td>
<td>3.84</td>
</tr>
<tr>
<td>Superior occipital gyrus</td>
<td>18</td>
<td>57</td>
<td>22, -88, 12</td>
<td>&lt;0.0005</td>
<td>4.04</td>
</tr>
<tr>
<td>Inferior occipital gyrus</td>
<td>18</td>
<td>257</td>
<td>32, -92, -4</td>
<td>&lt;0.0005</td>
<td>6.77</td>
</tr>
<tr>
<td>Middle occipital gyrus</td>
<td>19</td>
<td></td>
<td>44, -78, 4</td>
<td>&lt;0.0005</td>
<td>3.71</td>
</tr>
</tbody>
</table>
Appendix 3B. Regions that showed an effect of feature type in the univariate ANOVA (Figure 2.5B) for feature set 2 (p < 0.001, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>P-value (uncorrected)</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior frontal gyrus</td>
<td>48</td>
<td>1</td>
<td>-58, 18, 4</td>
<td>0.001</td>
<td>3.25</td>
</tr>
<tr>
<td>Olfactory cortex</td>
<td>48</td>
<td>3</td>
<td>-22, 6, -14</td>
<td>0.001</td>
<td>3.18</td>
</tr>
<tr>
<td>Middle cingulate cortex</td>
<td>23</td>
<td>6</td>
<td>-12, -26, 44</td>
<td>&lt;0.0005</td>
<td>3.37</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>19</td>
<td>-10, -32, -24</td>
<td>&lt;0.0005</td>
<td>4.02</td>
</tr>
<tr>
<td>Middle temporal gyrus</td>
<td>21</td>
<td>11</td>
<td>-60, -32, 6</td>
<td>&lt;0.0005</td>
<td>3.37</td>
</tr>
<tr>
<td>Middle temporal gyrus</td>
<td>22</td>
<td>8</td>
<td>-58, -34, 44</td>
<td>&lt;0.0005</td>
<td>3.36</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>18</td>
<td>26</td>
<td>-18, -70, 2</td>
<td>&lt;0.0005</td>
<td>4.15</td>
</tr>
<tr>
<td>Middle occipital gyrus</td>
<td>18</td>
<td>119</td>
<td>-22, -90, 20</td>
<td>&lt;0.0005</td>
<td>5.46</td>
</tr>
<tr>
<td>Inferior occipital gyrus</td>
<td>18</td>
<td>1001</td>
<td>-28, -92, -4</td>
<td>&lt;0.0005</td>
<td>5.86</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>17</td>
<td>4</td>
<td>-4, -78, -2</td>
<td>&lt;0.0005</td>
<td>5.48</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>18</td>
<td>-14, -86, -16</td>
<td>&lt;0.0005</td>
<td></td>
<td>5.43</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>19</td>
<td>-36, -82, -16</td>
<td>&lt;0.0005</td>
<td></td>
<td>4.22</td>
</tr>
<tr>
<td>Calcarine gyrus</td>
<td>17</td>
<td>0, -70, 10</td>
<td>&lt;0.0005</td>
<td></td>
<td>3.66</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>20</td>
<td>1</td>
<td>32, 10, -42</td>
<td>0.001</td>
<td>3.2</td>
</tr>
<tr>
<td>Thalamus</td>
<td>-</td>
<td>2</td>
<td>12, -12, 0</td>
<td>0.001</td>
<td>3.15</td>
</tr>
<tr>
<td>Middle cingulate cortex</td>
<td>23</td>
<td>5</td>
<td>14, -24, 42</td>
<td>&lt;0.0005</td>
<td>3.44</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>55</td>
<td>4, -30, -46</td>
<td>&lt;0.0005</td>
<td>4.01</td>
</tr>
<tr>
<td>Precuneus</td>
<td>17</td>
<td>1</td>
<td>20, -38, 22</td>
<td>0.001</td>
<td>3.1</td>
</tr>
<tr>
<td>Precuneus</td>
<td>27</td>
<td>7</td>
<td>16, -42, 4</td>
<td>&lt;0.0005</td>
<td>3.41</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>2</td>
<td>2</td>
<td>26, -42, 58</td>
<td>&lt;0.0005</td>
<td>3.31</td>
</tr>
<tr>
<td>Middle temporal gyrus</td>
<td>22</td>
<td>151</td>
<td>60, -44, 2</td>
<td>&lt;0.0005</td>
<td>5.1</td>
</tr>
<tr>
<td>Precuneus</td>
<td>23</td>
<td>23</td>
<td>14, -58, 22</td>
<td>&lt;0.0005</td>
<td>3.5</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>37</td>
<td>4</td>
<td>60, -58, -2</td>
<td>0.001</td>
<td>3.28</td>
</tr>
<tr>
<td>Precuneus</td>
<td>23</td>
<td>1</td>
<td>8, -66, 26</td>
<td>&lt;0.0005</td>
<td>3.37</td>
</tr>
<tr>
<td>Calcarine gyrus</td>
<td>17</td>
<td>88</td>
<td>18, -70, 10</td>
<td>&lt;0.0005</td>
<td>4.29</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>19</td>
<td>22, -66, -2</td>
<td>&lt;0.0005</td>
<td></td>
<td>3.38</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>18</td>
<td>24, -74, -8</td>
<td>&lt;0.0005</td>
<td></td>
<td>3.38</td>
</tr>
<tr>
<td>Precuneus</td>
<td>7</td>
<td>51</td>
<td>2, -76, 40</td>
<td>&lt;0.0005</td>
<td>3.68</td>
</tr>
<tr>
<td>Inferior occipital gyrus</td>
<td>19</td>
<td>1</td>
<td>44, -76, -4</td>
<td>0.001</td>
<td>3.18</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>1</td>
<td>20, -78, -18</td>
<td>0.001</td>
<td>3.12</td>
</tr>
<tr>
<td>Superior occipital gyrus</td>
<td>18</td>
<td>47</td>
<td>26, -90, 22</td>
<td>&lt;0.0005</td>
<td>3.92</td>
</tr>
<tr>
<td>Calcarine gyrus</td>
<td>18</td>
<td>8</td>
<td>12, -94, 12</td>
<td>&lt;0.0005</td>
<td>3.72</td>
</tr>
</tbody>
</table>
Appendix 4. Conjunction Contrast: Searchlight MVPA. Regions where the representation for the conjunction was different from the sum of its parts, shown at a more liberal threshold of $p < 0.001$ (uncorrected). See Figure 2.6A for whole-brain FWE-corrected results and Appendix 5 for a list of suprathreshold voxels.
**Appendix 5A.** Regions identified by the Conjunction Contrast in the searchlight MVPA (whole-brain FWE-corrected, p < 0.05).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>P-value (FWE)</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior orbitofrontal gyrus</td>
<td>47</td>
<td>16</td>
<td>-40, 28, -10</td>
<td>0.029</td>
<td>4.54</td>
</tr>
<tr>
<td>Olfactory cortex</td>
<td>11</td>
<td>302</td>
<td>-2, 22, -10</td>
<td>0.003</td>
<td>5.14</td>
</tr>
<tr>
<td>Superior orbital gyrus</td>
<td>11</td>
<td></td>
<td>-14, 24, -22</td>
<td>0.004</td>
<td>5.03</td>
</tr>
<tr>
<td>Insula</td>
<td>38</td>
<td></td>
<td>-28, 12, -20</td>
<td>0.016</td>
<td>4.70</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>20</td>
<td>121</td>
<td>-56, -18, -30</td>
<td>0.004</td>
<td>5.07</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>19</td>
<td>112</td>
<td>-26, -60, -10</td>
<td>0.013</td>
<td>4.76</td>
</tr>
<tr>
<td>Inferior occipital gyrus</td>
<td>18</td>
<td>800</td>
<td>-18, -98, -6</td>
<td>0.010</td>
<td>4.83</td>
</tr>
<tr>
<td>Superior occipital gyrus</td>
<td>17</td>
<td></td>
<td>-12, -98, 10</td>
<td>0.012</td>
<td>4.77</td>
</tr>
<tr>
<td>Inferior occipital gyrus</td>
<td>19</td>
<td></td>
<td>-38, -78, -10</td>
<td>0.015</td>
<td>4.71</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perirhinal cortex</td>
<td>36</td>
<td>214</td>
<td>30, -6, -30</td>
<td>0.001</td>
<td>5.47</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>19</td>
<td>2214</td>
<td>32, -80, -4</td>
<td>&lt;0.0005</td>
<td>5.93</td>
</tr>
<tr>
<td>Cuneus</td>
<td>17</td>
<td>41</td>
<td>2, -82, 16</td>
<td>0.026</td>
<td>4.57</td>
</tr>
</tbody>
</table>
Appendix 5B. Regions identified by the Conjunction Contrast in the searchlight MVPA (p < 0.001, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>P-value (uncorrected)</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior frontal gyrus</td>
<td>44</td>
<td>18</td>
<td>-62, 6, 14</td>
<td>0.001</td>
<td>3.26</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>3</td>
<td>333</td>
<td>-50, -16, 34</td>
<td>&lt;0.0005</td>
<td>4.03</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>43</td>
<td></td>
<td>-64, -8, 26</td>
<td>&lt;0.0005</td>
<td>3.31</td>
</tr>
<tr>
<td>Superior temporal gyrus</td>
<td>48</td>
<td>1</td>
<td>-54, -32, 24</td>
<td>0.001</td>
<td>3.09</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior orbital gyrus</td>
<td>11</td>
<td>4</td>
<td>22, 40, -8</td>
<td>0.001</td>
<td>3.15</td>
</tr>
<tr>
<td>Superior temporal gyrus</td>
<td>22</td>
<td>26</td>
<td>62, -18, 6</td>
<td>0.001</td>
<td>3.18</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>20</td>
<td>357</td>
<td>58, -22, -20</td>
<td>&lt;0.0005</td>
<td>4.10</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>3</td>
<td>245</td>
<td>40, -26, 36</td>
<td>&lt;0.0005</td>
<td>3.98</td>
</tr>
<tr>
<td>Middle temporal gyrus</td>
<td>21</td>
<td>347</td>
<td>52, -48, 10</td>
<td>&lt;0.0005</td>
<td>3.54</td>
</tr>
<tr>
<td>Superior parietal lobule</td>
<td>7</td>
<td>757</td>
<td>22, -60, 54</td>
<td>&lt;0.0005</td>
<td>3.64</td>
</tr>
<tr>
<td>Precuneus</td>
<td>5/7</td>
<td></td>
<td>16, -52, 52</td>
<td>&lt;0.0005</td>
<td>3.50</td>
</tr>
<tr>
<td>Inferior parietal lobule</td>
<td>40</td>
<td></td>
<td>40, -52, 48</td>
<td>&lt;0.0005</td>
<td>3.46</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>19</td>
<td>38459</td>
<td>32, -80, -4</td>
<td>&lt;0.0005</td>
<td>5.93</td>
</tr>
<tr>
<td>Perirhinal cortex</td>
<td>36</td>
<td></td>
<td>30, -6, -30</td>
<td>&lt;0.0005</td>
<td>5.47</td>
</tr>
<tr>
<td>Olfactory cortex</td>
<td>11</td>
<td></td>
<td>-2, 22, -10</td>
<td>&lt;0.0005</td>
<td>5.14</td>
</tr>
<tr>
<td>Inferior Temporal gyrus</td>
<td>20</td>
<td></td>
<td>-56, -18, -30</td>
<td>&lt;0.0005</td>
<td>5.07</td>
</tr>
<tr>
<td>Superior orbitofrontal gyrus</td>
<td>11</td>
<td></td>
<td>-14, 24, -22</td>
<td>&lt;0.0005</td>
<td>5.03</td>
</tr>
<tr>
<td>Inferior occipital gyrus</td>
<td>18</td>
<td></td>
<td>-18, -98, -6</td>
<td>&lt;0.0005</td>
<td>4.83</td>
</tr>
<tr>
<td>Superior occipital gyrus</td>
<td>17</td>
<td></td>
<td>-12, -98, 10</td>
<td>&lt;0.0005</td>
<td>4.77</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>19</td>
<td></td>
<td>-26, -60, -10</td>
<td>&lt;0.0005</td>
<td>4.76</td>
</tr>
<tr>
<td>Inferior occipital gyrus</td>
<td>19</td>
<td></td>
<td>-38, -78, -10</td>
<td>&lt;0.0005</td>
<td>4.71</td>
</tr>
<tr>
<td>Insula</td>
<td>38</td>
<td></td>
<td>-28, 12, -20</td>
<td>&lt;0.0005</td>
<td>4.70</td>
</tr>
<tr>
<td>Cuneus / Calcarine sulcus</td>
<td>17/18</td>
<td></td>
<td>2, -82, 16</td>
<td>&lt;0.0005</td>
<td>4.57</td>
</tr>
<tr>
<td>Inferior orbitofrontal cortex</td>
<td>47</td>
<td></td>
<td>-40, 28, -10</td>
<td>&lt;0.0005</td>
<td>4.54</td>
</tr>
<tr>
<td>Thalamus</td>
<td>-</td>
<td></td>
<td>18, -8, 2</td>
<td>&lt;0.0005</td>
<td>4.32</td>
</tr>
</tbody>
</table>
**Appendix 6A.** Regions identified by Viewpoint Contrast in the searchlight MVPA (whole-brain FWE-corrected, p < 0.05).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>P-value (FWE)</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal gyrus</td>
<td>25</td>
<td>2</td>
<td>-12, 18, -14</td>
<td>0.043</td>
<td>4.41</td>
</tr>
</tbody>
</table>

**Appendix 6B.** Regions identified by Viewpoint Contrast in the searchlight MVPA (p < 0.001, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>P-value (uncorrected)</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior frontal gyrus</td>
<td>47</td>
<td>166</td>
<td>-40, 38, -16</td>
<td>&lt;0.0005</td>
<td>3.57</td>
</tr>
<tr>
<td>Rectal gyrus</td>
<td>25</td>
<td>378</td>
<td>-12, 18, -14</td>
<td>&lt;0.0005</td>
<td>4.41</td>
</tr>
<tr>
<td>Middle temporal gyrus</td>
<td>21</td>
<td>26</td>
<td>-58, 4, -24</td>
<td>&lt;0.0005</td>
<td>3.34</td>
</tr>
<tr>
<td>Perirhinal cortex</td>
<td>36</td>
<td>38</td>
<td>-36, -4, -26</td>
<td>0.001</td>
<td>3.25</td>
</tr>
<tr>
<td>Middle cingulate cortex</td>
<td>23</td>
<td>14</td>
<td>-12, -24, 30</td>
<td>0.001</td>
<td>3.24</td>
</tr>
<tr>
<td>Insula</td>
<td>48</td>
<td>35</td>
<td>-28, -26, 18</td>
<td>&lt;0.0005</td>
<td>3.78</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>20</td>
<td>601</td>
<td>-52, -28, -20</td>
<td>&lt;0.0005</td>
<td>4.15</td>
</tr>
<tr>
<td>Posterior cingulate cortex</td>
<td>23</td>
<td>159</td>
<td>-14, -50, 22</td>
<td>&lt;0.0005</td>
<td>3.63</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>120</td>
<td>-36, -56, -18</td>
<td>&lt;0.0005</td>
<td>3.71</td>
</tr>
<tr>
<td>Cuneus</td>
<td>19</td>
<td>2</td>
<td>-16, -72, 36</td>
<td>0.001</td>
<td>3.11</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insula</td>
<td>48</td>
<td>3</td>
<td>26, -22, 36</td>
<td>0.001</td>
<td>3.12</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>19</td>
<td>393</td>
<td>32, -70, -4</td>
<td>&lt;0.0005</td>
<td>4.07</td>
</tr>
</tbody>
</table>
Appendix 7. Regions demonstrating view-invariant explicit conjunctive coding: analyses computing correlations across runs only (i.e., correlations within runs did not contribute to the analysis). These analyses revealed the same pattern of results as our main findings (Figure 3.3), indicating that our results were not driven by within-run correlations alone. (A) Regions demonstrating view-invariant conjunctive representations (displayed at $p < 0.005$, uncorrected; no VVS regions survived whole-brain FWE correction). These regions were almost entirely overlapping with those reported in our primary Viewpoint Contrast (Figure 3.3A) that computed both within- and across-run correlations. The only difference was that here activity in the anterior temporal lobe survived a correction of $p < 0.002$, whereas when we computed correlations both within and across runs it survived $p < 0.001$. This difference may be due to the
reduction in power caused by moving comparisons further apart in time, and thus admitting greater low frequency noise. (B) The strength of view-invariant conjunctive coding within and across ROIs. These results are very similar to the results shown in Figure 3.3B. Error bars indicate s.e.m.; *p < 0.05 for comparisons across ROIs (paired-samples t-tests indicated by brackets) and comparisons relative to zero within an ROI (one-sample t-tests indicated by an asterisk above a bar).
Appendix 8. Regions identified by the *Conjunctive Coding for Novel Objects* contrast in the searchlight MVPA (p < 0.001, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior orbitofrontal gyrus</td>
<td>11</td>
<td>98</td>
<td>-16 14 -18</td>
<td>3.44</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>20</td>
<td>363</td>
<td>-46 6 -36</td>
<td>4.25</td>
</tr>
<tr>
<td>Middle cingulate cortex</td>
<td>23</td>
<td>2875</td>
<td>0 -34 38</td>
<td>4.24</td>
</tr>
<tr>
<td>Thalamus</td>
<td>-</td>
<td></td>
<td>8 -10 12</td>
<td>3.85</td>
</tr>
<tr>
<td>Precuneus</td>
<td>29</td>
<td></td>
<td>6 -44 12</td>
<td>3.72</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>21</td>
<td>-10 -34 -40</td>
<td>3.19</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>365</td>
<td>-18 -52 -44</td>
<td>4.56</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td></td>
<td>-32 -38 -38</td>
<td>4</td>
</tr>
<tr>
<td>Precuneus</td>
<td>5</td>
<td>17</td>
<td>-14 -60 56</td>
<td>3.34</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>66</td>
<td>-8 -64 -14</td>
<td>3.3</td>
</tr>
<tr>
<td>Middle occipital gyrus</td>
<td>18</td>
<td>2053</td>
<td>-16 -104 8</td>
<td>4.38</td>
</tr>
<tr>
<td>Middle occipital gyrus</td>
<td>18</td>
<td></td>
<td>-38 -84 8</td>
<td>4.04</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>18</td>
<td></td>
<td>-20 -86 -6</td>
<td>3.93</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caudate</td>
<td>48</td>
<td>1</td>
<td>24 26 6</td>
<td>3.1</td>
</tr>
<tr>
<td>Putamen</td>
<td>48</td>
<td>482</td>
<td>24 16 -10</td>
<td>3.79</td>
</tr>
<tr>
<td>Rectal gyrus</td>
<td>11</td>
<td></td>
<td>12 34 -16</td>
<td>3.7</td>
</tr>
<tr>
<td>Superior orbitofrontal gyrus</td>
<td>11</td>
<td></td>
<td>20 36 -8</td>
<td>3.67</td>
</tr>
<tr>
<td>Middle temporal pole</td>
<td>48</td>
<td>95</td>
<td>38 12 -36</td>
<td>3.32</td>
</tr>
<tr>
<td>Inferotemporal cortex</td>
<td>20</td>
<td></td>
<td>42 -2 -38</td>
<td>3.25</td>
</tr>
<tr>
<td>Caudate</td>
<td>-</td>
<td>10</td>
<td>28 -10 32</td>
<td>3.5</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>38</td>
<td>16 -22 -34</td>
<td>3.42</td>
</tr>
<tr>
<td>Parahippocampal gyrus</td>
<td>36</td>
<td></td>
<td>20 -14 -34</td>
<td>3.41</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>20</td>
<td>102</td>
<td>54 -24 -28</td>
<td>3.47</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>41</td>
<td>20 -52 -40</td>
<td>3.42</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>37</td>
<td>1</td>
<td>58 -54 -2</td>
<td>3.1</td>
</tr>
<tr>
<td>Middle occipital gyrus</td>
<td>39</td>
<td>2061</td>
<td>34 -72 16</td>
<td>4.38</td>
</tr>
<tr>
<td>Middle occipital gyrus</td>
<td>19</td>
<td></td>
<td>28 -84 16</td>
<td>4.25</td>
</tr>
<tr>
<td>Inferior occipital cortex</td>
<td>18</td>
<td></td>
<td>24 -96 -2</td>
<td>3.43</td>
</tr>
</tbody>
</table>
**Appendix 9.** Regions identified by the *Conjunctive Coding for Trained Objects* contrast in the searchlight MVPA (p < 0.001, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal gyrus</td>
<td>11</td>
<td>28</td>
<td>-6 -40 -16</td>
<td>3.17</td>
</tr>
<tr>
<td>Caudate</td>
<td>25</td>
<td>182</td>
<td>-6 -18 -4</td>
<td>3.64</td>
</tr>
<tr>
<td>Inferior temporal cortex</td>
<td>20</td>
<td>283</td>
<td>-42 -4 -36</td>
<td>4.14</td>
</tr>
<tr>
<td>Superior temporal gyrus</td>
<td>42</td>
<td>334</td>
<td>-60 -42 20</td>
<td>3.51</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>10595</td>
<td>-32 -46 -14</td>
<td>5.04</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>19</td>
<td></td>
<td>-30 -66 0</td>
<td>4.86</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>18</td>
<td></td>
<td>-18 -70-4</td>
<td>4.58</td>
</tr>
<tr>
<td>Angular gyrus</td>
<td>40</td>
<td>15</td>
<td>-34 -56 36</td>
<td>3.3</td>
</tr>
<tr>
<td>Superior occipital gyrus</td>
<td>7</td>
<td>31</td>
<td>-12 -80 46</td>
<td>3.6</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle temporal pole</td>
<td>38</td>
<td>1</td>
<td>48 16 -34</td>
<td>3.09</td>
</tr>
<tr>
<td>Inferior frontal gyrus</td>
<td>48</td>
<td>52</td>
<td>40 10 18</td>
<td>3.82</td>
</tr>
<tr>
<td>Insula</td>
<td>48</td>
<td>210</td>
<td>46 2 2</td>
<td>3.37</td>
</tr>
<tr>
<td>Insula</td>
<td>48</td>
<td></td>
<td>42 10 -8</td>
<td>3.24</td>
</tr>
<tr>
<td>Inferior temporal cortex</td>
<td>20</td>
<td>258</td>
<td>58 -10 -36</td>
<td>3.9</td>
</tr>
<tr>
<td>Inferior temporal cortex</td>
<td>20</td>
<td></td>
<td>64 -24 -30</td>
<td>3.76</td>
</tr>
<tr>
<td>Putamen</td>
<td>-</td>
<td>1455</td>
<td>20 -12 6</td>
<td>4.14</td>
</tr>
<tr>
<td>Middle temporal gyrus</td>
<td>21</td>
<td></td>
<td>46 -36 -6</td>
<td>4.1</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>-</td>
<td></td>
<td>36 -34 2</td>
<td>4.03</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>2</td>
<td>205</td>
<td>26 -42 44</td>
<td>3.99</td>
</tr>
</tbody>
</table>
### Appendix 10. Regions identified by the *Conjunctive Coding for Trained > Novel Objects* contrast in the searchlight MVPA (p < 0.001, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>2931</td>
<td>-40 -50 -20</td>
<td>4.83</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>19</td>
<td></td>
<td>-22 -50 -18</td>
<td>4.59</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>20</td>
<td></td>
<td>-42 -26 -30</td>
<td>3.41</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>531</td>
<td>38 -42 -6</td>
<td>3.69</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>37</td>
<td></td>
<td>46 -52 -18</td>
<td>3.32</td>
</tr>
</tbody>
</table>

### Appendix 11. Regions identified by the *Conjunctive Coding for Novel > Trained Objects* contrast in the searchlight MVPA (p < 0.001, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle occipital cortex</td>
<td>17</td>
<td>272</td>
<td>-20 -102 8</td>
<td>3.79</td>
</tr>
</tbody>
</table>

### Appendix 12. Regions identified by the *View-Invariant Conjunctive Coding for Novel Objects* contrast in the searchlight MVPA (p < 0.005, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle temporal pole</td>
<td>36</td>
<td>3</td>
<td>-22 12 -38</td>
<td>2.76</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>3</td>
<td>-38 -40 -24</td>
<td>2.60</td>
</tr>
<tr>
<td>Middle occipital gyrus</td>
<td>18</td>
<td>1013</td>
<td>-26 -84 0</td>
<td>3.54</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>19</td>
<td></td>
<td>-28 -68 0</td>
<td>3.11</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle temporal pole</td>
<td>20</td>
<td>112</td>
<td>36 8 -36</td>
<td>2.93</td>
</tr>
<tr>
<td>Caudate</td>
<td>25</td>
<td>16</td>
<td>26 -26 8</td>
<td>2.81</td>
</tr>
<tr>
<td>Inferior occipital cortex</td>
<td>8</td>
<td></td>
<td>32 -92 -6</td>
<td>2.61</td>
</tr>
</tbody>
</table>
### Appendix 13. Regions identified by the View-Invariant Conjunctive Coding for Trained Objects contrast in the searchlight MVPA (p < 0.005, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior orbitofrontal cortex</td>
<td>47</td>
<td>26</td>
<td>-40 34 -20</td>
<td>2.87</td>
</tr>
<tr>
<td>Inferior frontal cortex</td>
<td>44</td>
<td>15</td>
<td>-58 18 10</td>
<td>2.9</td>
</tr>
<tr>
<td>Caudate</td>
<td>-</td>
<td>132</td>
<td>-14 10 18</td>
<td>3.64</td>
</tr>
<tr>
<td>Superior temporal pole</td>
<td>48</td>
<td>735</td>
<td>-46 8 -14</td>
<td>3.51</td>
</tr>
<tr>
<td>Caudate</td>
<td>-</td>
<td>80</td>
<td>-22 -14 34</td>
<td>3.21</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>48</td>
<td>6</td>
<td>-60 -14 20</td>
<td>2.63</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>1263</td>
<td>-38 -32 -30</td>
<td>3.31</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td></td>
<td>-14 -44 -22</td>
<td>3.3</td>
</tr>
<tr>
<td>Caudate</td>
<td>-</td>
<td></td>
<td>-22 -14 34</td>
<td>2.95</td>
</tr>
<tr>
<td>Inferior parietal cortex</td>
<td>40</td>
<td>83</td>
<td>-54 -52 38</td>
<td>2.98</td>
</tr>
<tr>
<td>Precuneus</td>
<td>7</td>
<td>1</td>
<td>-12 -52 38</td>
<td>2.58</td>
</tr>
<tr>
<td>Cuneus</td>
<td>23</td>
<td>26</td>
<td>-16 -60 22</td>
<td>2.85</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>19</td>
<td>7</td>
<td>-30 -84 -16</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior orbitofrontal cortex</td>
<td>11</td>
<td>2</td>
<td>22 28 -24</td>
<td>2.59</td>
</tr>
<tr>
<td>Superior orbitofrontal cortex</td>
<td>11</td>
<td>1</td>
<td>18 28 -26</td>
<td>2.6</td>
</tr>
<tr>
<td>Pallidum</td>
<td>--</td>
<td>1017</td>
<td>22 -6 4</td>
<td>3.76</td>
</tr>
<tr>
<td>Inferior temporal cortex</td>
<td>20</td>
<td>44</td>
<td>56 -20 -36</td>
<td>2.88</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>64</td>
<td>30 -44 -26</td>
<td>2.76</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td>49</td>
<td>38 -50 -4</td>
<td>2.89</td>
</tr>
<tr>
<td>Precuneus</td>
<td>19</td>
<td></td>
<td>32 -50 10</td>
<td>2.63</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>2</td>
<td>24 -68 -26</td>
<td>2.59</td>
</tr>
<tr>
<td>Calcarine gyrus</td>
<td>18</td>
<td>495</td>
<td>20 -96 -2</td>
<td>3.85</td>
</tr>
<tr>
<td>Calcarine gyrus</td>
<td>17</td>
<td></td>
<td>4 -94 -6</td>
<td>3.12</td>
</tr>
<tr>
<td>Middle occipital gyrus</td>
<td>17</td>
<td></td>
<td>-10 -96 0</td>
<td>2.62</td>
</tr>
</tbody>
</table>
**Appendix 14.** Regions identified by the *View-Invariant Conjunctive Coding for Trained > Novel Objects* contrast in the searchlight MVPA (p < 0.005, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior orbitofrontal cortex</td>
<td>38</td>
<td>806</td>
<td>-46 -28 -12</td>
<td>4.50</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>48</td>
<td>20</td>
<td>-66 -16 -22</td>
<td>2.82</td>
</tr>
<tr>
<td>Middle cingulate gyrus</td>
<td>23</td>
<td>266</td>
<td>0 -26 -46</td>
<td>3.59</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>5044</td>
<td>-42 -50 -36</td>
<td>4.33</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>37</td>
<td></td>
<td>-38 -52 -12</td>
<td>4.27</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>20</td>
<td></td>
<td>-36 -24 -32</td>
<td>3.73</td>
</tr>
<tr>
<td>Inferior parietal gyrus</td>
<td>39</td>
<td>98</td>
<td>-52 -60 46</td>
<td>3.01</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thalamus</td>
<td>-</td>
<td>2</td>
<td>18 -12 4</td>
<td>2.65</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td>4162</td>
<td>16 -46 -30</td>
<td>4.52</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-</td>
<td></td>
<td>16 -68 -30</td>
<td>3.85</td>
</tr>
<tr>
<td>Inferior temporal gyrus</td>
<td>37</td>
<td></td>
<td>40 -56 -6</td>
<td>3.53</td>
</tr>
</tbody>
</table>

**Appendix 15.** Regions identified by the *View-Invariant Conjunctive Coding for Novel > Trained Objects* contrast in the searchlight MVPA (p < 0.005, uncorrected).

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>BA</th>
<th>Cluster size (Number of voxels)</th>
<th>MNI coordinates</th>
<th>Z-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle occipital cortex</td>
<td>17</td>
<td>834</td>
<td>-18 -104 4</td>
<td>3.99</td>
</tr>
<tr>
<td><strong>Right Hemisphere:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior occipital cortex</td>
<td>18</td>
<td>121</td>
<td>24 -94 22</td>
<td>3.04</td>
</tr>
</tbody>
</table>
A. Mean Reaction Time by Set Size: One-Featured Objects (Target Present)

B. Mean Reaction Time by Set Size: Two-Featured Objects (Target Present)
E  Mean Reaction Time by Set Size: One-Featured Objects (Target Present)

F  Mean Reaction Time by Set Size: Two-Featured Objects (Target Present)
Appendix 16. Mean reaction times by set size (A-D) across 5 training sessions and (E-H) for trained and untrained objects on the last training session. Vertical bars represent standard error of the mean (SEM). See Appendix 17 for a description of an ANOVA performed with these data.
C. Average Visual Search Slope: One-Featured Objects (Target Absent)

- Day 1: Average slope = 150 ms/item
- Day 2: Average slope = 140 ms/item
- Day 3: Average slope = 130 ms/item
- Day 4: Average slope = 130 ms/item
- Day 5: Average slope = 140 ms/item
- Day 6 (Trained): Average slope = 150 ms/item
- Day 6 (Untrained): Average slope = 140 ms/item

D. Average Visual Search Slope: Two-Featured Objects (Target Absent)

- Day 1: Average slope = 240 ms/item
- Day 2: Average slope = 220 ms/item
- Day 3: Average slope = 210 ms/item
- Day 4: Average slope = 200 ms/item
- Day 5: Average slope = 190 ms/item
- Day 6 (Trained): Average slope = 230 ms/item
- Day 6 (Untrained): Average slope = 210 ms/item
Appendix 17. Mean search slopes by training session, displayed separately for: (A) One-featured objects: target present trials. (B) Two-featured objects: target present trials. (C) One-featured objects: target absent trials. (D) Two-featured objects: target absent trials. Vertical bars represent standard error of the mean (SEM). A 2 (one/two-featured objects) x 2 (target present/absent trials) x 2 (trained/novel objects) repeated-measures analysis of variance (ANOVA) of reaction times (RTs) revealed main effects of number of features [F(1,17) = 94.18, p < .01, η² = .85], target presence [F(1,17) = 92.27, p < .01, η² = .84], and training [F(1,17) = 6.64, p < .05, η² = .28], indicating that participants responded faster to one-featured objects, target present trials, and trained objects. None of the interactions between these factors were significant (all Fs < 4.20, all ps > 0.056).
Appendix 18. Mean search accuracy by training session, displayed separately for: (A) One-featured objects: target present trials. (B) Two-featured objects: target present trials. (C) One-featured objects: target absent trials. (D) Two-featured objects: target absent trials. Vertical bars represent standard error of the mean (SEM). A 2 (one/two-featured objects) × 2 (target present/absent trials) × 2 (trained/novel objects) repeated-measures analysis of variance (ANOVA) of accuracy percentage revealed main effects of number of features [F(1, 17) = 22.94, p < .01, η² = .57], indicating that participants were more accurate on one-featured objects trials than two-featured object trials. In addition, there was a significant three-way interaction between
the three factors \([F(1,17) = 6.56, \ p < .05, \ \eta^2 = .28]\). Subsequent comparisons revealed that this interaction reflected a two-way interaction between target (present/absent) and training (trained/novel objects) for two-featured objects \([F(1,17) = 4.63, \ p < .05, \ \eta^2 = .21]\), but not one-featured objects \([F(1,17) = .68, \ p > .05]\).
Copyright Acknowledgements

Chapters 2 and 3 in this thesis were adapted from the following article:


© Oxford University Press; the use of this article is by permission of the copyright holder.