

Role of recurrent computations in object completion

A DISSERTATION PRESENTED

BY

HANLIN TANG

TO

THE COMMITTEE ON HIGHER DEGREES IN BIOPHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE SUBJECT OF

BIOPHYSICS

HARVARD UNIVERSITY

CAMBRIDGE, MASSACHUSETTS

DECEMBER 2015

©2015 -- HANLIN TANG
ALL RIGHTS RESERVED.

Role of recurrent computations in object completion

ABSTRACT

Existing models of visual object recognition posit that recognition is orchestrated by a hierarchy of processing layers. In these models, neural computation proceeds in a largely feed-forward path up this hierarchy, without substantial feedback or recurrent processing. These feed-forward models provide a parsimonious account of experimental data, and have given rise to deep convolutional networks in computer vision that outperform previous approaches to object recognition. In this dissertation, we challenge these feed-forward theories by considering the problem of occlusion.

In natural vision, objects are often partially visible, either due to occlusion, limited viewing angles, or poor illumination. The vast majority of previous neurophysiological studies focus on the completion of simple contours, geometric shapes, or line drawings. These studies typically contrast neural responses to occluded objects against responses to unrecognizable scrambled counterparts, thus confounding object completion mechanisms with neural activity linked to perceptual awareness. We set out to provide conceptual advances by using naturalistic objects and measuring the selectivity and tolerance of neural responses when objects are only partially visible.

While we know that feedback and recurrent connections are prevalent throughout visual cortex, their underlying roles are unclear. We present three lines of evidence for the role of recurrence in recognition of occluded objects. We first recorded intracranial field potentials from electrodes surgically implanted in epilepsy patients and measured neural responses to whole and occluded objects. Responses along the ventral visual stream remained selective despite heavy occlusion. However, these visually selective signals emerged ~ 100 ms later than responses to whole objects. The processing delays were particularly pronounced in higher visual areas within the ventral stream, suggesting the involvement of additional recurrent processing. Second, we conducted psychophysical experiments to demonstrate that disrupting this recurrence with backward masking after ~ 75 ms significantly impaired recognition of occluded, but not whole, objects. Lastly, we augmented a computational model with recurrence that significantly outperformed existing feed-forward models and matched human performance.

Contents

1	INTRODUCTION	1
1.1	Nomenclature	2
1.2	Other projects	3
1.3	Visual system hierarchy	3
1.4	The challenge of object completion	5
1.5	Neural representation of occluded shapes	8
1.6	From amodal completion to recognition of occluded objects	10
2	NEUROPHYSIOLOGY RECORDINGS	13
2.1	Intracranial electrodes	14
2.2	Electrode localization	16
2.3	Biophysics of intracranial recordings	16
2.4	Comparison with other methods	18
3	DYNAMICS OF OBJECT COMPLETION	21
3.1	Experiment outline	23
3.2	Object selectivity was retained despite occlusion	28
3.3	Delayed responses to partial objects	31
3.4	Population analysis	36
3.5	Discussion	38
3.6	Methods	42
4	BACKWARD MASKING AND OBJECT COMPLETION	48
4.1	Results	51
4.2	Discussion	57
4.3	Methods	58
5	COMPUTATIONAL MODELS OF OBJECT COMPLETION	62
5.1	Performance of feed-forward models in recognizing occluded objects	63

5.2	Beyond feed-forward models	67
5.3	Discussion	73
6	CONCLUSION	80
APPENDIX A DYNAMICS OF COGNITIVE CONTROL		84
A.1	Introduction	85
A.2	Results	87
A.3	Discussion	99
A.4	Methods	109
APPENDIX B NEURAL REPRESENTATIONS OF MEMORABILITY		115
B.1	Methods	116
B.2	Correlations with memorability	119
B.3	Time scales of memorability	123
B.4	Adventures in electrical stimulation	126
B.5	Conclusion	129
APPENDIX C SUPPLEMENTAL INFORMATION		131
REFERENCES		155

Listing of figures

1.1	Example occluded stimuli	6
2.1	Example electrode placement	15
2.2	Electrode localization	17
2.3	Distribution of experimental time	19
3.1	Outline of neurophysiology experiment	24
3.2	Behavioral performance	25
3.3	Example physiological responses from the main experiment	26
3.4	Example physiological responses from the variant experiment	27
3.5	Neural responses remained visually selective despite partial information	30
3.6	Increased latency for object completion	33
3.7	Summary of latency measurements	37
4.1	Outline of psychophysics experiment	49
4.2	Overall behavioral performance	51
4.3	Backward masking disrupts performance	52
4.4	Effect of backward masking was observed in variant experiments	53
4.5	Method for comparing neural and behavioral data	54
4.6	Construction of the phys25 stimulus set	55
4.7	Correlations between neural and psychophysical measures	56
5.1	Feed-forward networks not robust to occlusion	66
5.2	Performance when trained on occluded objects	67
5.3	Correlation between distance and latency	68
5.4	Schematic of recurrent network model	70
5.5	Dynamics of recurrent neural network	72
5.6	Performance of recurrent neural network	73
5.7	Unfolding of recurrent neural network	74

A.1	Experimental task and behavioral performance	88
A.2	Behavioral data for each subject	90
A.3	Example responses from the anterior cingulate cortex	92
A.4	Location of conflict-selective electrodes	93
A.5	Gamma power in frontal cortex correlates with behavior	94
A.6	Responses during self-corrected error trials	96
A.7	Latency comparison across regions	98
A.8	Example electrode in the right dorsolateral Prefrontal Cortex	105
A.9	Example electrode in Orbitofrontal Cortex comparing responses in the theta and gamma Bands	106
A.10	Theta and beta band population results	107
A.11	Cross-frequency coupling analyses	108
B.1	Memorability experiment	117
B.2	Behavioral performance	118
B.3	Summary of recorded units	120
B.4	Example unit in the parahippocampal gyrus	121
B.5	Predicting memorability from neural data	122
B.6	Time scales of memorability	124
B.7	Stimulation paradigm	127
C.1	Example responses in the gamma band	133
C.2	Example responses in inferior temporal gyrus	134
C.3	d' metric, matched amplitude, and matched decoding comparisons	135
C.4	Eye-tracking data and analyses	136
C.5	Detailed summary of latency measurements	137

DEDICATED TO MY PARENTS.

Acknowledgments

The path to a doctorate is laid by the labor of others. My research depends entirely on the goodwill of patients from hospitals around the world: Boston Children's Hospital (BCH), Brigham and Women's Hospital (BWH), John Hopkins Medical Institution (JHMI), Taipei Veterans General Hospital (TVGH), and UCLA Medical Center. I have had the privilege of hearing their stories, fears, and aspirations. These patients come from all corners of the world, and from all walks of life, but are united in their willingness to volunteer their time, despite have undergone brain surgery, for the advancement of neuroscience. For this I am very grateful.

Behind these experiments is a large cast of clinicians who have volunteered their assistance, outside of their clinical duties, to support this research. I thank Sheryl Mangarano (BCH), Lixia Gao (BCH), Paul Dionne (BWH), Jack Connolly (BWH), Karen Walters (JHMI), and Dawn Eliashiv (UCLA). My experiments would not have been possible without the skill of neurosurgeons who performed the surgical implantations: Joseph Madsen (BCH and BWH), William Anderson (JHMI), Frederick Lenz (JHMI), Travis Tierney (BWH), and Itzhak Fried (UCLA).

At John Hopkins, Nathan Crone and his research group have generously supported this research with their time and effort. Jeff Liu, Victor Chen and other members of the Lenz and Anderson groups formed a lab-away-from-home during my many visits to Baltimore.

I am particularly indebted to research collaborators who helped collect some of the data presented in this dissertation when I was unable to be present: Laura Groomes, Ishita Basu, Hsiang-Yu Yu, Chien-Chen Chou, Ali Titiz, and Nanthia Suthana. In particular, Matias Ison designed and collected the physiological data for the memorability experiments, and Calin Buia collected part of the occlusion experiments. Working with epilepsy patients is difficult, time-consuming, and often frustrating, so a special thanks to these researchers who were willing to help out.

The extensive psychophysics experiments described in this thesis were largely carried out by an intrepid group of undergraduates. For the behavioral studies with occlusion: Walter Hardesty, Josue Ortega, and Ana Parades. For the memorability studies: Candace Ross, James Carroll, and Sarah Dowcett.

Jedediah Singer, Radhika Madhavan, and Arjun Bansal took me in as a fresh graduate student and mentored me on how to effectively work with patients and skillfully navigate the clinical environment.

The torch has now been passed on to William Lotter, Leyla Isik, and Joseph Olson. These new lab members have supported me through this final stretch and helped proofread this thesis. In particular, the computational part of this thesis would be non-existent without the insightful work of William Lotter.

James Hogle and Michele Jakalouv have put together an excellent graduate program, giving me exceptional freedom to explore the breadth of science. Members of my dissertation advisory and thesis committees generously donated their advice and time: Ken Nakayama, Margaret Livingstone, Sam Gershman, and Richard Born.

Of course, a Ph.D. is a marathon, not a sprint. I am fortunate to have so many friends in Boston. I have thoroughly enjoyed weekly lunches with the 'LMA kids,' the occasional meetings of the Biophysics Breakfast Club, and the Boston contingent of the integrated science group.

I am indebted to my girlfriend Christine, who has braved four years of long-distance relationship while patiently waiting for me to complete my graduate studies. She has tolerated last-minute cancellations due to the unpredictability of patient schedules and weekends where she travels to Boston just to work in lab with me. Throughout this, she has been supportive during the ups and downs of graduate work.

I would like to thank my family for always being there. My parents and brother have not only fully supported my decisions, but also encouraged me down this path. They are adamant that I follow my interests and stay happy, which perhaps more than anything, has been my guiding principle these last five years.

Brimming with an infectious passion for science, David Botstein and William Bialek's fantastic integrated science program at Princeton formed the intellectual foundations of my thinking process. Michael Berry, my senior thesis advisor, directed my enthusiasm toward the mysteries of neuroscience and transitioned me from student to scientist. While at the RAND Corporation, Mark Arena and John Schank taught me how to tell a story, and work with Steve Berner inspired my return to science.

Lastly but most importantly, the insights and enthusiasm of my thesis advisor, Gabriel Kreiman, permeate this dissertation. I am especially grateful to Gabriel for teaching me how to do science: ask the interesting questions, don't care too much about what has been done before, and follow the data to the story. From start to finish, Gabriel has been willing to do whatever necessary to guide my research, from funding trips to collect data in far-flung hospitals to personally reading the literature on a new field I wanted to explore. Gabriel treated me as a valued peer and persistently asked how he could help me do my best work. Because of his approach, throughout graduate school, I felt incredibly empowered; that my only obstacle was my own limitations as a scientist.

Prior Publications

This thesis is primarily based on the following publications:

- **Tang H** and Kreiman G. Neural representations during object completion. Book chapter in *Computational and Cognitive Neuroscience of Vision* (2016). (**Chapter 1**)
- **Tang H**, Buia C, Madhavan R, Crone N, Madsen J, Anderson WS, Kreiman G. Spatiotemporal dynamics underlying object completion in human ventral visual cortex. *Neuron*. Vol. 83(3):736-748 (2014). (**Chapters 2 and 3**)
- **Tang H***, Lotter W*, and Kreiman G. Recurrent computations during object completion. *In preparation*. (**Chapters 4 and 5**)
- **Tang H**, Yu H, Chou C, Crone N, Madsen J, Anderson W, and Kreiman G. Cascade of neural processing orchestrates cognitive control in human frontal cortex. *Submitted*. (**Appendix A**)
- **Tang H**, Singer J, Pivazyan G, Ison M, Romaine M, Frias R, Meller E, Boulin A, Carroll J, Perron V, Dowcett S, Arlellano M, Kreiman G. Computational prediction of episodic memory formation from movie events. *Submitted*. (**Appendix B**)

The author has also contributed to the following publications during graduate school:

- Madhavan R, Millman D, **Tang H**, Crone N, Lenz F, Tierney T, Madsen J, Kreiman G, Anderson WS. Decrease in gamma-band activity tracks sequence learning. *Frontiers in Systems Neuroscience*. Vol. 8:222 (2014).
- **Tang H** and Kreiman G. Face recognition: vision and emotions beyond the bubble. *Current Biology*. Vol. 21(21): 888-90. Commentary (2011).

*indicates equal contribution

1

Introduction

The statistical regularities of nature give rise to patterns: curves are usually continuous, faces have certain features, speech has a unique signature. We rely on these recognizable patterns of shape, sound, and behavior to interpret the sensory world around us. Often, these patterns are incomplete, yet we are still able to recognize partially visible objects or identify sounds from a noisy background. The neural circuits that mediate this pattern completion are a fundamental component of intelligence, and may also explain high-level cognitive phenomena such as predicting the trajectory of ball or a sequence of musical notes.

Pattern completion is particularly prominent in natural vision. Stimuli are often partially occluded, subject to poor illumination, or presented with limited viewing angles. While much progress has been made toward elucidating the mechanisms underlying recognition of whole objects, more

difficult conditions such as object occlusion remain poorly understood. In particular, occlusion presents a challenge to existing feed-forward theories of vision and computational algorithms. Solving these problems necessitates exploring new model architectures.

Understanding how the neural representations of visual signals are modified with occlusion is critical to this exploration and may also shed light on manifestations of pattern completion in other domains. The development of feed-forward models for visual recognition of whole objects has been driven by behavioral and physiological experiments establishing the hierarchy of feature tuning and robustness to image transformations. These findings form the core of modern computer vision algorithms. Similarly, systematically examining when and where neural representations that are robust to occlusion emerge can help extend our theoretical understanding of vision and develop the next generation of computational models in vision.

In this dissertation, I provide behavioral, neurophysiological, and computational evidence for the role of recurrent computations in the recognition of occluded objects. Chapter 1 reviews the prior approaches to understanding the neural representations under occlusion. Chapter 2 introduces our main method of invasively recording from the human brain, which provides the high spatiotemporal resolution necessary to examine the dynamics of object completion. Chapter 3 presents neural recordings suggesting the involvement of recurrent computations. Chapter 4 demonstrates that disrupting this recurrence significantly degrades behavioral performance. Chapter 5 explores the instability of existing feed-forward models to occlusion and proposes a recurrent neural network that reaches human-like performance on occluded object recognition tasks.

1.1 Nomenclature

Throughout this thesis, the term ‘occluded’ objects refers to any image where the object has missing features, regardless of the presence of an occluding shape. Similarly, we use interchangeably the terms

‘object completion’ and ‘recognition’ to refer to the recognition of occluded objects.

1.2 Other projects

The nature of our recording method is such that electrode locations are determined based on clinical need. Therefore, we often have patients without coverage of visual cortex. Since these patients are a rare resource, I have developed tasks for patients with frontal cortex coverage and/or medial temporal lobe coverage. The appendices detail two separate studies with human neural recordings on the dynamics of conflict signals during cognitive control ($N = 20$ subjects, Appendix A), and the neural representation of memorability in human medial temporal lobe ($N = 17$ subjects, Appendix B). Several other projects omitted from this thesis are experiments on the dynamics of visual imagery ($N = 4$ subjects), on receptive field sizes in human visual cortex ($N = 2$ subjects), and on the neural correlates of prediction ($N = 8$ subjects).

1.3 Visual system hierarchy

Object recognition is orchestrated through a semi-hierarchical series of processing areas along ventral visual cortex (Connor et al., 2007; DiCarlo et al., 2012; Felleman and Van Essen, 1991; Logothetis et al., 1995; Riesenhuber and Poggio, 1999; Tanaka, 1996). At each step in this hierarchy, the feature specificity of the neurons increases in complexity. For example, neurons in primary visual cortex (V1) respond selectively to bars of a particular orientation (Hubel and Wiesel, 1959), whereas neurons in inferior temporal cortex respond preferentially to complex shapes including faces and other objects (Desimone et al., 1984; Gross et al., 1969; Perrett, 1974; Richmond et al., 1983; Rolls, 1991). In addition to this increase in feature complexity, there is a concomitant progression in the degree of tolerance to object transformations such as changes in object position or scale (Hung et al., 2005; Ito et al., 1995; Logothetis et al., 1995). The selective and tolerant physiological responses characterized

in the macaque inferior temporal cortex have also been observed in the human inferior temporal cortex (Allison et al., 1999; Liu et al., 2009). The timing of these neural responses places important constraints on the number of possible computations involved in visual recognition. Multiple lines of evidence from human psychophysical measurements (Potter and Levy, 1969; Thorpe et al., 1996), macaque single unit recordings (Hung et al., 2005; Keysers et al., 2001), human EEG (Thorpe et al., 1996) and human intracranial recordings (Allison et al., 1999; Liu et al., 2009) have established that selective responses to whole objects emerge within 100-150 ms of stimulus onset in the highest echelons of the ventral visual stream.

Research over the last several decades characterizing the spatiotemporal dynamics involved in the neural representation of objects in these successive areas has led to the development of a theoretical framework to explain the mechanisms underlying object recognition. This theory suggests that, to a first approximation, processing of visual information traverses through the ventral stream in a feed-forward fashion, without significant contributions from long top-down feedback loops or within-area recurrent computations (Deco and Rolls, 2004; Fukushima, 1980; LeCun et al., 1998; Olshausen et al., 1993; Mel, 1997; Wallis and Rolls, 1997). Consistent with this notion, computational models of object recognition instantiating feed-forward processing provide a parsimonious explanation for the selectivity and tolerances observed experimentally (Serre et al., 2007b). The activity of these computational units at various stages of processing also capture the variance in firing rates from corresponding layers in the macaque visual system (Cadieu et al., 2014; Yamins et al., 2014). These feed-forward computational models have inspired the development of deep convolutional networks that perform significantly better than previous computer vision approaches to object recognition (Hinton and Salakhutdinov, 2006; LeCun et al., 1998; Russakovsky et al., 2015; Sun et al., 2014; Taigman et al.).

These purely feed-forward architectures do not incorporate any feedback or recurrent connections. However, at the anatomical level, feedback and recurrent connections figure prominently

throughout the visual system (Felleman and Van Essen, 1991). In fact, quantitative anatomical studies demonstrate that feedback and recurrent connections significantly outnumber feed-forward ones (Callaway, 2004; Douglas and Martin, 2004). These connections are largely absent in existing computational models because their underlying roles remain unclear. In addition to the role of feedback in attentional modulation, several investigators have suggested that these feedback and recurrent projections could play an important role during object recognition under conditions where the visual stimuli are impoverished (e.g. poor illumination, low contrast) or even partially missing (e.g. visual occlusion) (Carpenter and Grossberg, 2002; Hopfield, 1982; Mumford, 1992; Wyatte et al., 2012).

1.4 The challenge of object completion

Figure 1.1 shows examples of several images that induce object completion. In the natural world, objects can be partially occluded in multiple ways due to the presence of explicit occluders, shadows, camouflage, or illumination. Object completion is also an ill-posed problem: in general, there are an infinite number of ways to complete partially visible contours and objects. The visual system must be able to extract high level properties of occluded objects (identity, pose, intention, etc.) despite the existence of all these possible solutions.

1.4.1 Amodal completion of simple shapes

Occluded shapes are perceived as whole in the presence of an occluder (Figure 1.1A, left panel). However, this perception is dependent on the spatial arrangement; when the occluded shape and its occluder are separated, the circle appears notched (Figure 1.1A, right panel). Object completion is defined as amodal when there is an explicit occluder and the subject cannot see the contours behind the occluder despite being aware of the overall shape (Singh, 2004). In contrast, in the famous illusory triangle example (Figure 1.1B), Kanizsa describes the phenomenon known as modal

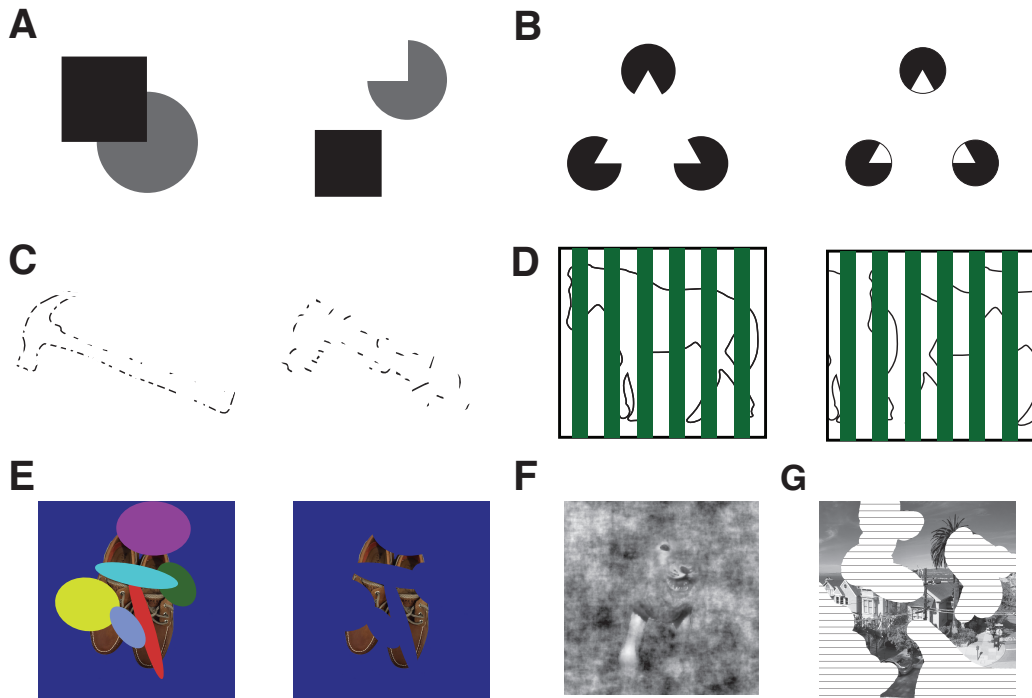


Figure 1.1: Example occluded stimuli

- (A) Occluded geometric shape (left) and its mosaic counterpart (right), similar to (Murray, 2004).
- (B) Example of modal completion inducing an illusory triangle (left). This percept is disrupted by adding edges to the inducers (right).
- (C) Line drawing of an object defined by disconnected segments and its fragmented counterpart, similar to (Doniger et al., 2000; Sehatpour et al., 2008).
- (D) Line drawing of an occluded object and its scrambled counterpart, similar to (Lerner et al., 2002).
- (E) Occluded object and its ‘deleted’ counterpart, similar to (Johnson and Olshausen, 2005).
- (F) Example partial image of an object seen through bubbles with a phase-scrambled background to equalize contrast, similar to (Tang et al., 2014).
- (G) Example partial image of a scene, similar to (Nielsen et al., 2006a).

completion whereby the object is completed by inducing illusory contours that are perceived by the observer (Kanizsa, 1979). Because these illusory inducers are rare in natural vision, in this chapter we focus on amodal completion. Even though occluded or partial objects such as the ones shown in Figures 1.1C-D are segmented, observers view the object as a single percept, not as disjointed segments. Amodal completion is important for achieving this single ‘gestalt’. Investigators have used

a variety of different stimuli to probe the workings of object completion, ranging from simple lines and geometric shapes to naturalistic objects such as the ones shown in Figure 1.1E-G.

Psychophysical studies of amodal completion have provided many clues to the underlying computations (Kellman et al., 2001; Sekuler and Murray, 2001). Amodal completion relies on an inferred depth between the occluder shape and the occluded object, which in turn generates a surface-based representation of the scene (Nakayama et al., 1995). In fact, presence of the occluder aids in identifying the occluded object, as powerfully illustrated by the Bregman's occluded B letters (Bregman, 1981). Grouping of different parts into a complete whole, and the 'completion' of missing lines and contours, represent an important component of object recognition. The ambiguities arise from the many combinations with which occluded edges, called 'inducers' can be paired together, as well as the infinite number of possible contours between two pairs of inducers (Kellman et al., 2001; Nakayama et al., 1995; Ullman, 1976). Despite the many possible solutions, the visual system typically arrives at a single (and correct) interpretation of the image.

The temporal dynamics of shape completion can constrain the computational steps involved in processing occluded images. Psychophysics experiments have measured the time course of amodal completion with a diverse array of experimental paradigms. The most common method contrasts an occluded shape against its mosaic parts (e.g. Figure 1.1A). For example, in the prime matching paradigm, subjects are first primed with a stimulus, and then asked to judge whether a pair of test stimuli represent the same or a different shape. Subjects are faster to correctly respond 'same' when the primed shape is the same as the test stimuli. When partly occluded objects are used as the prime, this priming effect depends on the exposure time (Sekuler and Palmer, 1992). At short durations (50 ms), occluded objects primed subject's responses toward mosaic shapes, suggesting that 50 ms is not enough time for amodal completion of the prime stimulus. At longer durations (100 ms or more), the priming effect switched to favor whole shapes. Therefore, the authors estimate that amodal completion for simple geometric shapes occurs at between 100 and 200 ms after stimulus onset, depending

on the amount of occlusion (Sekuler et al., 1994). A different set of behavioral experiments suggests approximately the same time scales for amodal completion: in several studies, subjects are asked to discriminate shapes in a timed forced-choice task. Response times to occluded shapes lagged those to whole shapes by about 75-150 ms (Murray et al., 2001; Shore and Enns, 1997).

1.5 Neural representation of occluded shapes

Essential aspects of shape completion can be traced back to the earliest stages in visual processing. An early study demonstrated that neurons in area V2 showed selective responses to illusory contours (Peterhans and von der Heydt, 1991; von der Heydt et al., 1984). One study has demonstrated that even V1 neurons can respond to occluded shapes (Sugita, 1999). The author recorded single cells in macaque V1 while presenting occluded moving bars (Sugita, 1999). Approximately 12% of orientation-selective cells responded to the moving oriented bar even when it was occluded, thus potentially describing the phenomenology of amodal completion. These cells responded strongly only when the occluder was presented in front of the moving bar (positive disparity), and not at zero or negative disparity. Notably, responses to the occluded bar were not different from those obtained when presenting the bar alone. These results have led to the suggestion that amodal completion is achieved by contextual modulation from outside the classical receptive field. While other studies have suggested that contextual modulation occurs with a delay of 50-70 ms with respect to the onset of the visually evoked responses (Bakin et al., 2000; Zipser et al., 1996), Sugita did not observe any latency delays for the amodally completed response. Instead, the author suggests that these contextual modulations may come from lateral connections or fast feedback from proximal areas. In another study, responses to illusory contours in V1 (modal completion) were delayed by about 55 ms compared to the response to real contours (Lee and Nguyen, 2001). Importantly, illusory contour responses appeared first in V2 before emerging in V1, suggesting that modal completion in V1

might require feedback modulation from V2. Complementing these studies, psychophysical studies on the effect of inferred depth and apparent motion on the perception of occluded surfaces also conclude that amodal completion effects manifest in early visual processing (Shimojo and Nakayama, 1990b,a).

These neurophysiology studies have focused on the occlusion of linear contours, where the inducers are close in proximity to the classical receptive field. However in natural vision we complete curvilinear contours over distances much longer than the width of classical V1 receptive fields. Often in these cases, correct completion of an object depends on the global context in which the object is embedded. Future studies are needed to examine whether and when V1 neurons respond to completed contours of varying curvature, length, and context.

As outlined above, V1 neurons feed into a cascade of semi-hierarchical processing steps through V2 and V4, culminating in the inferior temporal cortex (ITC) (Felleman and Van Essen, 1991). Few studies have examined the responses in intermediate visual areas to occluded shapes. A recent elegant study has begun to fill in this gap by characterizing how macaque V4 neurons respond to different curvatures when they are partially occluded by dots (Kosai et al., 2014). The authors report that neurons can maintain selectivity within a range of occlusion. While the response latency of these neurons were not delayed with the occlusion, the latency at which selectivity arose was delayed by hundreds of milliseconds.

Kovacs et al found that visually selective responses to complex shapes in ITC were similar between whole shapes and occluded shapes defined by adding noise, occluders or deleting shape parts (Kovacs et al., 1995a). Although selectivity to complex shapes was retained despite up to 50% occlusion, the absolute magnitude of the responses was modulated linearly with the amount of occlusion. Contrary to what (Kosai et al., 2014) find in V4, these authors observed delays of up to 50 ms in the response latency of occluded shapes. While it is tempting to attribute this discrepancy to differences in processing between V4 and IT, we note that the stimuli, occluding patterns, and monkey state (awake

versus anesthetized) differ between the two studies.

1.6 From amodal completion to recognition of occluded objects

Most studies of occluded object recognition have used simple shapes and contours, as described in the previous section. What remains unknown is how amodal completion of these simple components translates to recognition of the occluded objects that we encounter in natural vision. These naturalistic objects are characterized by complex textures, spatial arrangements, and color.

Two studies in macaque visual system used naturalistic stimuli. Nielsen et al examined the responses of ITC neurons to objects embedded in naturalistic scenes (Figure 1.1F) (Nielsen et al., 2006a). Using the bubbles paradigm (Gosselin and Schyns, 2001), the authors defined parts of an image that provided diagnostic value (i.e. provided information that aided recognition) versus other non-diagnostic parts. The authors first demonstrated that monkeys and humans show striking behavioral similarities in terms of what object parts are considered diagnostic (Nielsen et al., 2006b). For occluded scenes containing diagnostic parts, firing rates in inferotemporal cortex remained largely invariant to the amount of occlusion, in contrast to the findings of the Kovacs study with simpler stimuli (Kovacs et al., 1995a). However, for scenes that contained only non-diagnostic parts, the results from the Kovacs study were reproduced – the firing rate varied linearly with the amount of occlusion.

This comparison serves as a cautionary tale against extrapolating results based on geometric shapes to the processing of more naturalistic stimuli because the details of which features are revealed can play a very important role in dictating the effects of occlusion. Issa et al reached similar conclusions when demonstrating that ITC responses selective to faces were particularly sensitive to occlusion of certain parts (one eye) and that those parts could drive the responses almost as well as the whole face (Issa and Dicarlo, 2012). These results suggest that the robustness of the neural representation

to missing parts depends on the diagnosticity of the visible features.

A series of human scalp electroencephalography (EEG) studies have measured the latency at which responses differ between occluded objects and suitable control images. Using simple geometric stimuli, differences between occluded shapes and notched shapes emerged at 140-240 ms (Murray, 2004). Using more naturalistic stimuli (e.g. Figure 1.1E), other investigators report differential activity in the 130-220 ms (Chen et al., 2010) and 150-200 ms (Johnson and Olshausen, 2005) ranges. In a more difficult task with fragmented line drawings that are progressively completed, (Doniger et al., 2000) report that differences are only observed in the 200-250ms response window. Even though these studies use different stimuli with different comparisons, they all conclude that amodal completion effects can take 130-250ms to manifest.

The selection of an appropriate contrast condition is critical to the interpretation. Almost all human neuroimaging (Hegde et al., 2008; Komatsu, 2006; Lerner et al., 2002, 2004; Olson et al., 2004; Rauschenberger et al., 2004) and scalp EEG (Chen et al., 2010; Doniger et al., 2000; Johnson and Olshausen, 2005) studies with more complex objects have contrasted activity changes between an occluded object and an appropriately scrambled counterpart (e.g. Figure 1.1C,D). In these scrambled stimuli, the low-level features are maintained but disruption in their geometric arrangement renders the image unrecognizable. For example, investigators have reported differential activity in the lateral-occipital complex between occluded line drawings and their scrambled counterparts (Lerner et al., 2002). The authors reason that, since the occluded images elicit a larger response in the lateral-occipital complex (LOC) than scrambled images, the LOC could be involved in object completion.

However, LOC also demonstrates increased activity to whole objects compared to scrambled versions of those objects (Grill-Spector et al., 2001). Thus, the increased responses to whole objects may not be necessarily related to object completion mechanisms per se, but rather neural activity related to perceptual recognition.

Similarly, EEG and intracranial studies compared line drawings against their fragmented counter-

parts to measure the timing of and brain regions involved in object completion (Doniger et al., 2000; Sehatpour et al., 2008). Sehatpour et al worked with epilepsy patients who have intracranial electrodes implanted for clinical purposes. The authors take advantage of simultaneous recordings from multiple brain regions to show that line fragments elicited greater coherence in the LOC-Prefrontal Cortex-Hippocampus network compared to scrambled line fragments. They suggest that this network synchrony is responsible for the perceptual line closure of objects. Again it is challenging here to untangle effects of perceptual recognition from the involvement of closure mechanisms. Indeed, intracranial recordings with backward masking of whole objects have shown that perceptual recognition triggers a sustained neural response in visual cortex that 'ignites' a widespread network of processing (Fisch et al., 2009).

1.6.1 Motivation for our work

feed-forward models have been a mainstay in computational neuroscience for the last several decades. In this work, we sought to 'break' this model with occlusion. We then examined the visual system's response in a way that is conceptually different from previous experiments in human brain. Instead of comparing occluded line drawings against a scrambled counterpart, we used complex naturalistic objects and high spatiotemporal recordings to examine where and when visual information is preserved even with occlusion. The neural and behavioral experiments provided evidence for a role for recurrent processing, which we implemented in a proof-of-principle recurrent neural network that matched human performance on recognition of occluded objects.

2

Neurophysiology Recordings

Our understanding of human brain function hinges on our ability to interrogate neural circuits within human cortex. Human brain recordings pose a unique set of ethical and methodological challenges not found in animal models. Several non-invasive methods have been extensively used in the last few decades with tremendous success. Function magnetic resonance imaging (fMRI) measures blood flow, a plausible correlate of neuronal activity (Logothetis et al., 2001). This method allows researchers to measure activity in brain tissue from a wide spatial region, but lacks temporal precision (each scan can take ≥ 1 second, whereas visual events may be processed in hundreds of milliseconds). Electroencephalography (EEG) and magnetoencephalography (MEG) places electrodes near or on the scalp to measure respectively the electrical and magnetic fields generated by neural activity. This approach allows high temporal resolution recordings, but due to the distance of

the sensors from cortex, and the distorting properties of the skull and scalp, spatial localization and signal-to-noise are challenging issues.

2.1 Intracranial electrodes

Opportunities to invasively probe human brain with high spatiotemporal resolution techniques are rare. In this thesis, I describe studies in human epilepsy patients which allow such an approach. These patients have pharmacologically intractable epilepsy and therefore are candidates for surgical resection of the epileptogenic tissue (Penfield and Jasper, 1954). In order to map the seizure foci, and to avoid resecting functional regions, the clinical team will surgically implant subdural electrodes (Ojemann, 1997). The patients will then stay in the hospital for 1-2 weeks to allow the neurologists to record several seizure events. During this period, subjects can volunteer to participate in our research studies.

These intracranial electrodes have a 2.3mm diameter and are arranged into grids and strips with 1 cm separation. Each electrode is composed of a platinum-iridium alloy and has an impedance of approximately 60Ω . Because the electrode locations are driven by clinical considerations, brain coverage varies considerably across subjects. An example brain coverage of a single patient is shown in Figure 2.1. Note that electrode locations were driven by clinical considerations; the majority of the electrodes across the patient population were not in the visual cortex.

Implantation of these electrode serve two functions. First, a small subset of these electrodes are eventually identified as the seizure foci. Second, the electrode coverage plan is designed to cover and identify functional areas to avoid during the resection procedure. Therefore, the majority of the implanted electrodes are over functional tissue.

From these electrodes, we can record the intracranial field potential (IFP) from human cortex. Besides analyzing the field potential response, the signal can also be decomposed into various frequency

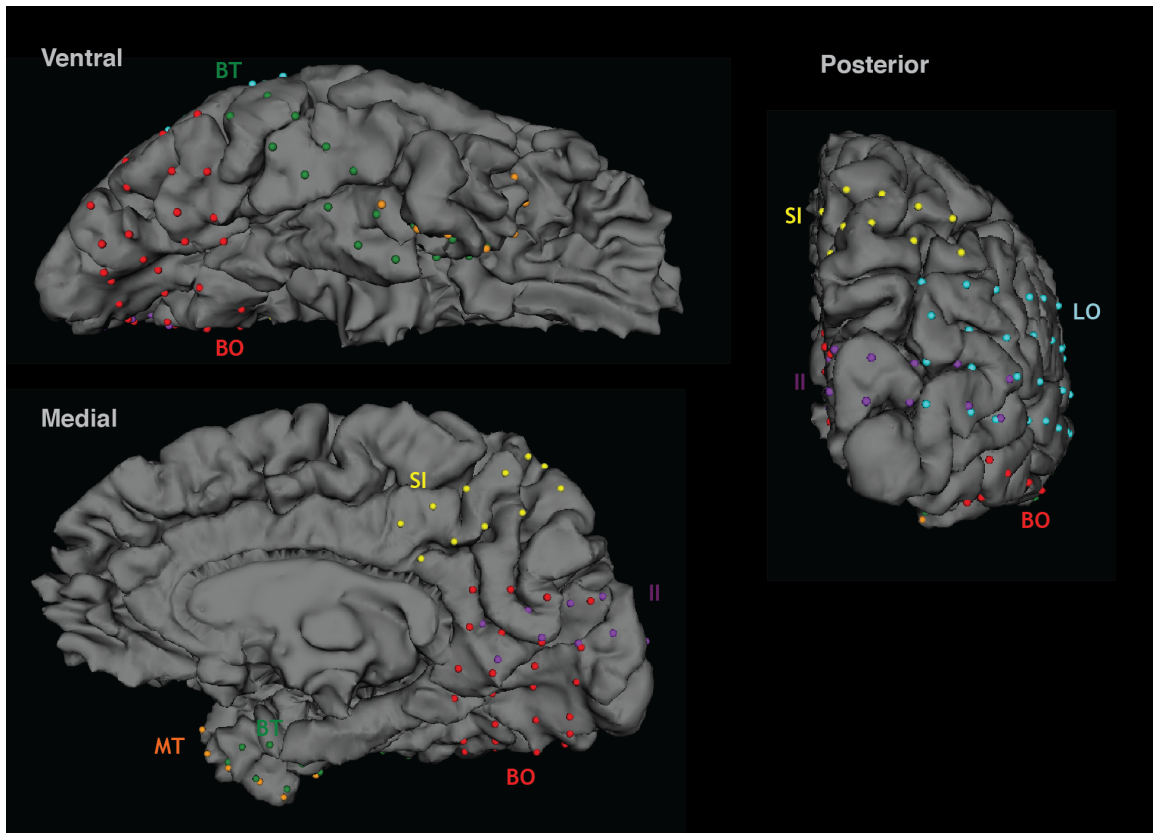


Figure 2.1: Example placement of intracranial electrodes

Example coverage from a single patient with intracranial electrodes from three views (ventral, medial, and posterior). Different strips and grids are have their own unique color and name. These electrodes can be placed on the lateral surface (where the craniotomy is typically located), slipped underneath to cover the ventral visual stream (e.g. BT and BO), or placed inter-hemispherically (e.g. II and SI).

bands. Power in these different bands is believed to underly a diverse set of high-level cognitive processes (Buzsaki et al., 2012), including the processing of visual signals (Davidesco et al., 2013; Vidal et al., 2010; Liu et al., 2009). In this study, we computed the power in the theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), low gamma (30-70 Hz), and high gamma (70-100 Hz) frequency ranges.

There are many different ways of computing spectral power. Because temporal resolution is critical to understanding visual processes, we calculated power by first applying a bandpass to the data (4th order Butterworth filter), then using the Hilbert transform to compute the magnitude of the analytical

representation of the signal. This approach sacrifices accuracy in favor of temporal resolution. Other methods, such as multi-taper Fourier transforms or wavelet analysis use time windows to increase frequency resolution at the expense of temporal resolution.

2.2 Electrode localization

Patients with implanted intracranial recordings typically receive a pre-operative MRI which shows the brain tissue, and then a CT scan after implantation that identifies the electrode locations. We use open-source software (freesurfer) to align the post-operative CT with the pre-operative MRI. This software also generates a 3D model of the brain surface from the MRI. Because these intracranial electrodes have substantial thickness, their implantation induces some brain compression, particularly near the craniotomy, which makes the electrodes appear 'inside' the brain in the pre-operative MRI (Figure 2.2). To account for this, we manually project the electrode locations back onto the surface of the cortex along the axis normal to the surface.

We perform this projection individually for each electrode, then for validation plot all the electrodes on the surface to check that they roughly follow a grid pattern. The result of this processing is for each electrode, a surface coordinate as well as a region label. For the region labels, we follow the brain atlas in (Destrieux et al., 2010). Depth electrodes are also localized to subcortical structures. We also use several scripts to visualize the electrode locations in a software called Slicer3D. The software and instructions for this processing pipeline are found at <https://github.com/hanlint/fs-coreg>.

2.3 Biophysics of intracranial recordings

The neural mechanisms underlying the local field potential we observe with these recordings are a matter of active research. Two components figure prominently in our study -- the broadband signal (e.g. 0.1-100 Hz) and high frequency gamma activity (e.g. 70-200 Hz). The broadband signal is

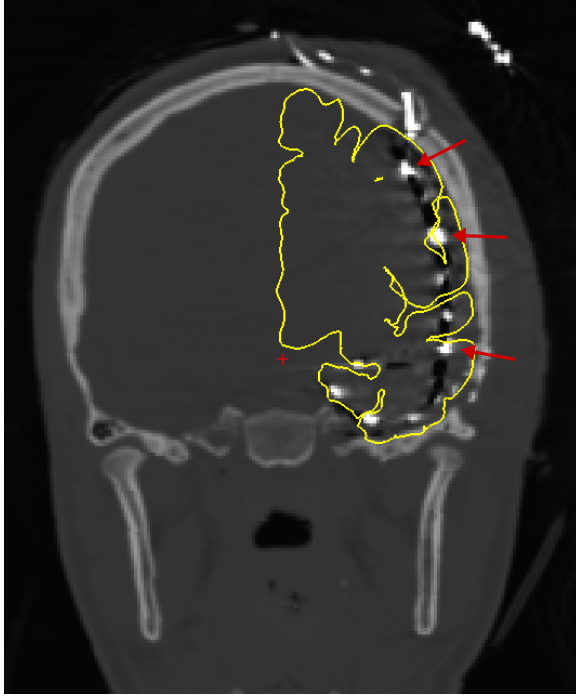


Figure 2.2: Electrode localization.

Coronal section of CT scan for a subject with the electrodes marked by the red arrows. The brain surface was extracted from the pre-operative MRI and shown in yellow. Because of compression of the brain by the intracranial electrodes, the electrodes need to be manually projected back out to the surface of cortex.

thought to represent the summation of postsynaptic activity (Logothetis et al., 2001; Mitzdorf, 1987; Buzsaki et al., 2012). Extracellular currents from many neurons must overlap in time to generate a measurable signal, and this overlap is strongest for slow events, such as synaptic currents. However, we note that recent biophysical simulations show that spikes can induce not only fast charge fluxes, but also a cascade of slower spiking currents at longer time scales that can also significantly contribute to the local field potentials (Reimann et al., 2013).

Multiple studies have correlated high frequency activity in the 70-200 Hz band, denoted the high gamma band, with the underlying population firing rate (Buzsaki et al., 2012; Ray and Maunsell, 2011; Nir et al., 2007; Manning et al., 2009). However, we note that these experiments were conducted with microwire recordings. Whether the same conclusions apply to the much larger intracranial used in our study is unclear.

Despite these ambiguities, the properties of visual cortex we observe with broadband signals and

Experiment	# of Subjects	Location
Recognition of occluded objects	18	Chapters 3-6
Dynamics of cognitive control	20	Appendix A
Neural representation of memorability	17	Appendix B
Fine spatial resolution recordings with microwire arrays	10	
Neural correlates of prediction	8	
Dynamics during visual imagery	4	
Receptive field sizes in human visual cortex	2	

Table 2.1: **Table of experiments**

Distribution of patients across my various experiments. Some patients may have participated in multiple experiments.

gamma band activity, such as object selectivity and tolerance to scale and position transformations, can match those found with single neuron recordings (Liu et al., 2009), and have formed the basis of many other research studies (Bouchard et al., 2013; Vidal et al., 2010; Singer et al., 2015; Davidesco et al., 2013; Fisch et al., 2009).

2.4 Comparison with other methods

There are several advantages and disadvantages to our recording method. Compared to experiments in macaques and rodents, human subjects allow for rapid training through verbal instruction (several minutes instead of several months). We can also ask more complex questions, particularly those that we detail in Appendix A and Appendix B. Other methods for recording from human brain are either too slow to visualize the dynamics of the visual system (fMRI) or may have a low signal-to-noise because of distortions introduced by the scalp tissue (EEG and MEG). In comparison, our recordings have high spatiotemporal resolution, but with several caveats.

First, data collection is limited to the inflow of patients at these hospitals. I have been fortunate to have access to patient populations for many local and international hospitals (Boston Children's Hospital, Brigham and Women's hospital, Massachusetts General Hospital, UCLA Medical Center,

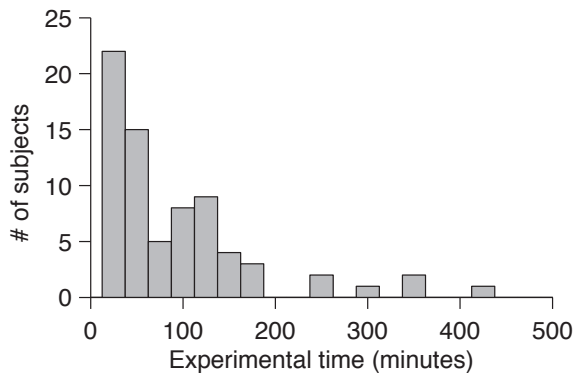


Figure 2.3: **Experimental time.**
The distribution of experimental time (in minutes) for the $N = 72$ patients that participated in my graduate research.

and Taipei Veteran's General Hospital). My graduate work is a product of data collected from ~ 72 patients across these hospitals, distributed over several experiments (Table 2.1).

Second, we do not have control over the location and type of electrodes that are implanted. Given these rare opportunities, it would be inefficient for example, to collect data on a visual task when the coverage is predominantly prefrontal cortex. As a consequence, I have developed several different tasks optimized for different brain coverage scenarios. The results from some of these other studies are described in Appendix A and Appendix B.

Third, even when patients are in the hospital, experimental time is extremely limited. Patients require time to recover from the trauma of brain surgery, may have family or friends visiting, or simply are not interested in volunteering for research. Clinicians also have a battery of tests that take precedence over research, and oftentimes experimental time has to be coordinated with other research groups. Even though patients are in the hospital for 1-2 weeks, the patients that participated in the research studies during my graduate work average 94 ± 83 (mean \pm SD) minutes of experimental time (Figure 2.3). In addition, the electrodes are explanted after this period, so the same patients cannot be retested with additional controls that may be needed. Therefore, experiments must be well-designed ahead of time and well-prepared for various contingencies that arise from collecting data in a clinical environment (e.g. need to pause experiment if clinicians need to converse with

patient or nurses need to deliver medicine, or switch to using keyboard input if the patient struggles with using a wireless gamepad).

Despite these challenges, the high spatiotemporal resolution of these invasive recordings, as well as the ability to probe complex questions with human subjects, and the fact that these are human brains and not a model organism, demonstrate that these recordings have the potential to transform our understanding of neural circuits.

3

Dynamics of object completion

Recordings in inferotemporal cortex of monkeys (Desimone et al., 1984; Hung et al., 2005) and humans (Liu et al., 2009) have revealed a significant degree of tolerance to object transformations. Visual recognition of isolated objects under certain transformations such as scale or position changes do not incur additional processing time at the behavioral or physiological level (Biederman and Cooper, 1991; Logothetis et al., 1995) and can be described using purely bottom-up computational models. While bottom-up models may provide a reasonable approximation for rapid recognition of whole isolated objects, top-down as well as horizontal projections abound throughout visual cortex (Callaway, 2004; Felleman and Van Essen, 1991). The contribution of these projections to the strong robustness of object recognition to various transformations remains unclear. In particular, recognition of objects from partial information is a difficult problem for purely feed-forward architectures

and may involve significant contributions from recurrent connections as shown in attractor networks (Hopfield, 1982; O'Reilly et al., 2013) or Bayesian inference models (Lee and Mumford, 2003).

As described in Chapter 1, previous studies have demonstrated that recognition of occluded shapes and lines can begin in early visual cortex (V1, V2, and V4) (Lee and Nguyen, 2001; Sugita, 1999; Kosai et al., 2014). However, how these processes contribute to recognition of complex naturalistic objects is unclear. Often, findings with shape stimuli do not translate to more complex stimuli*. Most studies of complex objects use line drawings (Sehatpour et al., 2008; Doniger et al., 2000; Lerner et al., 2004) or untextured geometric objects (Hegde et al., 2008; Olson et al., 2004). In addition, these studies typically contrast occluded objects against an unrecognizable scrambled counterpart (e.g. Figure 1.1), which may confound object completion mechanisms with activity related to perceptual recognition.

In this study, we instead use intracranial recordings to measure object selectivity while subjects recognized naturalistic objects from partial information. Importantly, we compare neurophysiological responses to occluded objects from different categories that are all perceptually recognizable. Even with very few features present (9-25% of object area shown), neural responses in the ventral visual stream retained object selectivity. These visually selective responses to partial objects emerged about 100ms later than responses to whole objects. The processing delays associated with interpreting objects from partial information increased along the visual hierarchy. These delays stand in contrast to the position and scale transformations that do not incur delays. Together, these results argue against a feed-forward explanation for recognition of partial objects and provide evidence for the involvement of highest visual areas in recurrent computations orchestrating pattern completion.

*See discussion in Chapter 1, page 10 comparing (Kovacs et al., 1995a) and (Nielsen et al., 2006a)

3.1 Experiment outline

We recorded intracranial field potentials (IFPs) from 1,699 electrodes in 18 subjects (11 male, 17 right-handed, 8-40 years old, Table C.1) implanted with subdural electrodes to localize epileptic seizure foci. In two subjects, eye positions were recorded simultaneously with the physiological recordings. Subjects viewed images containing grayscale objects presented for 150 ms. After a 650 ms delay period, subjects reported the object category (animals, chairs, human faces, fruits, or vehicles) by pressing corresponding buttons on a gamepad (Figure 3.1A). In 30% of the trials, the objects were unaltered (referred to as the ‘Whole’ condition). In 70% of the trials, partial object features were presented through randomly distributed Gaussian “bubbles” (Figure 3.1B, Methods, referred to as the ‘Partial’ condition) (Gosselin and Schyns, 2001). The number of bubbles was calibrated at the start of the experiment such that performance was 80% correct. The number of bubbles (but not their location) was then kept constant throughout the rest of the experiment. For 12 subjects, the objects were presented on a gray background (the ‘Main’ experiment). While contrast was normalized across whole objects, whole objects and partial objects had different contrast levels because of the gray background. In 6 additional subjects, a modified experiment (the ‘Variant’ experiment) was performed where contrast was normalized between whole and partial objects by presenting objects on a background of phase-scrambled noise (Figure 3.1B).

The performance of all subjects was around the target correct rate (Figure 3.2, $79\% \pm 7\%$, mean \pm SD). Performance was significantly above chance (Main experiment: chance = 20%, 5-alternative forced choice; Variant experiment: chance = 33%, 3-alternative forced choice) even when only 9-25% of the object was visible. As expected, performance for the whole condition was near ceiling ($95 \pm 5\%$, mean \pm SD). Subsequent analysis were performed on correct trials only.

Consistent with previous studies, multiple electrodes showed strong visually selective responses to whole objects (Allison et al., 1999; Davidesco et al., 2013; Liu et al., 2009). An example electrode from

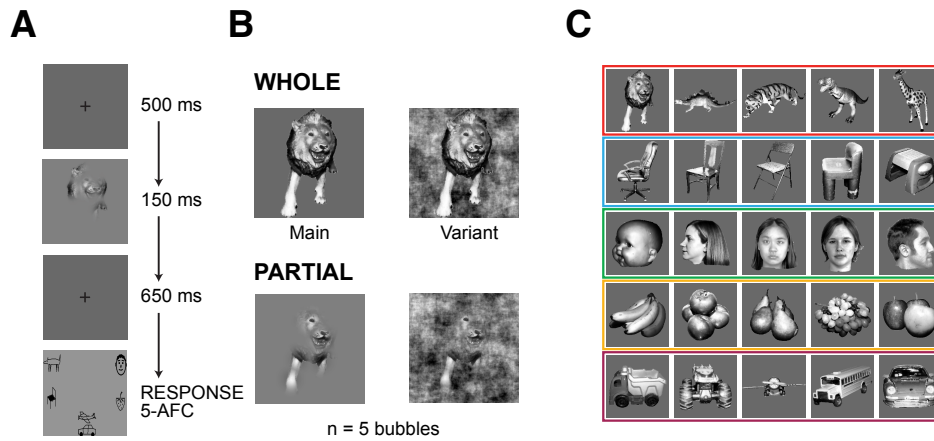


Figure 3.1: **Experimental design**

- (A) After 500 ms fixation, an image containing a whole object or a partial object was presented for 150 ms. Subjects categorized objects into one of five categories (5-Alternative Forced Choice) following a choice screen. Presentation order was pseudo-randomized.
- (B) Example images used in the task. Objects were either unaltered (Whole) or presented through Gaussian bubbles (Partial). For 12 subjects, the background was a gray screen (Main experiment), and for 6 subjects the background was phase-scrambled noise (Variant experiment). In this example, the object is seen through 5 bubbles (18% of object area shown). The number of bubbles was titrated for each subject to achieve 80% performance.
- (C) Stimuli consisted of 25 different objects belonging to five categories.

the ‘Main’ experiment, located in the Fusiform Gyrus, had robust responses to several exemplars in the Whole condition, such as the one illustrated in the first panel of Figure 3.3A. These responses could also be observed in individual trials of face exemplars (gray traces in Figure 3.3A, Figure 3.3B left). This electrode was preferentially activated in response to faces compared to the other objects (Figure 3.3C, left). Responses to stimuli other than human faces were also observed, such as the responses to several animal (red) and fruit (orange) exemplars.

The responses in this example electrode were preserved in the Partial condition, where only $11 \pm 4\%$ (mean \pm SD) of the object was visible. Robust responses to partial objects were observed in single trials (Figure 3.3A and 3.4B right). These responses were similar even when largely disjoint sets

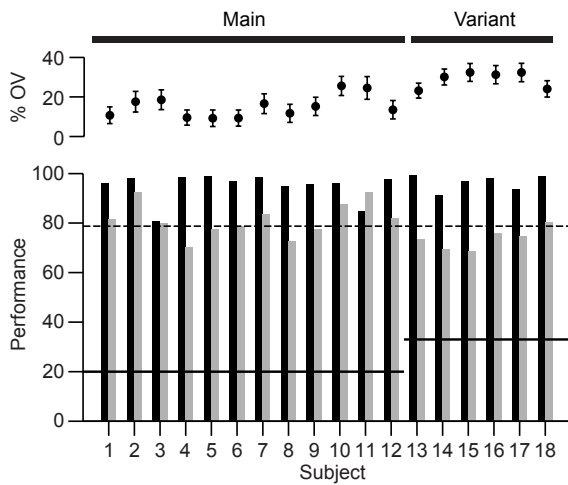


Figure 3.2: Behavioral Performance
 Top, percentage of the object visible (mean±SD) for each subject in the Main experiment (left) and the contrast-normalized Variant (right). Below, percentage of correct trials (performance) for Whole (black) and Partial (gray) objects.

of features were presented (e.g., compare Figure 3.3A, third and fourth images). Because the bubble locations varied from trial to trial, there was significant variability in the latency of the visual response (Figure 3.3B, right); this variability affected the average responses to each category of partial objects (Figure 3.3C, right). Despite this variability, the electrode remained selective and kept the stimulus preferences at the category and exemplar level (Figure 3.3C). The responses of an example electrode from the ‘Variant’ experiment support similar conclusions (Figure 3.4). Even though only $21\% \pm 4\%$ (mean±SD) of the object was visible, there were robust responses in single trials (Figure 3.4A-B), and strong selectivity both for whole objects and partial objects at the category and exemplar level (Figures 3.4C). While the selectivity was consistent across single trials, there was significantly more trial-to-trial variation in the timing of the responses to partial objects compared to whole objects (Figure 3.4B, top right).

To measure the strength of selectivity, we employed two approaches. The first approach (‘ANOVA’) was a non-parametric one-way analysis of variance test to evaluate whether and when the average category responses differed significantly. An electrode was denoted “selective” if, during 25 consecutive milliseconds, the ratio of variances across versus within categories (F-statistic) was greater than a significance threshold determined by a bootstrapping procedure to ensure a false discovery rate

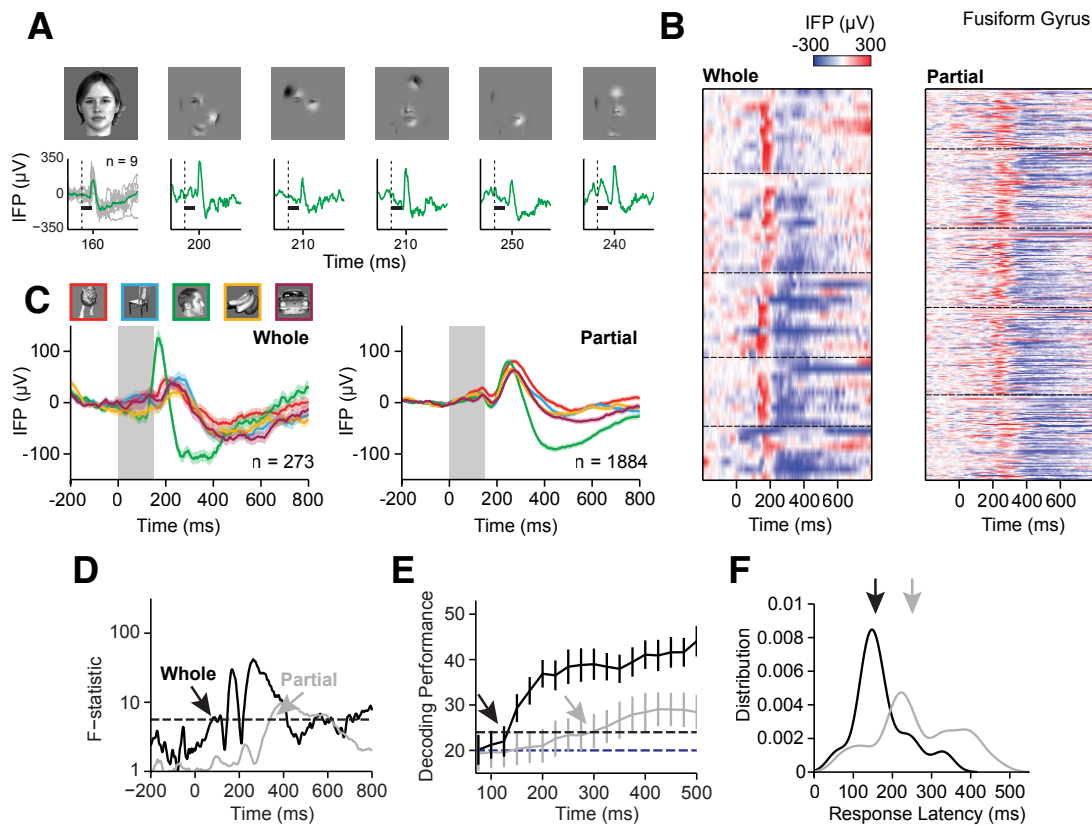


Figure 3.3: Example physiological responses from main experiment

Example responses from an electrode in the left Fusiform Gyrus.

- (A) Intracranial field potential (IFP) responses to an individual exemplar. For the Whole condition, the average response (green) and single trial traces (gray) are shown. For the Partial condition, example single trial responses (green, $n=1$) to different partial images of the same exemplar (top row) are shown. The response peak time is marked on the x-axis.
- (B) Raster of the neural responses for Whole (left) and Partial (right) objects for the preferred category (human faces). Rows represent individual trials. Dashed lines separate responses to the 5 face exemplars. The color indicates the IFP at each time point (bin size = 2 ms, see scale on top).
- (C) Average response to Whole (left) and Partial (right) objects belonging to five different categories. Shaded areas indicate s.e.m. The gray rectangle denotes the image presentation time (150 ms). The total number of trials is indicated on the bottom right of each subplot.
- (D) F-statistic at each time point for Whole (black) and Partial (gray) objects. Arrows indicate the first time point when the F-statistic exceeds the statistical threshold (black dashed line) for 25 consecutive milliseconds.
- (E) Decoding performance (mean \pm SD) for a five-way categorization task. Arrows indicate the first time when decoding performance reaches significance (black dashed line). Chance is 20% (blue dashed line).
- (F) Distribution of the visual response latency across trials for Whole (black) and Partial (gray) objects. The distribution is based on kernel density estimate (bin size = 6 ms). The arrows denote the distribution averages.

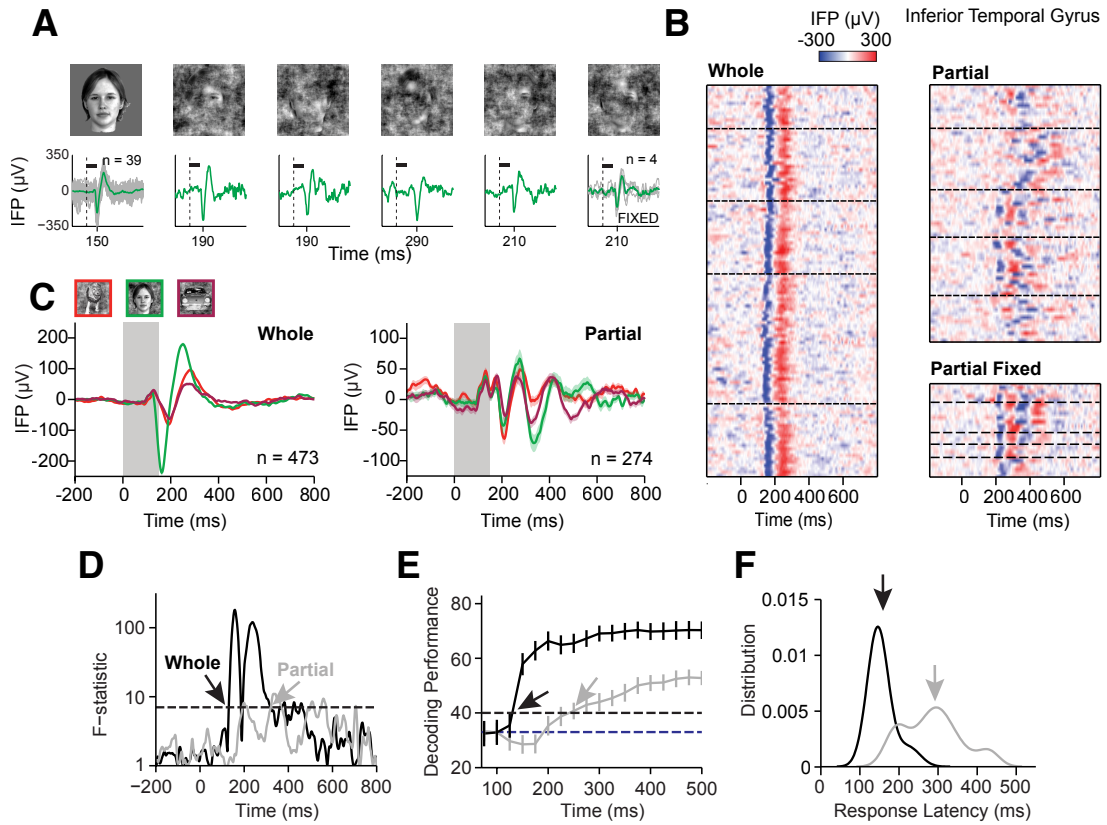


Figure 3.4: Example physiological responses from Variant experiment

Example responses from an electrode in the left Inferior Temporal Gyrus. The format and conventions are as in Figure 3.3, except that only three categories were tested, and the Partial Fixed condition was added in part A and B (Methods). Note that the statistical thresholds for the F-statistic and decoding performance differ from those in Figure 3.3 because of the different number of categories. More examples are shown in Figures C.1 and C.2.

$q < 0.001$ ($F = 5.7$) (Figure 3.3D, 3.4D). Similar results were obtained when considering d' as a measure of selectivity (Methods). The ANOVA test evaluates whether the responses are statistically different when averaged across trials, but the brain needs to discriminate among objects in single trials. To evaluate the degree of selectivity in single trials, we employed a statistical learning approach to measure when information in the neural response became available to correctly classify the object into one of the five categories (denoted ‘Decoding’; Figure 3.3E, chance = 20%; Figure 3.4E, chance = 33%). An electrode was considered “selective” if the decoding performance exceeded a threshold determined to ensure $q < 0.001$ (Methods).

3.2 Object selectivity was retained despite occlusion

Of the 1,699 electrodes, 210 electrodes (12%) and 163 electrodes (10%) were selective during the Whole condition in the ANOVA and Decoding tests, respectively. We focused subsequent analyses only on the 113 electrodes selective in both tests, (83 from the main experiment and 30 from the variant; Table 1). As a control, shuffling the object labels yielded only 2.78 ± 0.14 selective electrodes (mean \pm s.e.m., 1,000 iterations; 0.16% of the total). Similar to previous reports, the preferred category of different electrodes spanned all five object categories, and the electrode locations were primarily distributed along the ventral visual stream (Figure 3.5E-F) (Liu et al., 2009). As demonstrated for the examples in Figures 3.3 and 3.4, 30 electrodes (24%) remained visually selective in the Partial condition (Main experiment: 22; Variant experiment: 8) whereas the shuffling control yielded an average of 0.06 and 0.04 selective electrodes in the Main and Variant experiments respectively (Table 3.1).

The examples in Figure 3.3C and 3.4C seem to suggest that the response amplitudes were larger in the Whole condition. However, this effect was due to averaging over trials and the increased trial-to-trial variability in the response latency for the Partial condition. No amplitude changes are apparent

Experiment	Frequency Band	Whole	Shuffled	Both	Shuffled
Main	Broadband	83	(1.66±0.07)	22	(0.06±0.01)
Variant	Broadband	30	(1.12±0.12)	8	(0.04±0.03)
Main	Gamma	53	(1.56±0.05)	14	(0.04±0.01)

Table 3.1: **Number of selective electrodes**

For the experiment and frequency bands reported in the main text, this table shows the number of electrodes selective to whole images (‘Whole’) or to both whole and partial images (‘Both’). Also reported is the number of selective electrodes found when the object category labels were shuffled (mean±s.e.m., $n = 1000$ iterations).

in the single trial data (Figure 3.3B and 3.4B). The range of the IFP responses to the preferred category from 50 to 500 ms was not significantly different for whole versus partial objects (Figure 3.5A, $P = 0.68$, Wilcoxon rank-sum test). However, the strength of category selectivity was suppressed in the Partial condition. The median F-statistic was 23 for the Whole condition and 14 for the Partial condition (Figure 3.5B, $P < 10^{-4}$, Wilcoxon signed-rank test, an F-statistic value of 1 indicates no selectivity). The median decoding performance was 33% for the Whole condition and 26% for the Partial condition (Figure 3.5C, $P < 10^{-4}$, Wilcoxon signed-rank test). Because the Variant experiment contained only three categories, measures of selectivity such as the F-statistic or Decoding Performance are scaled differently from the Main experiment, so Figure 3.5A-D only shows data from the Main experiment. Analysis of the electrodes in the Variant experiment revealed similar conclusions.

The observation that even non-overlapping sets of features can elicit robust responses (e.g., third and fourth panel in Figure 3.3A) suggests that the electrodes tolerated significant trial-to-trial variability in the visible object fragments. To quantify this observation across the population, we defined the percentage of overlap between two partial images of the same object by computing the number of pixels shared by the image pair divided by the object area (Figure 3.5D, insert). We considered partial images where the response to the preferred category was highly discriminable from the response to the non-preferred categories (Methods). Even for these trials with robust responses, 45% of the 10,438 image pairs had less than 5% overlap, and 11% of the pairs had less than 1% overlap (Fig-

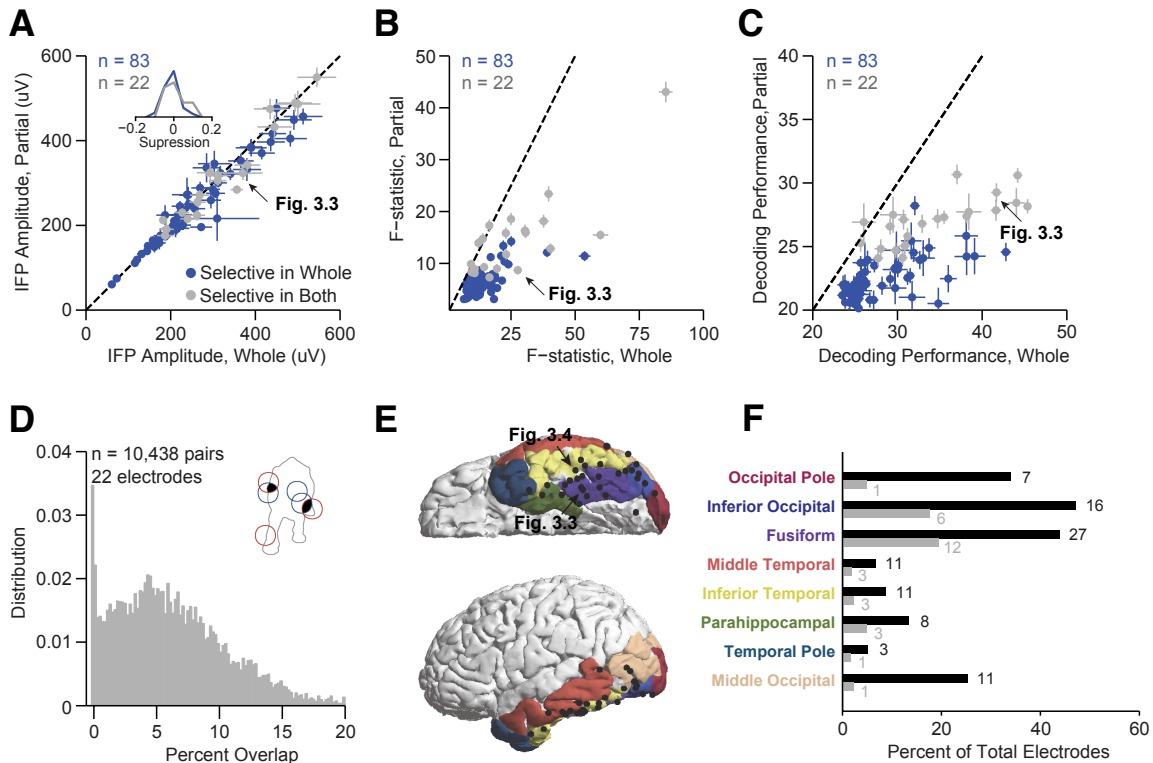


Figure 3.5: Neural responses remained visually selective despite partial information

- (A) Average IFP amplitude $A = \frac{1}{N} \sum_i \max IFP_i(t) - \min IFP_i(t)$ across trials (N) in response to partial versus whole objects for electrodes that were visually selective in the Whole condition (blue, $n = 61 + 22$), and electrodes that were visually selective in both conditions (gray, $n = 22$) (Main experiment). Most of the data clustered around the diagonal (dashed line). Inset, distribution of suppression index: $(A_{\text{whole}} - A_{\text{partial}})/A_{\text{whole}}$.
- (B) Comparison between selectivity for Partial versus Whole objects measured by the F-statistic. Most of the data were below the diagonal (dashed line). The arrow points to the example from Figure 3.3. Here we only show data from the Main experiment (F values are difficult to compare across experiments because of the different number of categories). Error bars are 99% CI.
- (C) Comparison between selectivity for Partial versus Whole objects measured by the single-trial decoding performance. Most of the data are below the diagonal (dashed line). Chance performance is 20%.
- (D) For all pairs of discriminable trials, the distribution of the percent overlap in shared pixels. The percent overlap between two pairs of trials (inset, red and blue bubbles) was defined as the number of shared pixels (black) divided by the total object area.
- (E) Locations of electrodes that showed visual selectivity in both Whole and Partial conditions. Electrodes were mapped to the same reference brain. Example electrodes from Figure 3.3 and 3.4 are marked by arrows. Colors indicate different brain gyri.
- (F) Percent of total electrodes in each region that were selective in either the Whole condition (black) or in both conditions (gray). Colors correspond to the brain regions in (E). The number of selective electrodes is shown next to each bar. Only regions with at least one electrode selective in both conditions are shown.

ure 3.5D). Furthermore, in every electrode, there existed pairs of robust responses where the partial images had <1% overlap.

To compare different brain regions, we measured the percentage of electrodes in each gyrus that were selective in either the Whole condition or in both conditions (Figure 3.5E-F). Consistent with previous reports, electrodes along the ventral visual stream were selective in the Whole condition (Figure 3.5F, black bars) (Allison et al., 1999; Davidesco et al., 2013; Liu et al., 2009). The locations with the highest percentages of electrodes selective to partial objects were primarily in higher visual areas, such as the Fusiform Gyrus and Inferior Occipital Gyrus (Figure 3.5F, gray bars, $P = 2 \times 10^{-6}$ and 5×10^{-4} respectively, Fisher's exact test). In sum, electrodes in the highest visual areas in the human ventral stream retained visual selectivity to partial objects, their responses could be driven by disjoint sets of object parts and the response amplitude but not the degree of selectivity was similar to that of whole objects.

3.3 Delayed responses to partial objects

In addition to the changes in selectivity described above, the responses to partial objects were delayed compared to the corresponding responses to whole objects (e.g. compare Whole versus Partial in the single trial responses in Figure 3.3A-B and 3A-B). To compare the latencies of responses to Whole and Partial objects, we measured both selectivity latency and visual response latency. Selectivity latency indicates when sufficient information becomes available to distinguish among different objects or object categories, whereas the response latency denotes when the visual response differs from baseline (Methods).

Quantitative estimates of latency are difficult because they depend on multiple variables, including number of trials, response amplitudes and thresholds. Here we independently applied different measures of latency to the same dataset. The selectivity latency in the responses to whole objects for

the electrode shown in Figure 3.3 was 100 ± 8 ms (mean \pm 99% CI) based on the first time point when the F-statistic crossed the statistical significance threshold (Figure 3.3D, black arrow). The selectivity latency for the partial objects was 320 ± 6 ms (mean \pm 99% CI), a delay of 220 ms. A comparable delay of 180 ms between partial and whole conditions was obtained using the single-trial decoding analyses (Figure 3.3E). Similar delays were apparent for the example electrode in Figure 3.4.

We considered all electrodes in the Main experiment that showed selective responses to both whole objects and partial objects ($n = 22$). For the responses to whole objects, the median latency across these electrodes was 155 ms, which is consistent with previous estimates (Liu et al., 2009). The responses to partial objects showed a significant delay in the selectivity latency as measured using ANOVA (median latency difference between Partial and Whole conditions = 117 ms, Figure 3.6A, black dots, $P < 10^{-5}$) or Decoding (median difference = 158 ms, Figure 3.6B, black dots, $P < 10^{-5}$). Similar effects were observed when considering two-class selectivity metrics such as d' (Figure C.3A-B).

We examined several potential factors that might correlate with the observed latency differences. Stimulus contrast is known to cause significant changes in response magnitude and latency across the visual system (Reich et al., 2001; Shapley and Victor, 1978). As noted above, there was no significant difference in the response magnitudes between Whole and Partial conditions (Figure 3.5A). Furthermore, in the Variant experiment, where all the images had the same contrast, we still observed latency differences between conditions (median difference = 73 ms (ANOVA), Figure 3.6A, and median difference = 93 ms (Decoding), Figure 3.6B, gray circles).

Because the spatial distribution of bubbles varied from trial to trial, each image in the Partial condition revealed different visual features. As a consequence, the response waveform changed from trial to trial in the partial condition (e.g. compare the strikingly small trial-to-trial variability in the responses to whole objects with the considerable variability in the responses to partial objects, Figure 3.4B). Yet, the latency differences between Whole and Partial conditions were apparent even

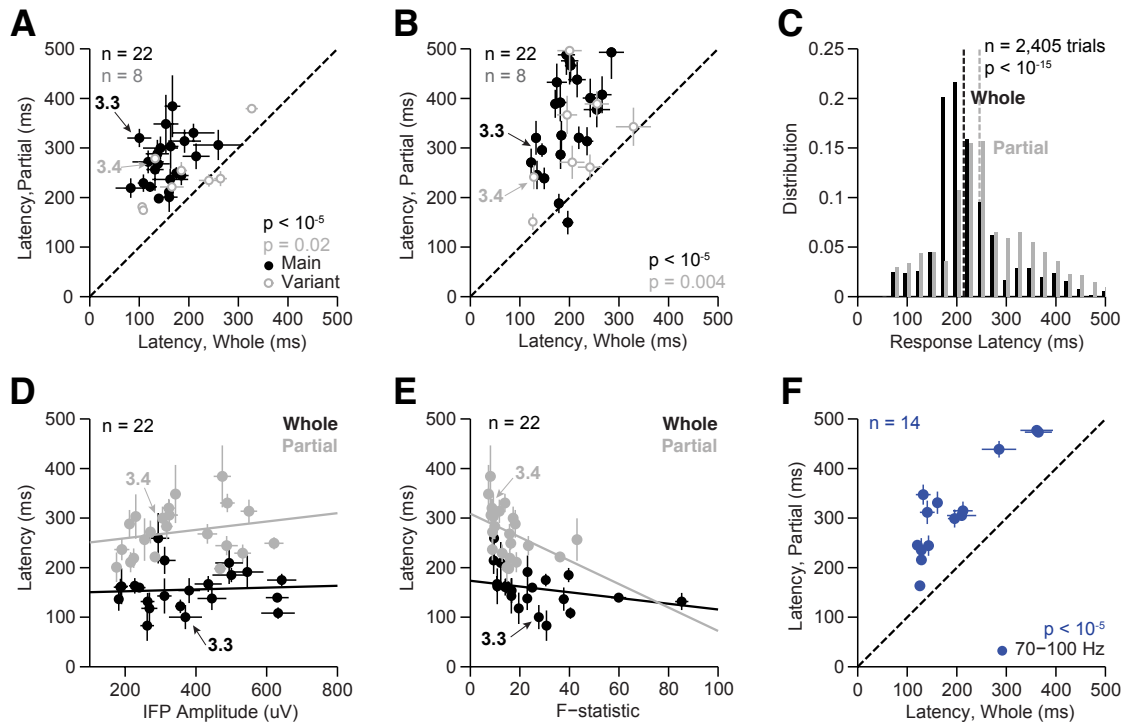


Figure 3.6: Increased latency for object completion

We considered several definitions of latency (see text).

- (A) Latency of selective responses, as measured through ANOVA (e.g. Figure 3.3D) for electrodes selective in both Whole and Partial conditions from the Main (black, $n=22$) and Variant (gray, $n=8$) experiments. The latency distributions were significantly different (signed-rank test, main experiment: $P < 10^{-5}$, variant experiment: $P = 0.02$).
- (B) Latency as measured by the decoding analysis (e.g. Figure 3.3E). These latency distributions were significantly different (signed-rank test, main experiment: $P < 10^{-5}$, variant experiment: $P = 0.004$).
- (C) Distribution of visual response latencies in single trials for Whole (black) and Partial (gray) objects (as illustrated in Figure 3.3F). These distributions were significantly different (rank-sum test, $P < 10^{-15}$). The vertical dashed lines denote the means of each distribution.
- (D) There was no significant correlation between selectivity latency (measured using ANOVA) and IFP amplitude (defined in Figure 3.5A) (Whole: $r = 0.13$, $P = 0.29$; Partial: $r = 0.15$, $P = 0.27$).
- (E) The correlation between selectivity latency and the selectivity as evaluated by the F-statistic was significant in the Partial condition ($r = -0.43$, $P = 0.03$) but not in the Whole condition ($r = -0.36$, $P = 0.06$). However, the latency difference between conditions was still significant when accounting for changes in the strength of selectivity (ANCOVA, $P < 10^{-8}$).
- (F) Latency of selective responses from electrodes using power in the 70-100 Hz (Gamma, blue) frequency bands. Statistical significance measured with the signed-rank test ($P < 10^{-5}$).

in single trials (e.g. Figure 3.3A, 3.4A). These response latencies depended on the sets of features revealed on each trial. In a subset of trials where we presented repetitions of partial objects with one fixed position of bubbles (the ‘Partial Fixed’ condition), the IFP timing was more consistent across trials (Figure 3.4C, right bottom), but the latencies were still longer for partial objects than for whole objects.

To further investigate the role of stimulus heterogeneity, we measured the response latency in each trial by determining when the IFP amplitude exceeded a threshold set as three standard deviations above the baseline activity (Figure 3.3F, 3.4F). The average response latencies in the Whole and Partial condition for the preferred category for the first example electrode were 172 and 264 ms respectively (Figure 3.3F, Wilcoxon rank-sum test, $P < 10^{-6}$). The distribution of response latencies in the Whole condition was highly peaked (Figure 3.3F, 3.4F), whereas the distribution of latencies in the Partial condition showed a larger variation, driven by the distinct visual features revealed in each trial. This effect was not observed in all the electrodes; some electrodes showed consistent, albeit delayed, latencies across trials in the Partial condition (Figure C.3). Across the population, delays were observed in the visual response latencies (Figure 3.6C, rank-sum test, $P < 10^{-15}$), even when the latencies were measured with only the most selective responses (Figure C.5).

We asked whether the observed delays could be related to differences in the IFP response strength or the degree of selectivity by conducting an analysis of covariance (ANCOVA). The latency difference between conditions was significant even when accounting for differences in IFP amplitude ($P < 10^{-9}$) or strength of selectivity ($P < 10^{-8}$). Additionally, subpopulations of electrodes with matched-amplitude or matched-selectivity still showed significant differences in the selectivity latency (Figure 3.6D and Figure 3.6E).

Even though the average amplitudes were similar for whole and partial objects (Figure 3.5A), the variety of partial images could include a wider distribution with weak stimuli that failed to elicit a response. To further investigate whether such potential weaker responses could contribute to the la-

tency differences, we performed two additional analyses. First, we subsampled the trials containing partial images to match the response amplitude distribution of the whole objects for each category. Second, we identified those trials where the decoder was correct at 500 ms and evaluated the decoding dynamics before 500 ms under these matched performance conditions. The selectivity latency differences between partial and whole objects remained when matching the amplitude distribution or the decoding performance ($P < 10^{-5}$, Figure C.3,C-D; $P < 10^{-7}$, Figure C.3,E-G).

Differences in eye movements between whole and partial conditions could potentially contribute to latency delays. We minimized the impact of eye movements by using a small stimulus size (5 degrees), fast presentation (150 ms) and trial order randomization. Furthermore, we recorded eye movements along with the neural responses in two subjects. There were no clear differences in eye movements between whole versus partial objects in these two subjects (Figure C.4), and those subjects contributed 5 of the 22 selective electrodes in the Main experiment. To further characterize the eye movements that subjects typically make under these experimental conditions, we also recorded eye movements from 20 healthy volunteers and found no difference in the statistics of saccades and fixation between Whole and Partial conditions (Figure C.4; note that these are not the same subjects that participated in the physiological experiments).

Several studies have documented visual selectivity in different frequency bands of the IFP responses including broadband and gamma band signals (Davidesco et al., 2013; Vidal et al., 2010; Liu et al., 2009). We also observed visually selective responses in the 70-100 Hz Gamma band (e.g. Figure C.1). Delays during the Partial condition documented above for the broadband signals were also observed when measuring the selectivity latency in the 70-100 Hz frequency band (median latency difference = 157 ms, $n = 14$ electrodes, Figure 3.6F).

3.4 Population analysis

To compare delays across different brain regions and different subjects, we mapped each electrode onto the same reference brain. Delays in the response latency between Partial and Whole conditions had a distinct spatial distribution: most of the delays occurred in higher visual areas such as the fusiform gyrus and inferior temporal gyrus (Figure 3.7A). There was a significant correlation between the latency difference and the electrode position along the anterior-posterior axis of the temporal lobe (Spearman's correlation = 0.43, permutation test, $P = 0.02$). In addition, the latency difference was smaller for electrodes in early visual areas (occipital cortex) versus late visual areas (temporal lobe), as shown in Figure 3.7B ($P = 0.02$, t-test). For the two gyri where we had $n > 5$ electrodes selective in both conditions, delays were more prominent in the Fusiform Gyrus than the Inferior Occipital Gyrus ($P = 0.01$, t-test).

The analyses presented thus far only measured selectivity latency for individual electrodes, but the subject has access to activity across many regions. To estimate the selectivity latency from activity across different regions, we combined information from multiple electrodes and across subjects by constructing pseudopopulations (Hung et al., 2005). For each trial, electrode responses were randomly sampled without replacement from stimulus-matched trials (same exemplar and condition) and then concatenated to produce one response vector for each pseudopopulation trial (Methods). This procedure involves several assumptions including independence and ignores potentially important correlations between electrodes within a trial (Meyers and Kreiman, 2011). Electrodes were rank-ordered based on their individual decoding performance, and varying population sizes were examined. Decoding performance using electrode ensembles was both fast and accurate (Figure 3.7C). Category information emerged within 150 ms for whole objects (black thick line) and 260 ms for partial objects (gray thick line), and reached 80% and 45% correct rate, respectively (chance = 20%). Even for the more difficult problem of identifying the stimulus exemplar (chance = 4%), decoding

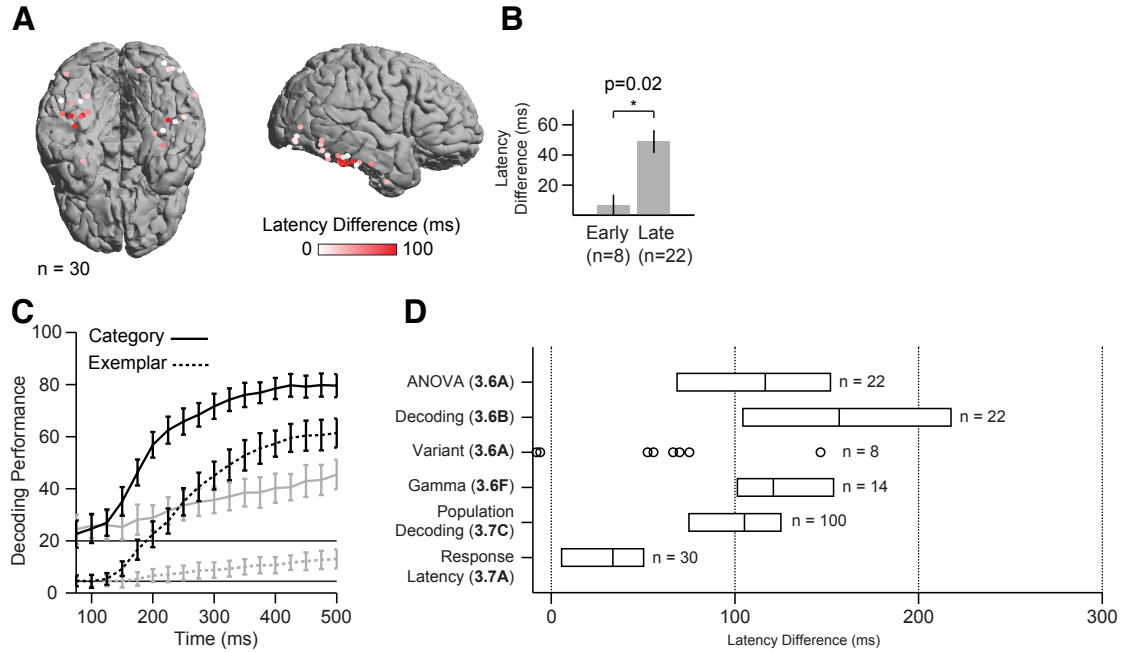


Figure 3.7: Summary of latency measurements

- (A) Brain map of electrodes selective in both conditions, colored by the difference in the response latency (Partial – Whole; see color scale on the bottom).
- (B) Comparison of response latency differences between electrodes in occipital (early visual) and temporal (late visual) lobes.
- (C) Decoding performance from pseudopopulation of 60 electrodes for categorization (thick lines) or exemplar identification (dotted lines) for Whole (black) or Partial (gray) conditions. Horizontal lines indicate chance for categorization (20%) and identification (4%). Error bars represent standard deviation.
- (D) Summary of latency difference (Partial-Whole) for various definitions of latency. Box plots represent the median and quartile. For the Variant experiment, individual electrodes are plotted since the number of electrodes n is small. For the Population decoding results, n denotes the number of repetitions using 60 electrodes.

performance emerged within 135 ms for whole objects (black dotted line) and 273 ms for partial objects (gray dotted line). Exemplar decoding accuracy reached 61% for whole objects and 14% for partial objects. These results suggest that, within the sampling limits of our techniques, electrode ensembles also show delayed selectivity for partial objects.

In sum, we have independently applied several different estimates of latency that use statistical

(ANOVA), machine learning (Decoding), or threshold (Response latency) techniques. These latency measures were estimated using information derived from both broadband signals and specific frequency bands, using individual electrodes as well as electrode ensembles, taking into account changes in contrast, signal strength and degree of selectivity. Each definition of latency requires different assumptions and emphasizes different aspects of the response, leading to variations in the absolute values of the latency estimates. Yet, independently of the specific definition, the latencies for partial objects were consistently delayed with respect to the latencies to whole objects (the multiple analyses are summarized in Figure 3.7D, see also Figure C.5).

3.5 Discussion

The visual system must maintain visual selectivity while remaining tolerant to a myriad of object transformations. This study shows that neural activity in the human occipitotemporal cortex remained visually selective (e.g. Figure 3.3) even when limited partial information about each object was presented (on average, only 18% of each object was visible). Despite the trial-to-trial variation in the features presented, the field potential response waveform, amplitude and object preferences were similar between the Whole and Partial conditions (Figures 3.3-3.4). However, the neural responses to partial objects required approximately 100 ms of additional processing time compared to whole objects (Figures 3.6). While the exact value of this delay may depend on stimulus parameters and task conditions, the requirement for additional computation was robust to different definitions of latencies including single-trial analyses, different frequency bands and different statistical comparisons (Figure 3.7D) and persisted when accounting for changes in image contrast, signal strength, and the strength of selectivity (Figure 3.6). This additional processing time was more pronounced in higher areas of the temporal lobe including inferior temporal cortex and the fusiform gyrus than in earlier visual areas (Figure 3.7A).

Previous human neuroimaging, scalp electroencephalography, and intracranial field potentials recordings have characterized object completion by comparing responses to occluded objects with feature-matched scrambled counterparts (Lerner et al., 2004; Sehatpour et al., 2008). Taking a different approach, neurophysiological recordings in the macaque inferior temporal cortex have examined how robust shape selectivity or encoding of diagnostic features are to partial occlusion (Issa and DiCarlo, 2012; Kovacs et al., 1995a; Missal et al., 1997; Nielsen et al., 2006a). Comparisons across species (monkeys versus humans) or across different techniques (intracranial field potential recordings versus fMRI) have to be interpreted with caution. However, the locations where we observed selective responses to partial objects, particularly inferior temporal cortex and fusiform gyrus (Figure 3.5E-F), are consistent with and provide a link between macaque neurophysiological recordings of selective responses and human neuroimaging of the signatures of object completion.

Presentation of whole objects elicits rapid responses that show initial selectivity within 100 to 200 ms after stimulus onset (Hung et al., 2005; Keysers et al., 2001; Liu et al., 2009; Thorpe et al., 1996; Optican and Richmond, 1987). The speed of the initial selective responses is consistent with a largely bottom-up cascade of processes leading to recognition (Deco and Rolls, 2004; Fukushima, 1980; Riesenhuber and Poggio, 1999; Rolls, 1991). For partial objects, however, visually selective responses were significantly delayed with respect to whole objects (Figures 3.6). These physiological delays are inconsistent with a purely bottom-up signal cascade, and stand in contrast to other transformations (scale, position, rotation) that do not induce additional neurophysiological delays (Desimone et al., 1984; Ito et al., 1995; Logothetis et al., 1995; Logothetis and Sheinberg, 1996; Liu et al., 2009).

Delays in response timing have been used as an indicator for recurrent computations and/or top-down modulation (Buschman and Miller, 2007; Keysers et al., 2001; Lamme and Roelfsema, 2000; Schmolesky et al., 1998). In line with these arguments, we speculate that the additional processing time implied by the delayed physiological responses can be ascribed to recurrent computations that

rely on prior knowledge about the objects to be recognized (Ahissar and Hochstein, 2004). Horizontal and top-down projections throughout visual cortex could instantiate such recurrent computations (Callaway, 2004; Felleman and Van Essen, 1991). Several areas where such top-down and horizontal connections are prevalent showed selective responses to partial objects (Figure 3.5E-F).

It is unlikely that these delays were due to the selective signals to partial objects propagating at a slower speed through the visual hierarchy in a purely feed-forward fashion. Selective electrodes in earlier visual areas did not have a significant delay in the response latency, which argues against latency differences being governed purely by low-level phenomena. Delays in the response latency were larger in higher visual areas, suggesting that top-down and/or horizontal signals within those areas of the temporal lobe are important for pattern completion (Figure 3.7A). Additionally, feedback is known to influence responses in visual areas within 100-200 ms after stimulus onset, as evidenced in studies of attentional modulation that involve top-down projections (Davidesco et al., 2013; Lamme and Roelfsema, 2000; Reynolds and Chelazzi, 2004). Those studies report onset latencies of feedback effects similar to the delays observed here in the same visual areas along the ventral stream. Cognitive effects on scalp EEG responses that presumably involve feedback processing have also been reported at similar latencies (Schyns et al., 2007).

The selective responses to partial objects were not exclusively driven by a single object patch (Figure 3.3A-B, 3A-B). Rather, they were tolerant to a broad set of partial feature combinations. While our analysis does not explicitly rule out common features shared by different images with largely non-overlapping pixels, the large fraction of trials with images with low overlap that elicited robust and selective responses makes this explanation unlikely (Figure 3.5D). The response latencies to partial objects were dependent on the features revealed: when we fixed the location of the bubbles, the response timing was consistent from trial to trial (Figure 3.4C).

The distinction between purely bottom-up processing and recurrent computations confirms predictions from computational models of visual recognition and attractor networks. Whereas recogni-

tion of whole objects has been successfully modeled by purely bottom-up architectures (Riesenhuber and Poggio, 1999; Rolls, 1991), those models struggle to identify objects with only partial information (Johnson and Olshausen, 2005; O'Reilly et al., 2013). Instead, computational models that are successful at pattern completion involve recurrent connections (Hopfield, 1982; Lee and Mumford, 2003; O'Reilly et al., 2013). Different computational models of visual recognition that incorporate recurrent computations include connections within the ventral stream (e.g. from ITC to V4) and/or from pre-frontal areas to the ventral stream. Our results implicate higher visual areas (Figure 3.5E) as participants in the recurrent processing network involved in recognizing objects from partial information. Additionally, the object-dependent and unimodal distribution of response latencies to partial objects (e.g. Figure 3.3F) suggest models that involve graded evidence accumulation as opposed to a binary switch.

The current observations highlight the need for dynamical models of recognition to describe where, when and how recurrent processing interacts with feed-forward signals. Our findings provide spatial and temporal bounds to constrain these models. Such models should achieve recognition of objects from partial information within 200 to 300 ms, demonstrate delays in the visual response that are feature-dependent, and include a graded involvement of recurrent processing in higher visual areas. We speculate that the proposed recurrent mechanisms may be employed not only in the context of object fragments but also in visual recognition for other types of transformations that impoverish the image or increase task difficulty. The behavioral and physiological observations presented here suggest that the involvement of recurrent computations during object completion, involving horizontal and top-down connections, result in a representation of visual information in the highest echelons of the ventral visual stream that is selective and robust to a broad range of possible transformations.

3.6 Methods

After 500 ms of fixation, subjects were presented with an image (256x256 pixels) of an object for 150 ms, followed by a 650 ms gray screen, and then a choice screen (Figure 3.1A). The images subtended 5 degrees of visual angle. Subjects performed a 5-alternative forced choice task, categorizing the images into one of five categories (animals, chairs, human faces, fruits, or vehicles) by pressing corresponding buttons on a gamepad (Logitech, Morges, Switzerland). No correct/incorrect feedback was provided. Stimuli consisted of contrast-normalized grayscale images of 25 objects, 5 objects in each of the aforementioned 5 categories (Figure 3.1C). In approximately 30% of the trials, the images were presented unaltered (the ‘Whole’ condition). In 70% of the trials, the visual features were presented through Gaussian bubbles of standard deviation 14 pixels (the ‘Partial condition, see example in Figure 3.1B) (Gosselin and Schyns, 2001). The more bubbles, the more visibility. Each subject was first presented with 40 trials of whole objects, then 80 calibration trials of partial objects, where the number of bubbles was titrated in a staircase procedure to set the task difficulty at $\sim 80\%$ correct rate. The number of bubbles was then kept constant throughout the rest of the experiment. The average percentage of the object shown for each subject is reported in Figure 3.2. Unless otherwise noted (below), the positions of the bubbles were randomly chosen in each trial. The trial order was pseudo-randomized.

Six subjects performed a variant of the main experiment with three key differences. First, contrast was normalized between the Whole and Partial conditions by presenting all objects in a phase-scrambled background (Figure 3.1B). Second, in 25% of the Partial condition trials, the spatial distribution of the bubbles was fixed to a single seed (the ‘Partial Fixed’ condition). Each of the images in these trials was identical across repetitions. Third, because experimental time was limited, only objects from three categories (animals, human faces and vehicles) were presented to collect enough trials in each condition.

3.6.1 Electrode localization

Electrodes were localized by co-registering the preoperative magnetic resonance imaging (MRI) with the postoperative computer tomography (CT) (Destrieux et al., 2010; Liu et al., 2009) . For each subject, the brain surface was reconstructed from the MRI and then assigned to one of 75 regions by Freesurfer. Each surface was also co-registered to a common brain for group analysis of electrode locations. The location of electrodes selective in both Whole and Partial conditions is shown in Table C.2. In Figure 3.7A, we computed the Spearman's correlation coefficient between the latency differences (Partial - Whole) and distance along the posterior-anterior axis of the temporal lobe. In Figure 3.5F, we partitioned the electrodes into three groups: Fusiform Gyrus, Inferior Occipital Gyrus, and Other. We used the Fisher's exact test to assess whether the proportion of electrodes selective in both conditions is greater in the Fusiform Gyrus versus Other, and in Inferior Occipital Gyrus versus Other.

3.6.2 Preprocessing

The signal from each electrode was amplified and filtered between 0.1 and 100 Hz with sampling rates ranging from 256 Hz to 1000 Hz at CHB (XLTEK, Oakville, ON, Canada), BWH (Bio-Logic, Knoxville, TN, USA) and JHMI (Natus, San Carlos, CA and Nihon Kohden, Tokyo, Japan). A notch filter was applied at 60 Hz. All the data were collected during periods without any seizure events. All studies described here were approved by each hospital's institutional review boards and were carried out with the subjects' informed consent.

3.6.3 Selectivity measures

All analyses in this manuscript used correct trials only. Noise artifacts were removed by omitting trials where the intracranial field potential (IFP) amplitude exceeded five times the standard deviation.

The responses from 50 to 500 ms after stimulus onset were used in the analyses.

ANOVA

We performed a non-parametric one-way analysis of variance (ANOVA) of the IFP responses. For each time bin, the F-statistic (ratio of variance across object categories to variance within object categories) was computed on the IFP response (Keeping, 1995). Electrodes were denoted ‘selective’ in this test if the F-statistic crossed a threshold (described below) for 25 consecutive milliseconds (e.g. Figure 3.3D). The latency was defined as the first time of this threshold crossing. The number of trials in the two conditions (Whole and Partial) was equalized by random subsampling; 100 subsamples were used to compute the average F-statistic. A value of 1 in the F-statistic indicates no selectivity (variance across categories comparable to variance within categories) whereas values above 1 indicate increased selectivity.

Decoding

We used a machine learning approach to determine if, and when, sufficient information became available to decode visual information from the IFP responses in single trials (Bishop, 1995). For each time point t , features were extracted from each electrode using Principal Component Analysis (PCA) on the IFP response from $[50 t]$ ms, and keeping those components that explained 95% of the variance. The features set also included the IFP range (max – min), time to maximum IFP, and time to minimum IFP. A multi-class linear discriminant classifier with diagonal covariance matrix was used to either categorize or identify the objects. Ten-fold stratified cross-validation was used to separate the training sets from the test sets. The proportion of trials where the classifier was correct in the test set is denoted the ‘Decoding Performance’ throughout the text. In the Main experiment, a decoding performance of 20% (1/5) indicates chance for categorization and 4% (1/25) indicates chance for identification. The number of trials in the Whole and Partial conditions was equalized by

subsampling; we computed the average Decoding Performance across 100 different subsamples. An electrode was denoted 'selective' if the decoding performance crossed a threshold (described below) at any time point t , and the latency was defined as the first time of this threshold-crossing.

Pseudopopulation

Decoding performance was also computed from an ensemble of electrodes across subjects by constructing a pseudopopulation, and then performing the same analyses described above (Figure 3.7C). The pseudopopulation pooled responses across subjects (Hung et al., 2005; Mehring et al., 2003; Pappas and Connor, 2002). The features for each trial in this pseudopopulation were generated by first randomly sampling exemplar-matched trials without replacement for each member of the ensemble, and then concatenating the corresponding features. The pseudopopulation size was set by the minimum dataset size of the subject, which in our data was 100 trials (4 from each exemplar). Because of the reduced data set size, four-fold cross-validation was used.

d-prime

We compared the above selectivity metrics against d' (Green and Swets, 1966). The value of d' was computed for each electrode by comparing the best category against the worst category, as defined by the average IFP amplitude. d' measures the separation between the two groups normalized by their standard deviation. The latency of selectivity for d' was measured using the same approach as the ANOVA (Figure C.3A-B).

Significance Thresholds

The significance thresholds for ANOVA, Decoding and d' , were determined by randomly shuffling the category labels 10,000 times, and using the value of the 99.9 percentile (ANOVA: $F = 5.7$, Decoding: 23%, $d' = 0.7$). This represents a false positive rate of 0.1% for each individual test. As

discussed in the text, we further restricted the set of electrodes by considering the conjunction of the ANOVA and Decoding tests. We evaluated this threshold by performing an additional 1,000 shuffles and measuring the number of selective electrodes that passed the same selectivity criteria by chance. In Table 3.1, we present the number of electrodes that passed each significance test and the number of electrodes that passed the same tests after randomly shuffling the object labels. The conclusions of this study did not change when using a less strict criterion of $q = 0.05$ (median latency difference for ANOVA: 123 ms, $n = 45$ electrodes selective in both conditions, Figure C.5).

3.6.4 Latency measures

We considered several different metrics to quantify the selectivity latency (i.e. the first time point when selective responses could be distinguished), and the visual response latency (i.e. the time point when a visual response occurred). These measures are summarized in Figure 3.7D and Figure C.5.

Selectivity latency

The selectivity latency represented the first time point when different stimuli could be discriminated and was defined above for the ANOVA, Decoding and d' analyses.

Response Latency

Latency of the visual response was computed at a per-trial level by determining the time, in each trial, when the IFP amplitude exceeded 3 standard deviations above the baseline activity. Only trials corresponding to the preferred category were used in the analysis. To test the multimodality of the distribution of response latencies, we used Hartigan's dip test. In 27 of the 30 electrodes, the unimodality null hypothesis could not be rejected ($P > 0.05$).

3.6.5 Other analyses

For each pair of partial object trials, the percent of overlap was computed by dividing the number of pixels that were revealed in both trials by the area of the object (Figure 3.5D). Because low degree of object overlap would be expected in trials with weak physiological responses, we focused on the most robust responses for these analyses by considering those trials when the IFP amplitude was greater than the 90th percentile of the distribution of IFP amplitudes of all the non-preferred category trials.

4

Backward masking and object completion

In the previous chapter, we presented neurophysiological evidence for the involvement of recurrent computations in the recognition of occluded objects. We therefore hypothesized that disrupting these recurrent computations would degrade recognition performance, but only for occluded objects and not for whole objects. To test this hypothesis, we performed a series of psychophysical experiments on healthy volunteers with backward masking.

In backward masking, an image is first presented on the screen for typically 10-100ms, denoted as the stimulus onset asynchrony (SOA), then immediately followed with a spatially overlapping noise pattern. At very short SOAs (<25 ms), backward masking has been shown to render the preceding image invisible (Breitmeyer and Ogmen, 2000; Op de Beeck et al., 2007). For intermediate SOAs (25-100 ms), however, backward masking is thought to disrupt recurrent computations while leaving

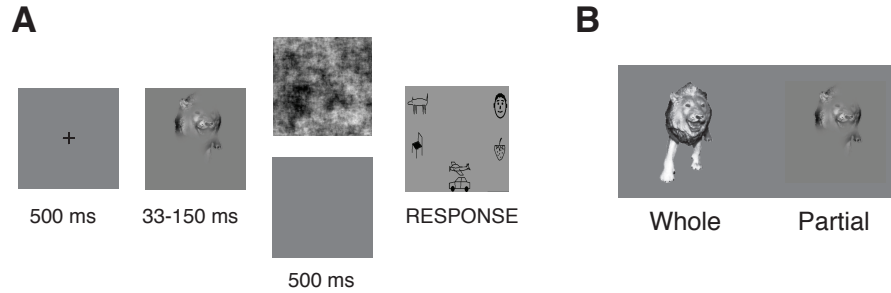


Figure 4.1: **Main psychophysics experiment.**

- (A) Subjects performed a forced-choice categorization task. After 500ms of fixation, stimuli were presented for variable exposure times (33 to 150 ms), denoted as the stimulus onset asynchrony (SOA). The image was followed by either a noise mask or a gray screen for 500 ms.
- (B) Stimuli were either presented unaltered ('Whole' condition) or rendered partially visible ('Partial' condition) by presenting features through Gaussian bubbles (Gosselin and Schyns, 2001).

the feed-forward stream intact (Lamme and Roelfsema, 2000; Serre et al., 2007b; Wyatte et al., 2012).

This conjecture is supported by several neurophysiological studies in macaques (Kovacs et al., 1995b; Rolls et al., 1999). In the Kovacs et al (1995) study, the authors demonstrated that pattern backward masking did not affect the firing rate or selectivity of the initial response of neurons in macaque IT (which we attribute to the initial feed-forward sweep of activity). However, the masking disrupted the neural representation at later times, thus shortening the length of the neural response. If feedback from a higher area (such as prefrontal cortex) attempted to interact with this residual activity, there would be a mismatch. Alternatively, if the recurrent computations reflected in this later representation in IT are important for recognition, backward masking would disrupt this information content as well. At longer SOAs (>150 ms), processing may already be complete, and the backward mask would have a minimal effect on recognition.

To examine the effect of disrupting recurrence, we probe this critical intermediate period with a series of experiments. In the main version of this experiment, subjects performed a categorization task where images are presented for a variable exposure time (33-150 ms). The images are followed with either a gray screen (unmasked) or a noise pattern (masked) for 500 ms (Figure 4.1). The stimuli

Experiment	# Categories	# Exemplars	Dataset	Task	Notes
Main	4	16	klab16	4-AFC	Partial images
Occluded	4	16	klab16	4-AFC	Occluded images
KLAB325	5	325	klab325	5-AFC	No exemplar repetitions
Phys	5	25	phys25	5-AFC	Physiology trials
PhysRT	5	25	phys25	2-AFC	Reaction Time

Table 4.1: **Psychophysics experiments**

Main In the main experiment, we used 16 stimuli from four categories (klab16 dataset).

Occluded To assess the effect of occluders, we ran a similar version with occluded images.

KLAB325 The previous experiments have very few exemplars, so the same exemplar is repeatedly presented under different occlusion patterns, possibly enabling learning effects for particular image fragments. Therefore, we constructed an expanded set of images and categories and designed an experiment without exemplar repetitions (klab325 dataset).

Phys In order to link behavior with neural activity, we designed a dataset consisting of images from which we have obtained neurophysiological responses (phys25 dataset). The exemplars and categories in this set are identical to those used in the intracranial recordings.

PhysRT We obtained reaction time measurements for each image in the phys25 dataset. For accurate reaction time measurements, we used a target/no-target task (see Methods) and the images were unmasked.

were either unaltered (‘Whole’) or rendered partially visible by presenting features through Gaussian bubbles (‘Partial’). Similar to the experiment described in the previous chapter, an initial calibration session was used to adjust the number of bubbles to reach 80% performance (see Methods). The number of bubbles was kept constant throughout the rest of the experiment, but the bubble locations were randomized. In the main experiment, the stimuli consisted of 16 objects belonging to four categories (animals, chairs, faces, vehicles). Throughout this chapter, images refer to particular combination of an object and bubble locations. For example, one object can be used to generate multiple occluded images.

We also designed several variant experiments that will be introduced throughout the course of this chapter. A summary is shown in Table 4.1 for reference.

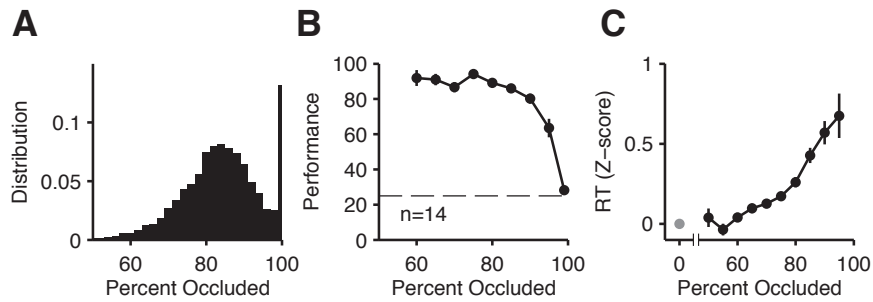


Figure 4.2: **Performance was robust to occlusion.**

- (A) Distribution of percentage of the object that was occluded for the trials in the main experiment. Bin size = 2%, last bar is >99% occlusion.
- (B) Performance as a function of the percent of the object occluded in the main experiment. Images were unmasked. Chance is 25% (dotted black line). Error bars indicate s.e.m. $n=14$ subjects.
- (C) In a task variant, subjects were instructed to respond as fast as possible in a two-alternative forced choice and we measured reaction times (see Methods and Table 4.1). To compare across subjects the reaction time (RT) data were normalized and shown here as a function of the percent occlusion. For each subject, the RT was normalized to the statistics of the RT to whole objects (gray dot).

4.1 Results

We first established a baseline of human performance on recognition of objects from partial information. Because the bubble locations were randomized from trial to trial, the partial images spanned a range of difficulty (Figure 4.2A). We measured performance against percent occlusion for ‘Partial unmasked trials (Figure 4.2B). Human performance was robust even in heavy occlusion (e.g. 60% performance at 95-99% occlusion, where chance is 25%). Because the backward masking and fixed delay period did not allow us to accurately measure reaction time, we also ran a separate reaction time task with these same type of images (Methods). Consistent with previous studies on simple geometric shapes (Shore and Enns, 1997), the reaction time was delayed for more occluded images (Figure 4.2C, one standard deviation is approximately 170 ms). These results suggest a robust recognition capability in the intact human visual system.

We then examined performance in trials with backward masking where recurrence or feedback

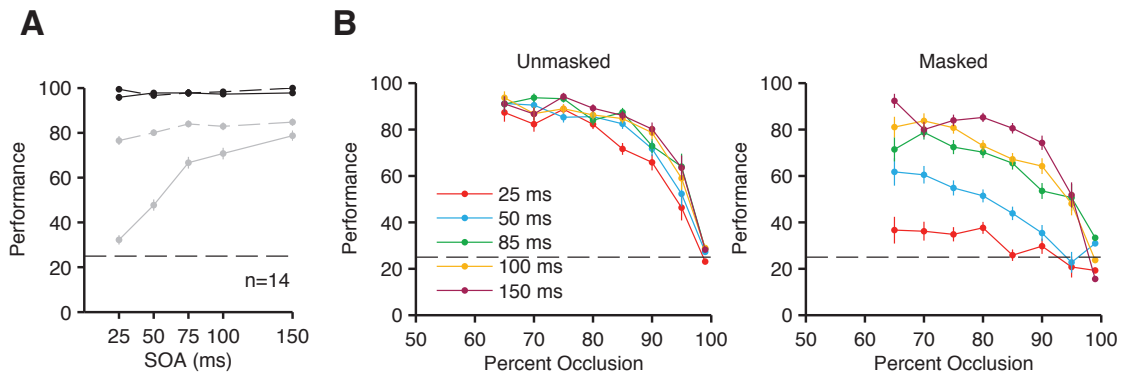


Figure 4.3: **Backward masking disrupts performance.**

- (A) Recognition performance as a function of the stimulus onset asynchrony (SOA) for whole objects (black) or partial objects (gray). Solid lines indicate the masking condition and dashed lines indicate the unmasked condition. For partial images, performance was significantly degraded by masking (solid gray line) compared to the unmasked trials (dotted gray line). However, performance on whole images (black lines) was not affected by backward masking. Horizontal dashed line indicates chance level (25%). Error bars indicating s.e.m. are too small to be visible.
- (B) Performance across different amounts of occlusion for the unmasked (left panel) and masked (right panel) trials. Different colors mark the different SOAs.

may be disrupted. For whole objects, performance was near ceiling regardless of SOA and masking (Figure 4.3A, compare black solid and dotted lines), which is consistent with the theory that recognition of whole objects is mediated by the initial feed-forward sweep (Serre et al., 2007b). However, when partial images were followed with a backward mask, performance was significantly degraded (Figure 4.3A, gray solid line) compared to unmasked performance (gray dotted line). We performed a two-way ANOVA on performance with SOA and Masking as factors and found a significant interaction ($F(4) = 28.2, P < 10^{-17}$). Whereas performance on unmasked trials was not significantly affected by the SOA, the effect of backward masking was strongly dependent on SOA. Backward masking disrupted performance across a wide range of image difficulty, as demonstrated in Figure 4.3 (compare unmasked (left) versus masked (right) panels across a range of occlusion amounts).

We also measured the effect of backward masking in two experimental variants (Figure 4.4). Previous literature has suggested that the presence of occluders might improve recognition by in-

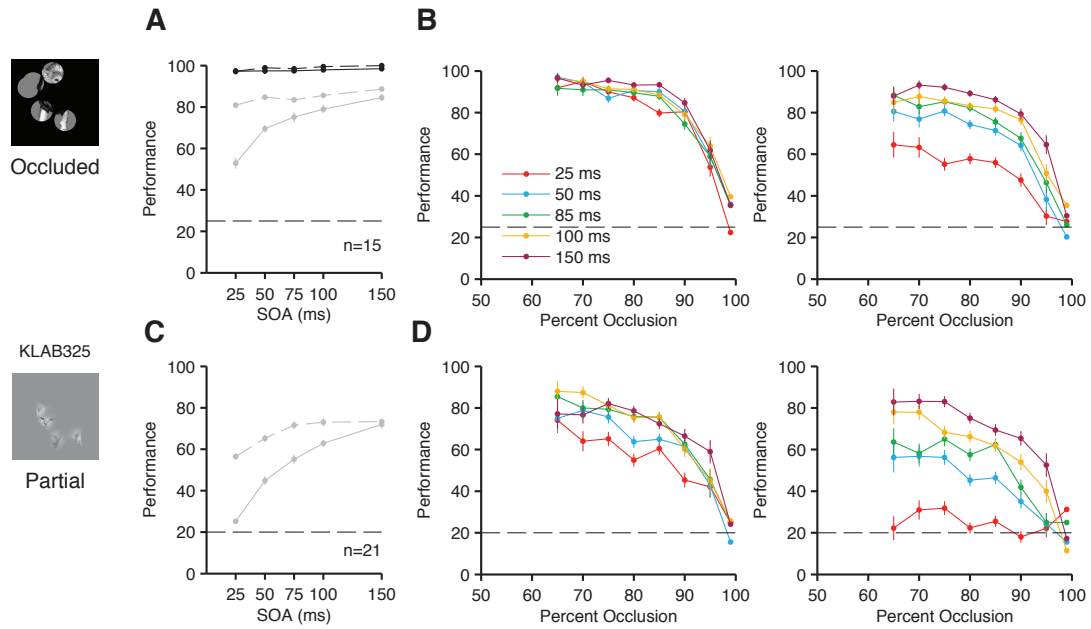


Figure 4.4: Effect of backward masking was observed in variant experiments.

Same conventions as Figure 4.3, but for two variant experiments.

- (A) Performance for different SOAs during the Occluded experiment.
- (B) Performance as a function of SOA and percent occlusion during the Occluded experiment.
- (C) Performance for different SOAs during the KLAB325 experiment. Note that chance here is 20% (five-way categorization).
- (D) Performance as a function of SOA and percent occlusion during the KLAB325 experiment.

ducing amodal completion mechanisms (see Chapter 1, (Johnson and Olshausen, 2005; Bregman, 1981)). When the objects were presented behind an occluding shape (Occluded experiment), performance was slightly higher, but the effect of backward masking persisted ($F(4) = 14.8, P < 10^{-9}$, ANOVA). One concern with the *klab16* stimulus set is that the same object is shown repeatedly under different occluding patterns, possibly leading to memorization of specific exemplar features. Therefore, we designed an expanded stimulus set of 325 objects belonging to five categories (*klab325* set), and performed a variant experiment where each object exemplar is only presented once with masking and once without masking (KLAB325 experiment). The other importance difference is that, in

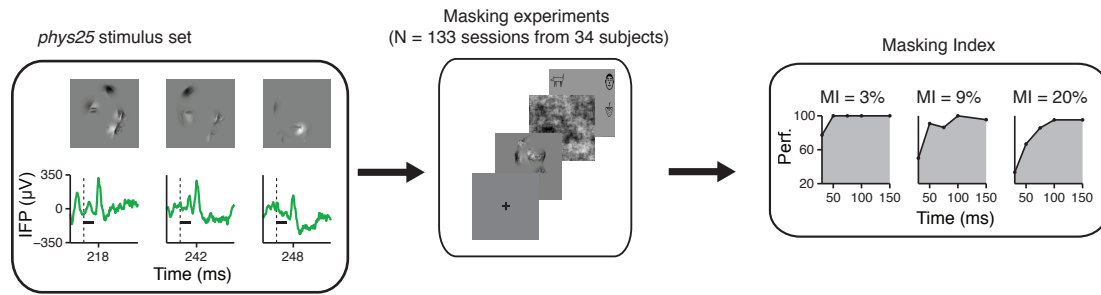


Figure 4.5: **Outline of experiment linking neural responses with backward masking**

We first constructed the *phys25* stimulus set consisting of $n = 650$ partial images where we have recorded neural responses (left). This set of images was then used in extensive backward masking experiments (middle) to obtain, for each partial image, a curve of performance against SOA (right). We defined the masking index (MI) as $1 - AUC$, where AUC is the area under the curve (gray region) divided by total area.

contrast to the previous experiments, subjects were never shown the corresponding whole object. While performance was lower compared to the main experiment, the backward masking effect was still consistent in this non-repeating stimulus set ($F(4) = 13.5$, $P < 10^{-8}$, ANOVA).

If backward masking were interrupting recurrence, then partial images that elicited slower neurophysiological responses would also be more affected by the backward mask. To test this hypothesis, we ran a backward masking experiment with partial images where we had previously recorded neural responses from epilepsy patients (Chapter 3). Due to feasibility constraints, we cannot collect backward masking data on all the images shown in those experiments. We therefore first constructed the *phys25* stimulus set consisting of $n = 650$ trials from $n = 2$ electrodes that responded selectively to partial images (see Methods). Figure 4.6A illustrates one electrode from fusiform gyrus that responded preferentially to faces (green curve). When partially visible images were presented to this electrode, the responses were tolerant and selective (Figure 4.6B, first and second row). From each electrode, we selected images from $n = 325$ trials for psychophysical experiments. These partial images were selected such that they elicited strong responses, and represented a wide range of response latencies (Figure 4.6C).

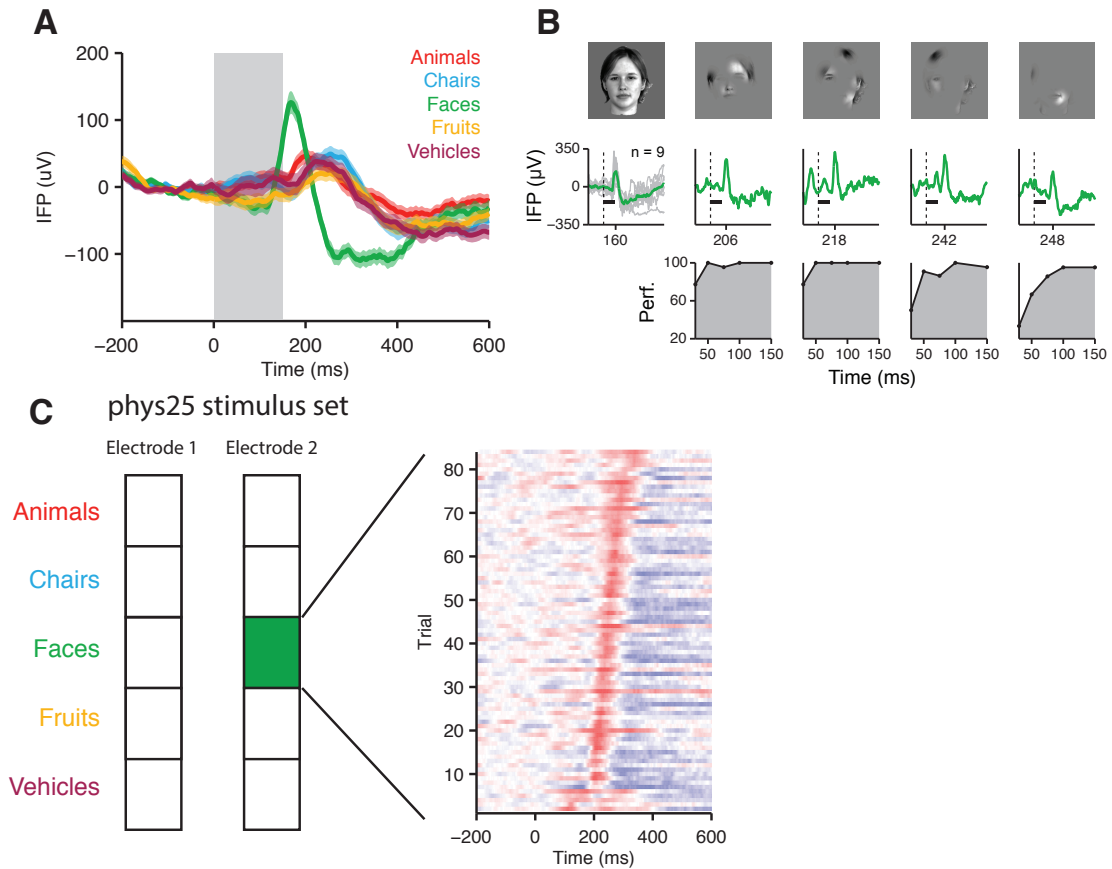


Figure 4.6: *phys25* stimulus set

- (A) Intracranial recordings while the subject performed a five-way categorization task (see Figure 3.3). For an electrode in the left Fusiform gyrus, here we show the average intracranial field potential (IFP) of each object category for whole objects. This electrode preferred faces (green line).
- (B) For the same electrode in (A), the IFP responses for an exemplar object. For the whole condition, the single trial responses (gray, $n = 9$) and average response (green) are shown. For the partial condition, single-trial responses (green, $n = 1$) to several partial images of the same object are shown. The latency of the peak response is marked on the x-axis. For each partial exemplar image, we conducted a separate psychophysics experiment to measure the effect of backward masking at various SOAs (bottom row).
- (C) The *phys25* stimulus set consisted of $n = 650$ trials from all five categories drawn from two electrodes. The right shows a raster of the neural responses from partial trials of the preferred category (faces) that were selected for this dataset. These trials elicited strong responses at a wide range of latencies (200-300 ms). Responses are sorted by response latency in this raster.

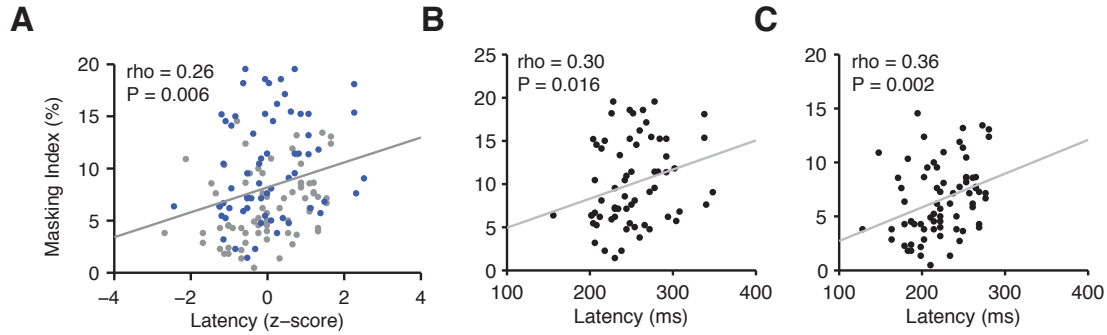


Figure 4.7: **Correlations between neural response latency and masking effect**

- (A) Partial images that elicited slower neural responses were also more susceptible to the disruptions caused by the backward mask. Data from the two electrodes were combined by standardizing the latency. Statistical significance was assessed by regressing latency against masking index with electrode number and percent occlusion as additional factors. Even when controlling for low-level effects and inter-electrode variability, the masking index was a significant predictor ($P = 0.006$).
- (B) For the first electrode in the *phys25* stimulus set, the correlation between the physiological response latency (x-axis) and the effect of backward mask, quantified as the masking index (y-axis). Each dot is a single partial image from the preferred category. There was a significant correlation ($r = 0.30$, $P = 0.016$, Pearson's correlation).
- (C) For the second electrode in the *phys25* stimulus set, we also found a significant correlation between latency and the masking index ($r = 0.36$, $P = 0.002$).

After constructing this stimulus set, We presented the images to psychophysics subjects at various SOAs with backward masking (n=132 sessions from 33 subjects). This allowed us to construct, for each of the selected images from the neurophysiology experiment, a performance curve of backward masking versus SOA (see Figure 4.6B, third row). To quantify the effect of masking, we defined the masking index (MI) as 1-AUC, where AUC is the normalized area under the curve (gray area in Figure 4.6B divided by the entire area). The greater the MI, the larger the effect of backward masking on this particular image.

For partial images from the preferred category, the masking index correlated with the response latency (Figure 4.7A, Pearson's correlation coefficient = 0.26). To determine statistical significance, we performed the following regression:

$$\text{latency} \sim 1 + \text{masking_index} + \% \text{occlusion} + \text{electrode}.$$

The inclusion of the factors $\% \text{occlusion}$ and electrode controls for low-level confounds and inter-electrode variability. The masking index was a significant predictor of the neural response latency in this regression ($P = 0.006$). The correlations were also statistically significant in each electrode individually (Figure 4.7B-C, $P = 0.02$ and $P = 0.002$, respectively). As expected, response latency was not significantly correlated with the masking index for the non-preferred categories ($P = 0.62$ and $P = 0.08$, respectively).

4.2 Discussion

We have established with our psychophysical experiments the robustness of occluded object recognition. Our observation that reaction time increases with the amount of occlusion is intuitive, and extends the findings from geometric shapes to naturalistic objects (Shore and Enns, 1997). Under the supposition that backward masking disrupts recurrence, our results provide behavioral evidence for the necessity of recurrence in object recognition.

As we have noted previously, the role of the backward masking is still a matter of debate. Here we provided neurophysiological evidence that backward masking disrupts recurrence. In particular, partial images that require more neural processing (e.g. longer response latencies) were also more vulnerable to backward masking, even when controlling for low-level factors such as the amount of occlusion. These results are consistent with findings from macaque visual cortex with backward masking (Kovacs et al., 1995b; Rolls et al., 1999).

Because of feasibility constraints, we were not able to measure masking index for all the trials from the physiology experiment. The physiology experiment consisted of 31,130 trials, and measuring masking index on just 650 of those trials required 132 sessions from $n = 33$ psychophysics

subjects. Importantly, we *a priori* selected physiology images to construct the *phys25* stimulus set before running any psychophysics experiments. There was no cherry-picking of the trials post-hoc. Therefore, while we only presented correlations from two electrodes, those were the only electrodes whose images were tested in the psychophysics experiments.

While performance on occluded objects is slightly better than that of partial objects, the differences are only apparent at low SOAs and high amounts of occlusion. This is consistent with (Johnson and Olshausen, 2005) and perhaps counter to the intuition from the famous example of Bregman's Bs (Bregman, 1981). Perhaps for impoverished stimuli in that example, where there is little context to guide the amodal completion process, explicit occluders are important. For more naturalistic stimuli, however, and higher-level tasks such as categorization, they may not be entirely necessary.

The *klab325* stimulus set has two important changes from *klab16*. First, exemplars are never repeated, so subjects are unable to learn specific exemplar fragments. Second, the whole images corresponding to the partial objects are never shown to the subject, which is reflective of natural viewing, where we often encounter novel but occluded objects. While this is a more difficult task, the effect of backward masking persists.

Now that we have provided both neural (Chapter 3) and psychophysical evidence for the role of recurrent computations, we next turn to exploring the computational contribution of this recurrence to object completion in the next chapter.

4.3 Methods

A total of $n = 83$ volunteers with normal or corrected-to-normal vision participated in the psychophysics experiments reported in this study. All subjects gave informed consent, and the studies were approved by the Boston Children's Hospital institutional review board.

4.3.1 Experimental design

Subjects were asked to categorize each image by pressing corresponding buttons on a gamepad. Each trial was initiated by a fixating on a cross for at least 500 ms. After fixation, subjects were presented with the image of an object for a variable time (33 ms, 50 ms, 100 ms, or 150 ms), which we denote as the stimulus onset asynchrony (SOA). The image is followed by either a noise mask or a gray screen for 500 ms, after which a choice screen appears and the subject indicates the response. The image (256 x 256 pixels) subtended approximately 5 degrees of the visual field.

Each subject performed an initial training period to familiarize themselves with the task and the stimuli. They were presented with 40 trials of whole objects, then 80 calibration trials of occluded images. During the calibration trials, the number of bubbles was titrated using a staircase procedure to achieve a task difficulty of 80% correct rate. The number of bubbles (but not their positions) were then kept constant for the rest of the experiment. Results from this familiarization and calibration phase were not included in the analysis. The rest of the experiment consisted of 1,200 trials, with 600 unmasked images, followed by 600 masked images. While the positions of the bubbles were randomly chosen in each trial and the trial order was pseudo-randomized, the same set of images were presented in the masked and unmasked conditions.

In the main experiment ($n = 14$ subjects), stimuli consisted of contrast-normalized grayscale images of 16 objects belonging to four categories (animals, human faces, fruits, or vehicles). The noise mask was generated by scrambling the phase of the images, while retaining the spectral coefficients. In approximately 15% of the trials, the objects were presented unaltered (the ‘Whole’ condition). In the other 85%, the objects were occluded by presenting visual features through Gaussian bubbles (the ‘Partial condition’, standard deviation = 14 pixels, see (Gosselin and Schyns, 2001)).

4.3.2 Variant experiments

Several variants of the main experiment were performed. In the KLAB325 experiment ($n = 21$ subjects), the stimuli were expanded to a set of 325 objects belonging to 5 categories. Importantly, each object was only presented twice to the subject (once in the masked condition, and once in the unmasked condition), so the subject could not memorize the features of particular exemplars. In a second variant ($n = 15$ subjects), an occluding shape was used instead of Gaussian bubbles to generate occluded images (Occluded experiment).

4.3.3 *phys25* stimulus set

In the Phys experiment (132 sessions from $n = 33$ subjects), the exact same partial images shown to neurophysiology subjects were used as stimuli for the psychophysics subjects. The neurophysiological data was previously reported in Chapter 3.

To construct the *phys25* stimulus set, we *a priori* selected 650 trials for psychophysical testing drawn from two electrodes that were visually selective for partial images from the neurophysiology dataset. Only trials where the amplitude of the elicited neural response was in the top 50 percentile were included. To maximize the power of our analysis, trials from the preferred category were selected such that the latency of the neural response, defined as the time of the peak, spanned a long interval. The images from these trials were then used as stimuli for the psychophysics experiments with masking and variable exposure time to construct the masking curves illustrated in Figure 4.6. We collected this data from 132 sessions in $n = 33$ subjects. We were unable to include more trials or electrodes because the resulting length of the psychophysical experiments would be infeasible to collect.

4.3.4 Reaction time experiment

Because the masking experiments had a fixed delay period, we could not compare behavioral reaction times. Therefore, we performed a reaction-time variant of the experiment with the *phys25* stimulus set (PhysRT experiment). To obtain a more accurate measure of the reaction time, we used a two-alternative forced choice task. At the beginning of each block, subjects were cued to a target category. Images were presented for 150 ms (unmasked), and subjects performed a target/non-target task by pressing the right or left button on a gamepad.

5

Computational models of object completion

There has been significant progress over the last decade in developing computational models of object recognition (Deco and Rolls, 2004; DiCarlo et al., 2012; Kreiman, 2013; Riesenhuber and Poggio, 1999; Serre et al., 2007b; Krizhevsky et al., 2012). To a first approximation, these models propose a hierarchical sequence of linear filtering and non-linear pooling operations inspired by the basic principles giving rise to simple and complex cells in primary visual cortex (Hubel and Wiesel, 1962). Concatenating multiple such operations together resulted in some of the initial models for object

This chapter is a product of joint work with William Lotter, who developed the recurrent neural network model described in this chapter and performed some of the computational analysis.

recognition (Fukushima, 1980). Recently, these ideas have also seen wide adoption in the computer science literature in the form of deep convolutional neural networks (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014). Both biologically inspired models and deep convolutional neural networks (CNNs) optimized for performance share similar core architectures.

We have provided neural (Chapter 3) and behavioral (Chapter 4) evidence for recurrent computations in the recognition of occluded objects. In this Chapter, we examine the computational role of this recurrence. We first demonstrate that existing feed-forward networks fail to generate representations that are robust to occlusion. To understand how the brain may solve this problem, we show that correlations between the model representation and our neural response latencies are consistent with an attractor network. As a proof-of-principle, we present a recurrent neural network that significantly improves recognition of occluded objects, and matches the pattern of human performance.

5.1 Performance of feed-forward models in recognizing occluded objects

The canonical steps in feed-forward computational models are inspired by the observation of simple and complex cells in primary visual cortex of anesthetized cats. In their classic study, Hubel and Wiesel discovered 'simple' cells tuned to bars oriented at a particular orientation (Hubel and Wiesel, 1959). They also described 'complex' cells, which were also tuned to a preferred orientation, but exhibited a degree of tolerance to spatial translation of the stimulus. They hypothesized that to generate this spatial invariance, the complex cells pool over simple cells whose receptive fields tile the visual space with a max-like operation. This complex cell would then respond to an oriented bar regardless of its spatial location. Both hierarchical models of biological vision such as HMAX (Riesenhuber and Poggio, 1999; Serre et al., 2007b) and CNNs are composed of alternating layers of tuning and pooling with increasingly more complicated tuning functions as one ascends this hierarchy. Whereas biologically-inspired models such as HMAX have about 2-4 layers, state-of-the-art computer vision

models have moved to complex topologies with up to 20 layers and different mixtures of tuning and pooling layers (Russakovsky et al., 2015). Performance of feed-forward models such as HMAX on object recognition datasets match the pattern of human performance (Serre et al., 2007a). Additionally, the activity of individual layers in deep learning networks can capture the variance of neural firing rates in the corresponding layers of macaque cortex (Cadieu et al., 2014; Yamins et al., 2014).

While impressive, these algorithms have a number of limitations. They are sensitive to image transformations such as occlusion or rotation (Pepik et al., 2015) and can be easily fooled by nonsensical images (Nguyen et al., 2014). To measure the performance of these models on the recognition of occluded objects, we used the *klab325* stimulus set of 325 exemplars belonging to five categories (see Chapter 4). This allowed us to compare the model results against human performance. We tested several different models:

AlexNet-fc7 AlexNet is a convolutional neural network that has been pre-trained on ImageNet, a dataset of 1.2 million high-resolution images (Krizhevsky et al., 2012). Here we used the features from the last fully-connected layer, which we denote as the fc7 layer and has 4096 features.


AlexNet-pool5 Here we used the last pooling layer from AlexNet, the pool5 layer. This layer has 9216 features.

HMAX HMAX has two layers of pooling, with parameters inspired from primate physiology experiments. The final layer of HMAX has 1000 features.

Pixels Using the raw pixels ($256^2 = 65,536$ features) provided a baseline performance with which to compare other computational models, and also measured the effect of low-level differences such as overall contrast.

We assessed the stability of these models to occlusion by measuring how well a classifier trained

on whole objects generalizes to categorizing partial objects. Some may object here that we encounter occluded objects in everyday life, so this requirement may seem artificial. However, having a robust representation would in fact be advantageous since the same decision boundaries can be used for recognition of both whole and occluded objects. Otherwise, our visual system would be constantly switching decision boundaries. As an analogy, when we measure scale invariance in these models, we require that models are able to accurately classify objects at one scale when trained on objects from a different scale (Serre et al., 2007a; Hung et al., 2005). Natural scenes are also composed of objects at various scales, making a scale-invariant representation even more important. Similarly, models that are invariant to occlusion should be able to generalize from whole to occluded objects (see 5.3.2 for an extended discussion).

 We trained the decision boundaries of a support vector machine (SVM) classifier (linear kernel) on whole objects of the *klab325* stimulus set, then measured classification performance on partial images. In order to compare with human performance, the test set contained images from the 13,000 trials shown to human subjects during the KLAB325 experiment (see Chapter 4). Importantly, cross-validation was performed over objects, meaning that the objects used to train the decision boundary did not appear as partial images in the test set.

Compared to human performance, feed-forward models such as HMAX and AlexNet fail to generate representations that are robust to occlusion (Figure 5.1A). While model performance was significantly above chance, even the best performing model (fc7, red line) has significant gaps with human performance, particularly for more heavily occluded images. These results are similar to other experimental simulations with convolutional neural networks (Pepik et al., 2015; Wyatte et al., 2012).

To visualize the effect of occlusion on the model representations, we used multi-dimensional scaling (MDS) to project the features from the fc7 layer of AlexNet to two dimensions (Figure 5.1B). Multi-dimensional scaling is an algorithm that attempts to find a low-dimensional representation of

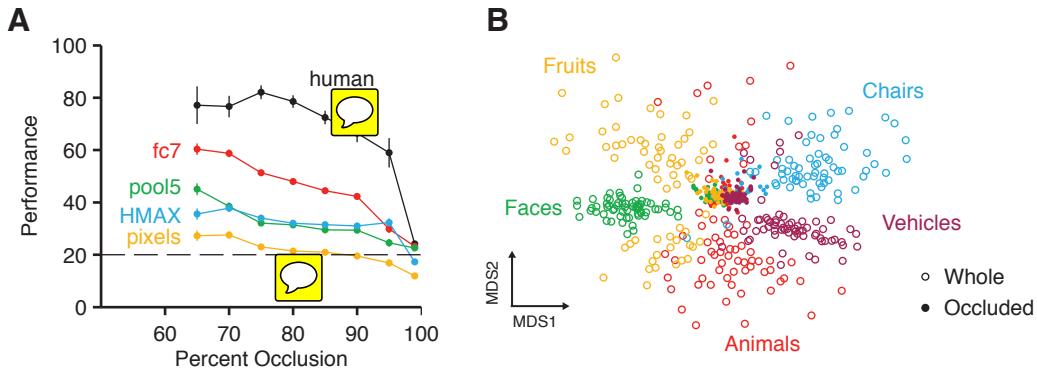




Figure 5.1: Feed-forward networks are not robust to occlusion.

- (A) Performance of various computational models (colors) compared to human performance (black). Performance was measured by training on whole objects and testing on the same partial objects shown to human psychophysics subjects (see main text for more detail and model descriptions). Human performance here is  the unmasked condition with 150ms exposure time.
- (B) Multi-dimensional scaling (MDS) was applied to feature vectors from the final layer of AlexNet (fc7 layer). Both whole objects (open circles) and partial objects (closed circles) from different categories are separable in this space, but the boundaries learned on whole objects do not generalize to the space of partial objects. 

the features which preserves the distance between points^{*}. In this space, partial objects from different categories were more similar to each other than to whole objects from their corresponding categories. These visualizations help explain why the decision boundaries trained on whole objects do not generalize to categorization of partial objects.

This is not to say that the fc7 responses to partial objects does not contain discriminable information. In fact, when we used the fc7 features, but trained the decision boundaries on partial ob-

^{*}In other words, points that are closer together in the high-dimensional space are still closer together in this low-dimensional embedding. Formally, given a list of I objects in \mathbb{R}^M , MDS finds an embedding $f : \mathbb{R}^M \rightarrow \mathbb{R}^N$ that seeks to minimize a given cost function. Here, we used Stress, which is defined as the difference between the pairwise distances in the original space, D and the pairwise distances in the reduced space, D' :

$$\text{Stress} \sim \|D - D'\|^2$$

Note that if we use a related cost function, called the strain, which instead measures the difference in the scalar products $\langle x, y \rangle$, then MDS is equivalent to Principal Component Analysis (PCA).

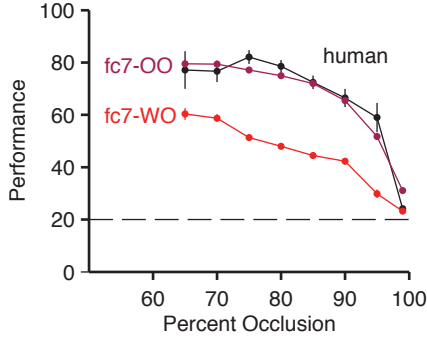


Figure 5.2: **Performance when trained on occluded objects**
 While the fc7 representation is not robust (red line, fc7-WO, trained on whole objects), the representation does contain discriminable information about occluded objects (purple line, fc7-OO, trained on occluded objects). However, this information is not formatted in a representation is robust to occlusion.

jects, performance on categorizing other partial objects reached human performance (Figure 5.2). However, this information was not formatted in a robust representation. This approach was also not successful because a classifier trained on partial objects did not perform well on recognition of whole objects, only reaching 91% performance. In comparison, training and testing on whole objects yielded 97.8% performance. For more complex datasets, the performance gap was even more pronounced (Pepik et al., 2015). Therefore, we concluded that new architectures are required to solve this problem.

5.2 Beyond feed-forward models

Several theories on the role of feedback connections emphasize inference, but these ideas have largely not been operationalized into object recognition models. Predictive coding models are generative models of object recognition (Rao and Ballard, 1999). In these models, higher visual areas send their predictions to lower levels, which then return only the mismatch between the predicted activity and the actual activity. This creates an efficient system where each layer only sends forwards signals that deviate from the receiving layer's predictions. The higher layers then attempt to generate the correct hypothesis of the image by reducing the incoming prediction errors. A related model proposes that visual cortex is essentially performing bayesian inference where feed-forward inputs combine with top-down priors for recognition (Lee and Mumford, 2003; Yuille and Kersten, 2006).

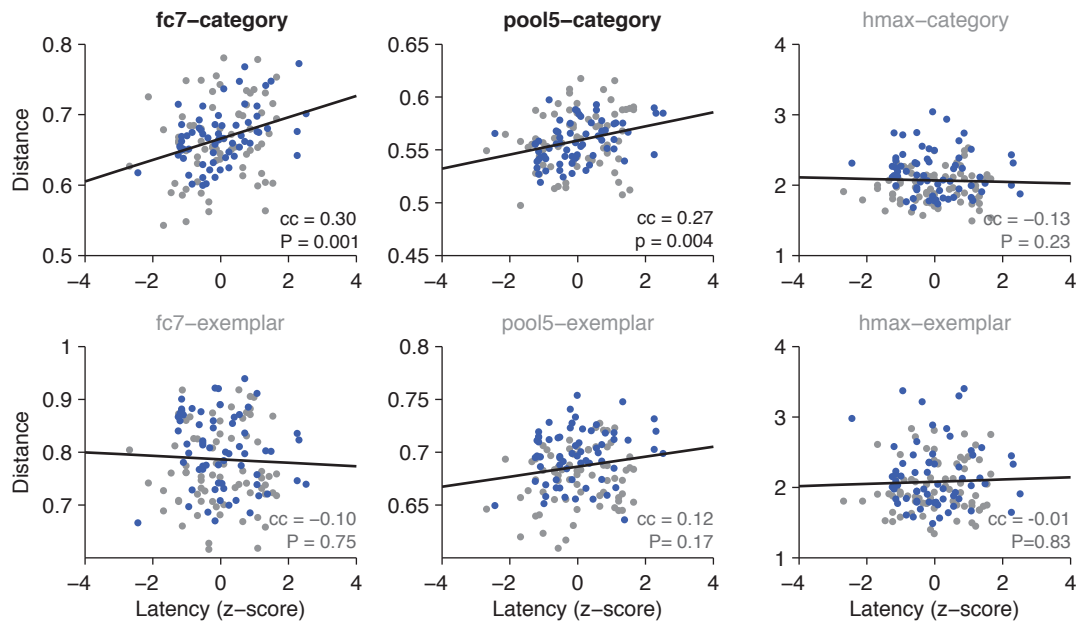


Figure 5.3: **Correlation between distance and latency.**

For partial images that were shown to subjects where we had intracranial recordings (*phy25* stimulus set, see Table 4.4), the correlation between the neural response latency and the distance from that partial image to its whole counterpart. Each dot is a partial image with responses recorded either from electrode #1 (blue dots) or electrode #2 (gray dots). Columns represent three different models (AlexNet-fc7, AlexNet-pool5, and hmax). The first row is distance in feature space from the partial image to its category mean, and the second row is distance from the partial image to its corresponding whole object. Bold title indicates a significant effect from a linear regression (see text).

Computer vision approaches have used recurrence to improve recognition on existing datasets, but with stimuli that are mostly unoccluded (Liang and Hu, 2015). Models have also been proposed that improve recognition of occluded faces or digits (Tang et al., 2012; Zhou et al., 2009; Jia and Martinez, 2008). These models take generative approaches, or use restricted boltzmann machines to perform image reconstruction and denoising. These approaches work by performing pixel space reconstruction, whereas our visual system is more concerned with extracting high-level labels from occluded objects (identity, emotion, gaze, etc.) than performing the expensive computations of filling in the occluded parts.

Alternatively, attractor networks such as Hopfield networks (Hopfield, 1982), when seeded with the complete pattern as attractors, can take a partial input pattern and dynamically evolve towards the correct attractor. Interestingly, this type of dynamical convergence towards the attractor state could account for the type of delays observed in the behavioral and physiological experiments.

To explore this possibility, we compared neural, behavioral, and computational measures on the same images. We used the same dataset from Chapter 4 with partial images where we have recorded both neurophysiological and behavioral responses (*phys25* dataset, see Table 4.4). We computed, for each partial image in the preferred category of the electrode, the euclidean distance to its corresponding whole object in a variety of model feature representations (fc7, pool5, and hmax). This distance represents how far the representation was 'pushed' because of occlusion. Suppose we have a particular whole object $o \in \text{faces}$ whose representation in Alexnet-fc7 is $\mathbf{h}_{\text{whole}}^o$. We then have a partial image of that object whose features we denote $\mathbf{h}_{\text{partial}}^o$. We also computed the average feature vector across all whole faces as $\bar{\mathbf{h}}_{\text{whole}}$. We measured two distances: the distance to the original exemplar $\|\mathbf{h}_{\text{partial}}^o - \mathbf{h}_{\text{whole}}^o\|^2$, and the distance to the category mean $\|\mathbf{h}_{\text{partial}}^o - \bar{\mathbf{h}}_{\text{whole}}\|^2$. We found a significant correlation between the distance to the category mean and the neurophysiological response latency for features in the fc7 and pool5 layers, but not the top layer of HMAX (Figure 5.3, top row). Images that were 'pushed' farther away from the category mean by occlusion also elicited slower neurophysiological responses.

To assess the significance of these effects, we first combined trials from both electrodes by z-scoring the response latency. For each distance metric, We used a linear regression on the latency with several predictors:

$$\text{latency} \sim 1 + \text{distance} + \%occlusion + \text{pixel_distance} + \text{electrode} + \text{masking_index}.$$



We included the factors `%occlusion` and `pixel_distance` to regress out any variation explained by

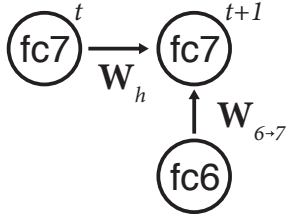


Figure 5.4: Schematic of recurrent network model. We added a recurrent loop within the top-level representation, fc7. At time $t + 1$, the fc7 layer takes two inputs: the constant drive from fc6, $\mathbf{W}_{6 \rightarrow 7} \mathbf{h}_t^{\text{fc6}}$, and the fc7 activity at the previous time step, $\mathbf{h}_t^{\text{fc7}}$, modulated by the weight matrix \mathbf{W}_h . These inputs are summed and passed through a linear rectification $\text{ReLu}(x) = \max(x, 0)$. The weight matrix governs the temporal evolution of the fc7 representation.

low-level effects of occlusion, and we included electrode to account for the inter-electrode variability in our dataset. We also used `masking_index`, which refers to the measure of image difficulty that was determined with psychophysical experiments (Chapter 4) to control for overall recognition difficulty. We found that distance in the fc7 and pool5 layers strongly correlated with the response latency beyond what can be explained by these low-level factors ($P = 0.001$ and $P = 0.004$, respectively). Note that here we have used partial images from the preferred category of the electrodes. As expected, images from the non-preferred categories were not significantly correlated (fc7, $P = 0.34$; pool5, $P = 0.78$). In addition, we did not observe significant correlations in any model when using distance to the original exemplar as a factor (Figure 5.3, bottom row). This link between distance in feature space and the resulting response latency is consistent with an attractor network approach. In this scheme, images that are 'pushed' farther would take longer to converge to the appropriate attractor.

As a proof-of-principle, we augmented the top layer of AlexNet with recurrence to generate a robust representation through an attractor-like mechanism (Figure 5.4). We denote the activity of the fc7 layer of AlexNet at time t as $\mathbf{h}_t^{\text{fc7}}$. Our recurrent loop was implemented by computing the activity in the next time step $t + 1$ as

$$\mathbf{h}_{t+1}^{\text{fc7}} = \text{ReLu} \left(\mathbf{W}_h \mathbf{h}_t^{\text{fc7}} + \mathbf{W}_{6 \rightarrow 7} \mathbf{h}_t^{\text{fc6}} \right), \quad (5.1)$$

where we used the rectified linear unit as our activation function, defined as $\text{ReLu}(x) = \max(x, 0)$. The first term inside the activation function consists of a weight matrix \mathbf{W}_h that describes the dy-

namics with which the fc7 layer activity evolves over time. The second term captures the constant input from the previous layer, fc6. The model was trained to adjust \mathbf{W}_h to minimize the euclidean distance between a set of partial images and its corresponding whole object.

Put more formally, our dataset consisted of a set of 325 whole objects, which we denote O . The representation in the fc7 layer for the o -th whole object is $\mathbf{h}_{\text{whole}}^o$. For each object $o \in O$, we also generated $N = 40$ partial images, indexed by i : $\{\mathbf{h}^{i=1,o}, \mathbf{h}^{i=2,o} \dots \mathbf{h}^{i=N,o}\}$, for a total of 13,000 partial images. These were the same partial images with which we had previously collected behavioral results (klab325 dataset, see Chapter 4). The fc7 representation at time t for the i -th partial image of the o -th object is $\mathbf{h}_t^{i,o}$. At each cross validation fold, we selected a subset of objects $M \subset O$ and minimized the cost function

$$\sum_{o \in M} \sum_{i=1}^{N=40} \left\| \mathbf{h}_{t=4}^{i,o} - \mathbf{h}_{\text{whole}}^o \right\|^2. \quad (5.2)$$

The model strives to evolve the fc7 representation of a partial image over $t = 4$ time steps to match the representation of its whole counterpart. Another way to understand this model is that occlusion transforms the representation of an object from $\mathbf{h}_{\text{whole}}^o \rightarrow \mathbf{h}_{\text{partial}}^{i,o}$, and this model dynamically reverses that transformation to restore the original representation. Importantly, for each cross-validation fold, we selected a subset of objects $M \subset O$ and their associated partial images to train the weight matrix \mathbf{W}_h , and then validated the model performance on partial images drawn from a separate set of objects $T \subset O \setminus M$.

To visualize the dynamics of this recurrent neural network, we used a variant of stochastic neighborhood embedding (t-SNE) to embed the high-dimensional fc7 features into a two-dimensional space[†] (Van der Maaten and Hinton, 2008). The trajectories of images are visualized in Figure 5.5.

[†]Here we use t-SNE instead of the multi-dimensional scaling (MDS) approach from earlier because t-SNE overcomes the crowding problem of MDS (see Figure 5.1 where the partial images are crowded near the center, making visualization difficult). t-SNE accomplishes this by favoring local structure accuracy as opposed to global distance accuracy.

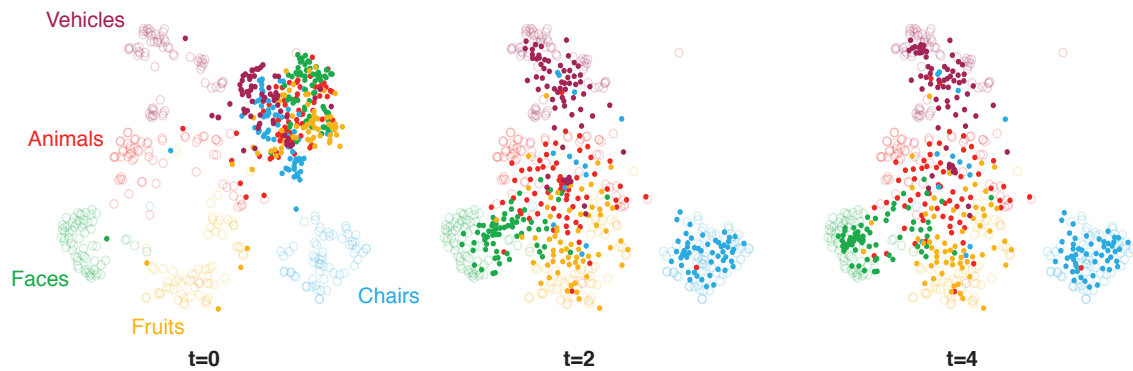


Figure 5.5: **Dynamics of recurrent neural network.**

We applied dimensionality reduction using t-SNE to visualize the time evolution of the fc7 representation in the recurrent neural network model (Van der Maaten and Hinton, 2008). Whole objects (open circles) and partial images (closed circles) are colored according to their category. For visualization purposes, only one partial image of each object is shown. Over time, the partial images approach the correct category in the clusters of whole images.

Before any recurrent computations have taken place, at $t = 0$, the partial images are clustered together (closed circles), separated from the clusters of whole objects from each category (open circles). As time progresses, the cluster of partial images are pulled apart, and dynamically attracted toward their respective categories. For example, at $t = 4$, the representation of partial chairs (closed blue circles) are now overlapping with the cluster of whole chairs (open blue circles). While it may appear that some categories such as faces (green) take longer to converge than others (i.e. partial images of chairs do not have much movement from $t = 2 \rightarrow 4$), this may be a consequence of the visualization method.

These dynamics created a representation at the final time step that is robust to occlusion. Performance when the classifier is trained on whole objects but tested on partial objects approached human performance (Figure 5.6A). When just measuring overall performance, the model saturated after the first time step (Figure 5.6B). However, we also measured the model by its similarity in the pattern of responses with humans tested on the same images. Even though overall performance was

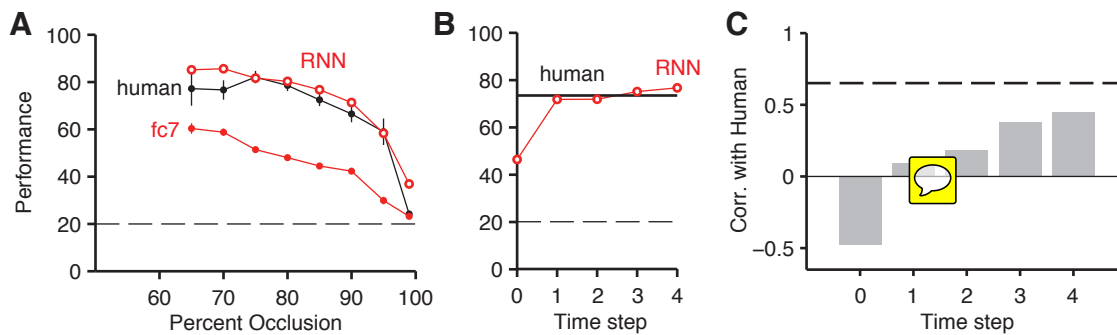


Figure 5.6: **Performance of recurrent neural network.**

- (A) Performance of the fc7 layer of the recurrent neural network (RNN, red open circles) approaches human performance (black lines), and is a significant improvement from the original fc7 layer (fc7, red closed circles). Dashed line indicates chance.
- (B) Overall performance of the RNN over the four time steps compared to human performance (black line). Performance saturates after the first time step.
- (C) Correlation in the pattern of responses between the model and humans. Dashed line indicates the upper bound of human-human similarity obtained by computing how well half of the subject pool correlates with the other half. Over time, the model becomes more human-like, even though overall performance is conserved.

saturated, over time the model became more human-like (Figure 5.6C). For each time step in the model, we computed the average correct rate on partial images from each of the 325 exemplars and correlated this vector with the human pattern of performance. The upper bound (dashed line) represents human-human similarity, defined as the correlation in the response patterns between half of the subject pool and the other half.

5.3 Discussion

In this chapter, we have demonstrated that existing convolutional neural networks do not have representations that are robust to occlusion (Figure 5.1). In fact, the stability of higher-level representations in AlexNet directly correlates with neural response latency; partial images that are moved a greater distance from the category mean in this space elicit slower neural responses, even after con-

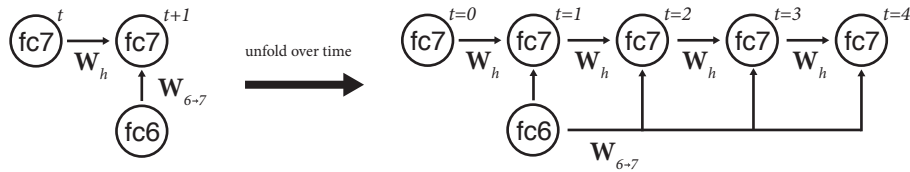


Figure 5.7: **Unfolding of recurrent neural network.**

Our recurrent neural network can be unfolded over time into a feed-forward network that is mathematically equivalent.

trolling for low-level differences (Figure 5.3). This link is consistent with an attractor-type model, where the farther the distance, the longer the system needs to converge. As a proof of principle, we augmented AlexNet with recurrent connections and trained the weights to carry out this attractor function. This recurrent neural network (RNN) dynamically restored the original feature representation (Figure 5.5), and reached human-like performance (Figure 5.6).

5.3.1 Feed-forward versus recurrent networks

We emphasize that our claim here is not that feed-forward networks can never perform recognition of occluded objects. Feed-forward and recurrent networks are topologically equivalent. In fact, to train our RNN, we unfolded the network in time to produce a feed-forward equivalent (Figure 5.7), then used back-propagation to train our weights. This approach, called back-propagation through time (BPTT), is the standard way of training recurrent networks (Werbos, 1988). However, while mathematically equivalent, there are several reasons why the brain may prefer to use recurrent networks instead of adding additional feed-forward layers to the visual system to solve a challenge like occlusion.

The energetic and material costs of creating and maintaining additional neurons are high, so RNNs are more materially efficient, at the cost of computational expressivity[‡]. Similarly, over the course of

[‡]While the RNN in Figure 5.7 is restricted to one weight matrix \mathbf{W}_h that governs the time evolution of the

learning, the visual system may need to gain invariance to a number of transformations, and adjusting synaptic weights in a RNN is easier than adding and reconfiguring new layers of additional processing in visual cortex. In addition, the number of neurons in a feed-forward network needed to mimic a RNN scales linearly with the number of time steps. For recurrence that may take place over many time steps, this scaling problem would prove impossible to overcome.


Given that behavioral and neural evidence suggests the involvement of recurrence, we demonstrate a role for recurrence through a simple RNN model. The fact that our RNN monotonically matches the pattern of human responses over time is further evidence that the dynamics of our network may be well aligned with brain function.

5.3.2 Training on occluded objects


In our data, existing feed-forward models fail to extrapolate from whole objects to occluded objects. However, the same models support a classifier trained on occluded objects that can match human performance when tested on occluded objects (Figure 5.2). The devil's advocate may argue that we've already solved the problem by augmenting the training set with occluded objects, so why at all develop recurrent models? In addition, training the RNN also requires occluded objects as training examples. Given the occluded objects and a feed-forward network, wouldn't the simplest solution be to just use occluded objects to train the decision boundaries instead of implementing a more complicated recurrent architecture?

We have ample evidence that this is not how the brain approaches the problem. If this were true, for example, we would not observe delayed response latencies, and backward masking would not degrade performance. We also have several theoretical arguments as to why this is the case. This system, the topologically equivalent feed-forward architecture on the right could in principle be trained to use different weights for different time steps, thus increasing expressivity. We note, however, that the brain can circumvent the restriction on RNNs we impose here by reconfiguring the weight matrix over short time scales with short term plasticity, biochemical modifications, and other synaptic mechanisms (Shepherd, 1998).

alternative approach, compared to the recurrent approach, would mean that the visual representation in areas like IT would not be invariant to occlusion, but rather this invariance would arise even later in the processing stream, such as prefrontal cortex. However, other regions and processes may need to access the visual representation in areas such as IT (e.g. the dorsal stream, motor planning, etc.), giving an advantage to systems that develop robust representations earlier in the ventral visual stream.

Additionally, when we trained the classifier on occluded objects, performance was significantly impacted when we tested on whole objects (performance was 91.5%, whereas training and testing on whole objects reached 97.8% performance). Previous research also demonstrates that augmenting the training data with occluded examples does not significantly improve performance on more complex datasets such as PASCAL3D (Pepik et al., 2015). Therefore, simply training with a feed-forward network on occluded objects is not a scalable approach --- a completely new architecture is required. 


5.3.3 Correlations between neural data and models

Previous studies of correspondence between neural data and computational models have shown that the features from deep convolutional networks can explain the ance of neural firing rates in macaque IT (Yamins et al., 2014; Cadieu et al., 2014). However, these comparisons have used the average firing rate of neurons over the initial feed-forward window, which is a static representation of IT responses. Our results demonstrate a very different property of these networks. When objects are transformed, the stability in the feature space correlates with the temporal evolution of the neural response. In addition, we examine responses well beyond the initial feed-forward sweep, and demonstrate that even here, distance in the feature space of the computational model has biological relevance.

Intriguingly, we noted in the results section that the response latency only correlated with the distance from the category mean, but not distance to its own exemplar. We speculate here on several possible explanations. First, in the neural recordings, the subjects were performing a categorization

task, which may modulate the response latencies. Based on this theory, one would predict that in a different task, the response latency would correlate with distance to the relevant task-defined categories. For example, in a young vs. old task for faces, the response latency would correlate with distance from the partial image to the average feature vector of all young whole faces. The second, and perhaps less exotic, theory is that the fc7 representation for individual exemplars may be a noisy estimate of the actual neural representations. Therefore, using distance to the exemplar injects unwanted variability into the distance measure, and is thus a less faithful representation of distance in the brain than distance to the category.

Response latency could correlate with feature distance due to low-level effects. We have controlled for these effects by including several terms in our regression: the percent of the object occluded, distance in pixel space, and a measure of behavioral difficulty obtained with extensive psychophysical experiments. Because of the need to control for behavioral difficulty, we have limited our correlation analysis to a subset of trials where we have psychophysical measurements with backward masking.

However, if one would tolerate a slightly less controlled analysis by not including this behavioral measure, an exciting future direction would be to compute correlations using all the trials from the many different electrodes along the ventral visual stream, as well as features from various processing stages in these convolutional neural networks. Examining when and where distance in feature space corresponds to the temporal dynamics can provide clues as to where these recurrent computations are being implemented. 

5.3.4 Attractor networks

Theoreticians have proposed a wide array of attractor network models (Hopfield, 1982; Seung, 1997), and their role in various neural functions (Ben-Yishai et al., 1995; Carandini and Ringach, 1997; Rolls, 2007). In this chapter we demonstrate that a simple implementation of these ideas can generate a stable representation at the top layer of the network. While this proof-of-principle matches human

performance, there are several next steps for improvement.

For example, here we have trained our network on a relatively small dataset because of the desire to measure correspondence with human performance. However, a much larger dataset is required to test the applicability of our proposed network to real-world computer vision challenges. In particular, we could generate occluded examples from ImageNet to train the recurrent weights of our RNN. These larger datasets could create even more generalizable dynamics.

While here we consider one particular occlusion transformation, the brain is unlikely to engage different recurrent weights when different transformations (occlusion, blur, pixelation, illumination changes, etc.) are detected. One would search for a general attractor model that would match human performance (and errors) across a wide range of transformations that are encountered in natural vision. Our RNN model works because same-category objects are transformed similarly by occlusion, so that the temporal dynamics learned on one set of the objects generalizes to other objects of the same class. While this may be true for occlusion, the capacity of the feature space to tolerate a multitude of transformations, often in combination, is unclear.

One weakness of our dataset is that we do not use occluding patterns, but rather substitute the missing object parts with the gray background. Our model would not be able to handle the complex occlusions we observe in nature where both the occluding element and the underlying image are objects (e.g. leaves covering a stop sign). In this scenario, depending on the amount of occlusion, the model may very well evolve towards the representation of the occluding pattern instead. Solving this problem may require implementing biologically-inspired mechanisms to generate a surface representation that separates the two elements for further processing (Nakayama et al., 1995).

While our behavioral and neural recordings provide evidence for recurrent computations, our techniques cannot differentiate between recurrent connectivity within layers and feedback across layers (e.g. from prefrontal cortex, hippocampus, or other structures). Top-down connections from higher cortex has been shown to bias activity in visual cortex (Miller, 2000; Buschman and Miller,

2007; Reynolds and Chelazzi, 2004). In the context of our model, top-down activity could be used to determine the proper task-dependent attractor locations (e.g. see discussion on correlations between latency and distance in section 5.3.3). Or, for ambiguous occluded stimuli that could give rise to multiple interpretations (such as the occluded stimuli discussed in the previous paragraph), feedback could use contextual cues to 'nudge' the representation down the correct pathway in an RNN (e.g. if one is driving, attend to the stop sign instead of the leaves).

6

Conclusion

Object completion is a difficult task for computational theories of vision, yet an ability we effortlessly deploy everyday in natural vision. In this thesis, I have presented neural, behavioral, and computational evidence for the role of recurrent computations in the occluded objects. Chapter 3 demonstrates that even under heavy occlusion, neural representations along the visual stream remain visually selective. However, responses to partial objects emerge with a 50-100 millisecond delay compared to that of whole objects. These delays are localized to higher visual cortex, and dependent on the visible features, suggesting the involvement of recurrent and/or feedback computations. Psychophysics experiments reveal that disrupting these computations with backward masking significantly degrades performance for occluded objects, but not whole objects (Chapter 4). Partial images that elicited a slower neurophysiological response are more susceptible to backward masking, and are

a farther distance from their category means in the feature space of computational models. Chapter 5 explains why existing feed-forward models of vision fail to generate representations that are robust to occlusion, and proposes a plausible recurrent neural network that reaches human-like performance on recognition of occluded objects.

This work makes several important conceptual contributions. From a computational perspective, feed-forward neural networks have been at the center of biological theories of vision over the past several decades. While we have always known that feedback and recurrent connections are prevalent throughout visual cortex, feed-forward networks and corresponding convolutional neural networks have had tremendous success in recognizing objects and explaining the variance of neural responses. Here we present a theoretical challenge to these theories by considering the problem of occlusion. Delays in the neural response latency with occlusion cannot be explained by feed-forward networks. In addition, computational simulations demonstrate that representations at the highest layers of these convolutional networks are not robust to occlusion. These results call for novel computational architectures beyond feed-forward vision.

From the perspective of a visual neuroscientist, our approach differs substantially from previous human experiments. These experiments typically examine object completion by contrasting responses between, for example, an occluded line drawing against an unrecognizable scrambled counterpart (Lerner et al., 2002; Doniger et al., 2000; Sehatpour et al., 2008; Chen et al., 2009). In these comparisons, it is difficult to tease apart the contribution of object completion mechanisms from the increased activity triggered by perceptual recognition. Instead, perhaps a different way to think about the problem is to measure how information content in the neural code is (or is not) is changed with occlusion (Kosai et al., 2014; Kovacs et al., 1995b). Critical to this approach is our use of a recording methodology with single-trial resolution. Previous studies with EEG, which report trial-averaged responses, indicate that occluded images elicit responses with smaller amplitudes compared to responses to whole images (Chen et al., 2009; Doniger et al., 2000). However, with our single-trial

resolution we show that the amplitudes are in fact, not suppressed, and that the lower observed amplitude in EEG studies stems from averaging over response with variable response latencies.

In these experiments, we measured both the latency of selectivity, which reflects the time at which category selective information emerged, and the response latency, which denotes the time at which the response amplitude differed significantly from the baseline activity. While the latency of selectivity was consistently delayed along the ventral visual stream, delays in the response latency were spatially localized. Responses in early visual cortex (occipital lobe) were not delayed, consistent with recordings in awake macaque V4 (Kosai et al., 2014). The response latencies in high visual areas (temporal lobe), however, were significantly delayed, matching findings from anesthetized macaque IT (Kovacs et al., 1995a). This provides a unified explanation linking these two macaque studies, which were previously not comparable because they used different shape stimuli, occluding patterns, and experimental paradigms. The consistency we observe also provides one example where findings with simple geometric shapes extrapolates to those with naturalistic objects.

We have demonstrated several important links between behavior, neurophysiology, and computation. In our neural recordings, we observed a wide distribution of response latencies to occluded stimuli. Our experiments show that this latency was dependent on the set of visible features (repeated presentations while keeping the visible features constant quenched the variability in latency), but we were unable to determine the nature of that dependence based on the image alone. One might hypothesize that more informative features, when visible, would elicit a faster neural response. However, our analyses did not find any significant correlations. Instead, the response latencies were correlated with two measures derived from other approaches, one psychophysical (susceptibility to backward masking), and one computational (distance in feature space of a convolutional neural network). These intriguing correlations are significant even when controlling for low-level effects such as the amount of occlusion, and are consistent with an attractor mechanism for object completion. I believe the combination of neural recordings with behavioral psychophysics and computational

modeling (i.e. brains, minds, and machines) presented in this dissertation is useful and rarely found in studies in human neuroscience.

Given that strong behavioral and neurophysiological evidence exists for object completion in human brain, and that surface representations are important for organizing the visual scene, models of object recognition would be remiss to exclude these features in favor of purely feature-based recognition. An important step towards theories that fully capture natural biological vision would be to integrate traditional feed-forward models with recurrent and feedback mechanisms, including amodal completion, attractor networks, surface generation, and top-down modulation based on priors and context. The challenge of recognizing occluded objects stands as the first test of these future integrative theories.



Dynamics of cognitive control

Note: Since electrode locations are determined by clinical needs, patient coverage can vary widely. This appendix details a separate study I undertook during my Ph.D. for patients with frontal cortex coverage.

Rapid and flexible interpretation of conflicting sensory inputs in the context of current goals is a critical component of cognitive control that is orchestrated by frontal cortex. The relative roles of distinct subregions within frontal cortex are poorly understood. To examine the dynamics underlying cognitive control across frontal regions, we took advantage of the spatiotemporal resolution of intracranial recordings in epilepsy patients while subjects resolved color-word conflict. We observed differential activity preceding the behavioral responses to conflict trials throughout frontal

cortex; this activity was correlated with behavioral reaction times. These signals emerged first in anterior cingulate cortex (ACC) before dorsolateral prefrontal cortex (dlPFC), followed by medial frontal cortex (mFC) and then by orbitofrontal cortex (OFC). These results disassociate the frontal subregions based on their dynamics, and suggest a temporal hierarchy for cognitive control in human cortex.

A.1 Introduction

Flexible control of cognitive processes is fundamental to daily activities, including the execution of goal-directed tasks according to stimulus inputs and context dependencies. An important case of cognitive control arises when input stimuli elicit conflicting responses and subjects must select the task-relevant response despite competition from an often stronger but task-irrelevant response (Miller, 2000; Miller and Cohen, 2001). A canonical example of this type of conflict is the Stroop task: subjects are asked to name the font color of a word where the semantic meaning conflicts with the color signal (e.g. the word “red” shown in green versus red). Such incongruent inputs lead to longer reaction times, attributed to weaker signals (color processing) that must be emphasized over the automatic processing of word information (Stroop, 1935). The Stroop task is frequently used in cognitive neuroscience and clinical psychology and forms the foundation for theories of cognitive control.

Neurophysiological, neuroimaging, and lesion studies have ascribed a critical role in cognitive control to networks within frontal cortex (Miller, 2000), yet the neural circuit dynamics and mechanisms responsible for orchestrating control processes remain poorly understood. Lesion studies (Cohen and Servan-Schreiber, 1992; Perrett, 1974), human neuroimaging measurements (Egner and Hirsch, 2005; MacDonald, 2000), and macaque single unit recordings (Johnston et al., 2007) implicate the dorsolateral prefrontal cortex (dlPFC) in providing top-down signals to bias processing

in favor of the task-relevant stimuli (Botvinick et al., 2001; Miller and Cohen, 2001). The medial frontal cortex (mFC) also participates in cognitive control, possibly in a conflict monitoring capacity (Botvinick et al., 2001; Ridderinkhof et al., 2004; Rushworth et al., 2004). Recordings and lesions studies in the macaque anterior cingulate cortex (ACC) (Ito et al., 2003; Nakamura et al., 2005) suggest that ACC neurons are principally involved in monitoring for errors and making between-trial adjustments (Brown and Braver, 2005; Ito et al., 2003; Johnston et al., 2007)—an idea that has received support by a recent study in the human ACC (Sheth et al., 2012). Recent work has also demonstrated that the supplementary motor area and the medial frontal cortex play an important role in monitoring for errors (Bonini et al., 2014). An alternative and influential theoretical framework posits that the ACC monitors for potential conflicts and subsequently directs the dlPFC to engage control processes (Botvinick et al., 2001; Shenhav et al., 2013). Several human neuroimaging studies are consistent with this notion (Botvinick et al., 1999; Kerns, 2006; Kerns et al., 2004; MacDonald, 2000) but the relative contributions of dlPFC, mFC, and ACC to cognitive control remain a matter of debate (Aarts et al., 2008; Cole et al., 2009; Fellows and Farah, 2005; Mansouri et al., 2007; Milham et al., 2003; Rushworth et al., 2004).

Previously, some neuroimaging studies have suggested that these frontal cortex regions can be differentiated based on the presence or absence of conflict signals (MacDonald, 2000). The challenge in dissociating the relative roles of these regions during Stroop-like tasks is that increased task difficulty recruits a host of executive functions (attention, decision-making, uncertainty, cognitive control). These functions are associated with neural activity spanning tens to hundreds of milliseconds that are hard to untangle with the low temporal resolution of existing neuroimaging techniques (Shenhav et al., 2013). Human single neuron studies provide millisecond resolution but have focused on individual regions (Sheth et al., 2012). We took advantage of the high spatiotemporal resolution of intracranial recordings in human epilepsy patients and the ability to record simultaneously from multiple regions to directly investigate the dynamics of conflict responses during cognitive control.

We hypothesized that subregions of frontal cortex could be differentiated based on the temporal profile of their conflict responses. We recorded intracranial field potentials from 1,397 electrodes in 15 subjects while they performed the Stroop task or a variation in which they were asked to read the word instead of focusing on its color.

We observed conflict-selective activity throughout several regions in frontal cortex: ACC, mFC, dlPFC, and also orbitofrontal cortex. Several analyses link these signals to cognitive control. Neural responses were increased for incongruent compared to congruent trials, and these signals correlated with behavioral reaction time, depended on the task, and exhibited adaptation over trials. We compared pairs of simultaneously recorded electrodes to disassociate these different regions based on the timing of these conflict responses rather than their presence or absence. Conflict responses emerged first in the ACC and subsequently emerged in dlPFC and mFC and finally in OFC. These observations propose a plausible flow of signals underlying cognitive control.

A.2 Results

We recorded field potentials from 15 epilepsy patients implanted with intracranial electrodes in frontal cortex as they performed the Stroop task (Fig. A.1A, Table A.1). After 500 ms of a fixation cross, subjects were presented with one of three words (Red, Blue, Green), which were colored either red, blue, or green. We refer to congruent trials (C) where the font color matched the word (60% of the trials) compared to incongruent trials (I) where the font color conflicted with the word (40% of the trials). Within each trial type, the word-color combinations were counter-balanced and randomly interleaved. The stimuli were presented for 2 seconds (in two subjects, for 3 seconds). Subjects were asked to respond verbally and either name the color (Stroop task), or read the word (Reading task) in separate blocks. Performance during congruent trials was essentially at ceiling (Fig. A.2).

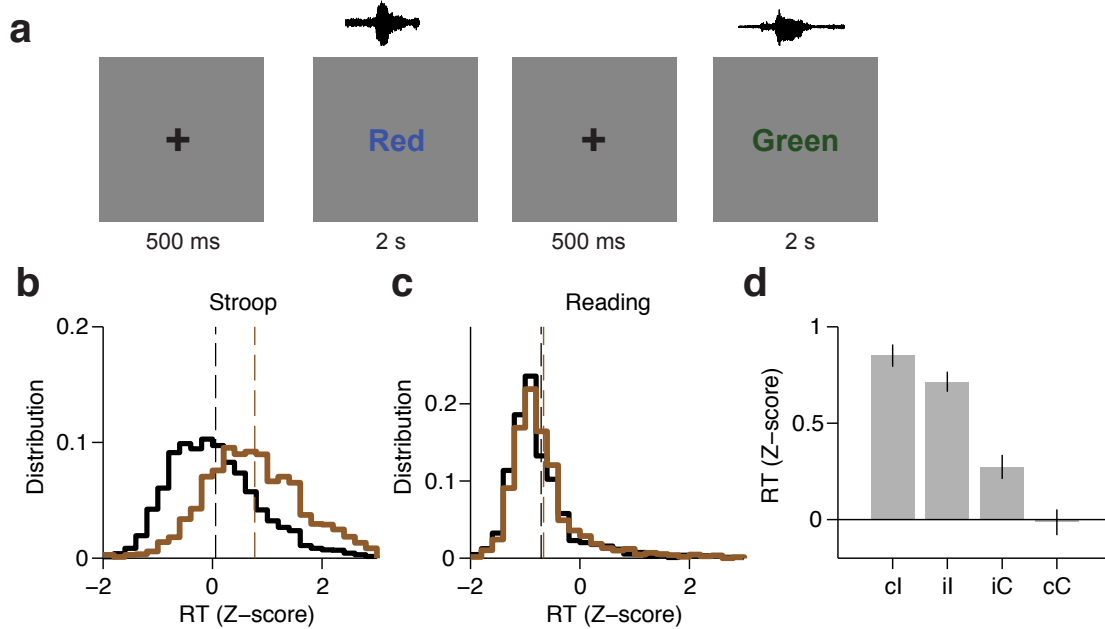


Figure A.1: Experimental task and behavioral performance

(A) Subjects were presented with one of three words (Red, Blue or Green); each word was randomly colored red, blue, or green. Trials were incongruent (I) when the word and color did not match, and were congruent (C) otherwise. The word-color combinations were counter-balanced and randomly interleaved. Subjects performed the Stroop task (name the color), and the Reading task (read the word) in separate blocks.

(B) Distribution of z-scored behavioral reaction times (speech onset) across all subjects ($n=15$) for congruent (black) or incongruent (brown) trials during the Stroop task. Bin size = 0.2. Dashed lines indicate average reaction times. Fig. A.2 shows the percentage correct and reaction times for each individual subject.

(C) Distribution of z-scored reaction times during the Reading task.

(D) Z-scored reaction time across subjects for different trial histories during the Stroop Task (ci: incongruent trial preceded by congruent trial; ii: incongruent trial preceded by incongruent trial; iC: congruent trial preceded by incongruent trial; cC: congruent trial preceded by congruent trial). Error bars indicate s.e.m.

An ANOVA conducted on subjects' performance with stimulus type (congruent or incongruent) and task (Stroop or Reading) as repeated measures revealed a significant interaction between stimulus type and task ($F = 22.9$, $P < 0.001$). For the Stroop task, subjects made more errors during incongruent trials (average error rate: $5 \pm 3\%$, $P < 0.001$ paired t-test), as demonstrated in previous studies (Bugg et al., 2008; Egner and Hirsch, 2005; Kerns et al., 2004). There was no difference in the number of error trials during the Reading task ($P = 0.76$, paired t-test). Subsequent analyses

focused on correct trials only unless otherwise stated. Subjects' reaction times also had a significant interaction between stimulus type and task ($F = 65.2, P < 10^{-5}$, ANOVA). Consistent with previous observations (Stroop, 1935), subjects' response times during the Stroop task were delayed for incongruent trials compared to congruent trials (Fig. A.1B, average delay: 215 ± 93 ms, $P < 0.001$, paired t-test). The reaction time delays were shorter in the Reading task (Fig. A.1C, average delay: 22 ± 31 ms, $P = 0.02$, paired t-test). Trial history also has a strong effect on reaction time (known as Gratton effect in the literature (Gratton et al., 1992)). A repeated measures ANOVA revealed an interaction between previous and current trial type ($F = 19.5, P < 0.001$). Incongruent trials that were preceded by a congruent trial (cI trials) elicited slower reaction times compared to incongruent trials that were preceded by an incongruent trial (iI trials) (Fig. A.1D, average reaction time difference: 34 ± 14 ms, $P = 0.03$, paired t-test). A similar Gratton effect was found for iC versus cC trials (Fig. A.1D, average reaction time difference: 72 ± 136 ms, $P < 0.001$, paired t-test).

We recorded intracranial field potentials from 1,397 electrodes (average 93 ± 31 electrodes per subject) while subjects performed the Stroop and Reading tasks. The number of electrodes per subject and the location of these electrodes was strictly dictated by clinical needs. Therefore, there was a wide distribution of electrode locations, as is typical in this type of recordings (Liu et al., 2009). We excluded electrodes in epileptogenic regions. We focused on the neural signals in the gamma band (70-120 Hz) given their prominence in sensory, motor and cognitive phenomena (Crone et al., 1998b; Liu et al., 2009; Oehrns et al., 2014); results for other frequency bands are shown in Fig. A.9, A.10, and A.11. Presentation of the visual stimuli evoked rapid neural responses in visual cortical areas, as expected from previous studies (e.g. (Tang et al., 2014)). Other electrodes were activated for different motor (verbal) outputs (e.g. (Bouchard et al., 2013; Crone et al., 1998b)).

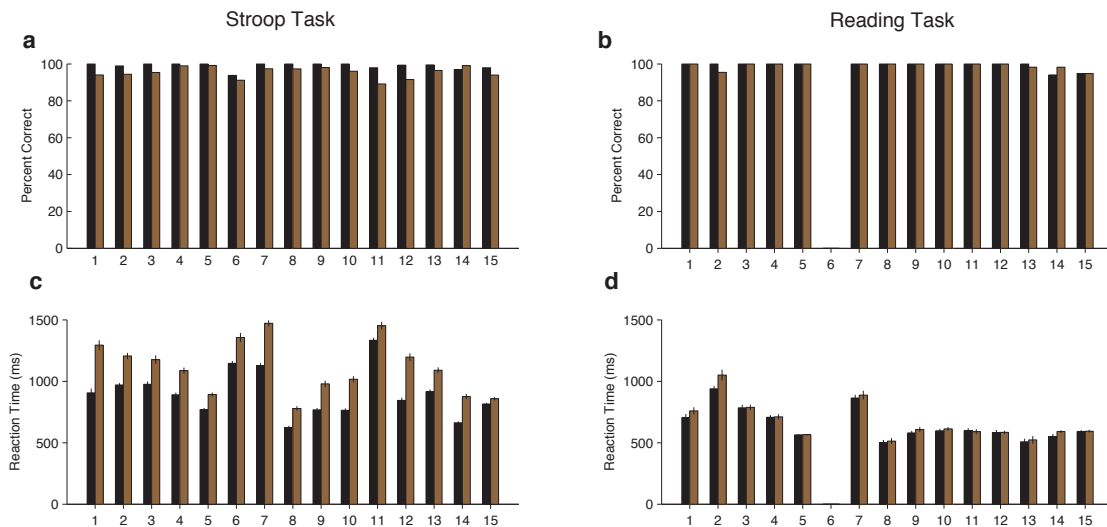


Figure A.2: Behavioral data for each subject

(A-B) Percent correct for each subject for the Stroop task (A) or Reading task (B) during congruent (black) or incongruent (brown) trials. Subjects made more errors for incongruent trials compared to congruent trials during the Stroop task ($P < 0.001$, signed-rank test). One subject (Subject 6) did not participate in the Reading task.

(C-D) Average behavioral reaction time (speech onset) for each subject for the Stroop task (C) or Reading task (D). Error bars indicate s.e.m. Subjects had delayed responses for incongruent trials compared to congruent trials during the Stroop task ($P < 0.001$, signed-rank test).

A.2.1 Conflict responses in frontal cortex

We focused on 469 electrodes located in areas within frontal lobe which have been previously implicated in executive function: medial frontal cortex (mFC, $n = 111$), orbitofrontal cortex (OFC, $n = 156$), dorsolateral prefrontal cortex (dlPFC, $n = 168$) and the anterior cingulate cortex (ACC, $n = 34$). Table A.2 provides a list of all electrode locations and their distribution across subjects. We applied a non-parametric analysis of variance (ANOVA) to measure whether and when the physiological responses differed between congruent and incongruent trials. An electrode was considered conflict-selective if the F-statistic was greater than a significance threshold computed by a permutation test with $P = 0.001$ for 50 consecutive milliseconds (Methods). The latency was defined as the

first time of this threshold-crossing.

Figure A.3 shows an example electrode from the left Anterior Cingulate Cortex that responded differentially between congruent and incongruent trials during the Stroop task. These signals were better aligned to the speech onset than to the stimulus onset, as shown in the response-aligned view (compare Fig. A.3A-C with Fig. A.3D-F). During the Stroop task, the response-aligned signals were significantly stronger for the incongruent (brown) trials compared to the congruent (black) trials (Fig. A.3D, $P < 10^{-5}$, ANOVA), and were invariant to the particular word/color combinations (Fig. A.3G). Incongruent trials could be discriminated from congruent trials at a latency of 669 ± 31 ms (mean \pm s.e.m.) before the onset of the response (Fig. A.3D). This conflict response was also specific to the Stroop task; there was a significant interaction between congruency and task ($F = 13.5$, $P = 0.007$, ANOVA). The same stimuli did not elicit differential activity during the Reading task (Fig. A.3F). We assessed the correlation between the neural signal strength and behavioral reaction times in single trials. We computed the maximal gamma power during each incongruent trial (using the average gamma power yielded similar results). The gamma power was positively correlated with the behavioral reaction times (Fig. A.3H, $\rho = 0.25$, $P = 0.02$).

Any differences between congruent and incongruent trials in the stimulus-aligned analyses can be confounded by the reaction time differences; therefore, we focus subsequent analyses on the response-aligned signals. More example electrodes are shown in Fig. A.8 (dlPFC) and Fig. A.9 (OFC).

Using the aforementioned criteria, we identified $n = 51$ conflict selective frontal cortex electrodes during the Stroop task, with contributions from 13 subjects (Table S2). These electrodes were distributed throughout different subregions within frontal cortex (Fig. A.4A). To evaluate whether random variation in the signals could give rise to apparent conflict-selective electrodes, we randomly shuffled the congruent/incongruent trial labels 10,000 times and applied the same statistical criteria (Methods). Across our population, we found $n = 4.4 \pm 0.03$ false positive electrodes (mean \pm s.e.m.,

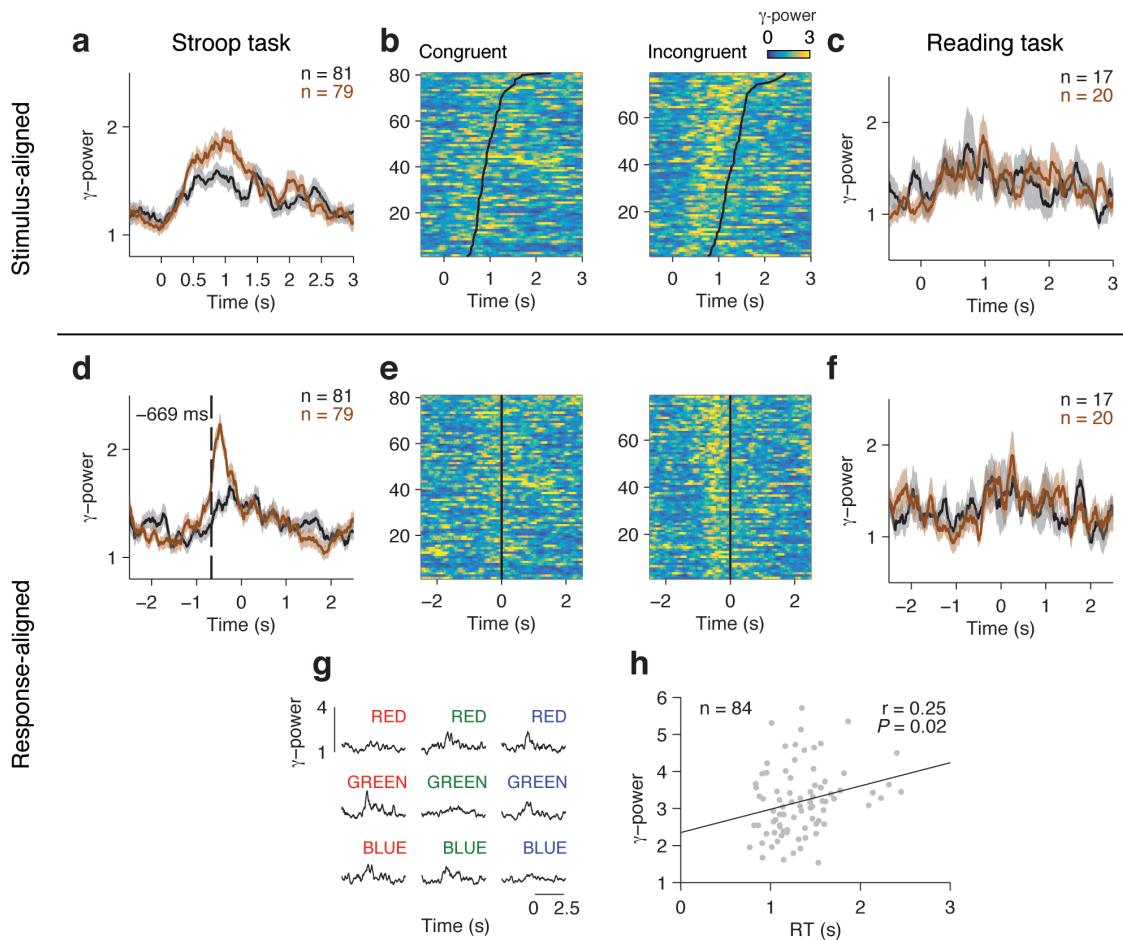


Figure A.3: Example responses from the anterior cingulate cortex

(A) Average gamma power signals aligned to the stimulus onset from an electrode during the Stroop task, for congruent (black) or incongruent (brown) stimuli. Gamma power was normalized to the baseline period (500 ms prior to stimulus onset). Shaded areas indicate s.e.m. The total number of trials for each condition is indicated in the upper right.

(B) Single-trial data for congruent (left) and incongruent (right) trials. Each row is a trial, and the color indicates the normalized gamma power (color scale on upper right). Trials are sorted by behavioral response time (black line).

(C) Same as (A), but showing data from the Reading task.

(D-F) Same as in (A-C), but aligning the data to behavioral response time. Gamma power was better aligned to the behavioral response, and was stronger for incongruent compared to congruent trials. The dashed line indicates the response-aligned latency, defined as the first time point at which incongruent and congruent trials can be discriminated.

(G) Signals elicited by each of the 9 possible stimulus combinations.

(H) There was a correlation between gamma power and behavioral reaction times during incongruent trials ($\rho = 0.25$, $P = 0.02$, permutation test).

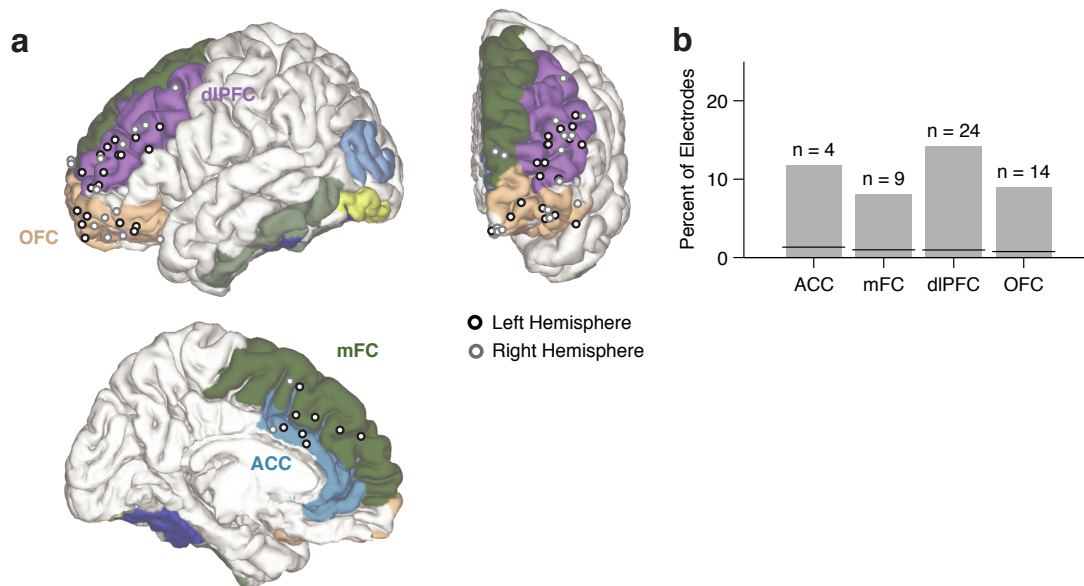


Figure A.4: **Electrode locations**

(A) Location of conflict-selective electrodes (black/gray) shown on a reference brain, with each region colored (Methods). Electrodes from the right hemisphere were mapped to the left hemisphere for display purposes. (B) Percent of total electrodes in each region that were selective for conflict. Chance levels were computed using a permutation test (black line). The number of observed electrodes was significantly above chance for all regions ($P < 0.01$, permutation test, Methods).

out of 469 electrodes), which corresponds to a false discovery rate (FDR) of $q = 0.01$, compared against our observation of $n = 51$ electrodes. The number of conflict-selective electrodes within each subregion was significantly greater than expected by chance (Fig. A.4B, $P < 0.01$, all regions). We repeated the analyses during the Reading task. In contrast with the Stroop task, we only observed $n = 3$ conflict-selective frontal cortex electrodes during the Reading task (out of 469 electrodes), a number that is within the false positive rate.

To account for within-subject and across-subject variation, we used a multilevel model (Aarts et al., 2014) to conduct a group analysis of the physiological responses, with electrodes nested within subjects (Methods). Across the population, we observed a significant interaction between the factors congruency and task on the gamma power ($\chi^2 = 9.2$, $P = 0.002$). Consistent with the single

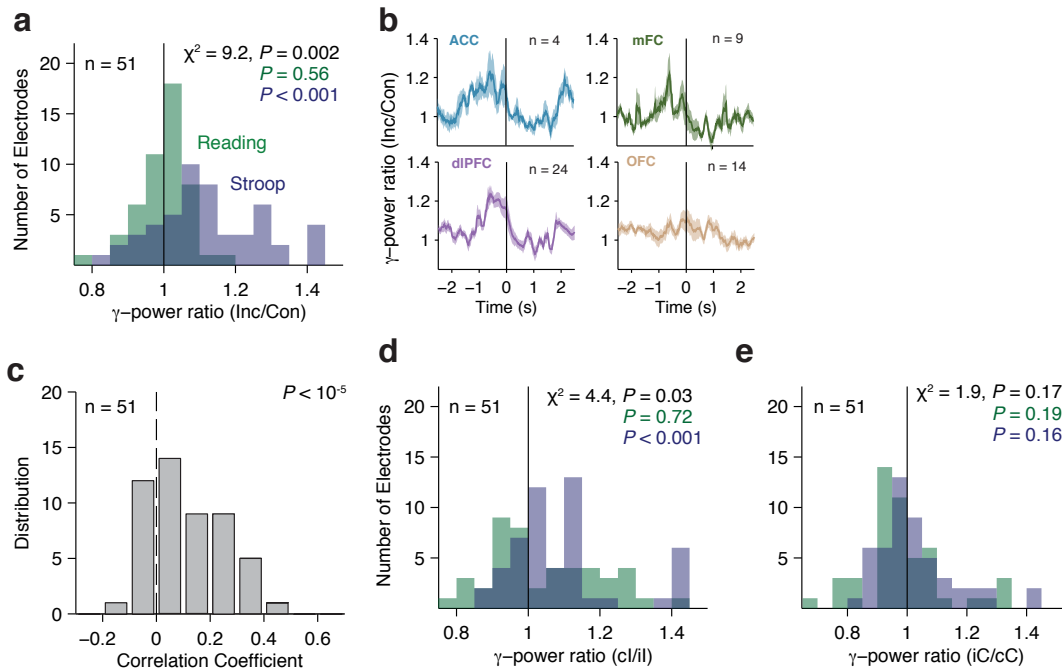


Figure A.5: Gamma power in frontal cortex correlates with behavior

(A) Distribution of gamma power ratio (Incongruent/Congruent) for the Stroop task (blue) and Reading task (green). Bin size = 0.05. Gamma power showed a significant interaction between Congruency and Task ($P = 0.002$, multilevel model). Power was larger for incongruent versus congruent trials during the Stroop task ($P < 0.001$, $n = 51$ frontal cortex electrodes) but not during the Reading task (green, $P = 0.56$). Gamma power ratios > 1.4 were included in the last bar for graphical purposes only.

(B) Normalized gamma power averaged across electrodes from each of the four different frontal cortex regions during the Stroop task. Responses were normalized by dividing the power during incongruent trials by the power during congruent trials. Data are aligned to the behavioral response onset ($t=0$).

(C) Distribution of Pearson correlation coefficients between the gamma power and behavioral reaction time during incongruent trials for $n = 51$ frontal cortex electrodes. These correlations were significantly positive ($P < 10^{-5}$, sign-rank test). Bin size = 0.1.

(D) For incongruent trials, there was a significant interaction between trial history and task ($P = 0.03$, multilevel model). Gamma power was larger for incongruent trials preceded by congruent trials (cI) compared to incongruent trials preceded by incongruent trials (iI), particularly during the Stroop task (blue, $P = 0.001$), compared to the Reading task (green, $P = 0.72$).

(E) For congruent trials, there was no interaction between trial history and task ($P = 0.17$, multilevel model).

electrode example, gamma power was greater for incongruent compared to congruent trials, but only during the Stroop task (Fig. A.5A, Stroop: $P < 10^{-3}$, Reading: $P = 0.56$). We computed the

average response in each region (Fig. A.5B). Each electrode's response was normalized by dividing the power during incongruent trials by the power in congruent trials (dividing the brown curve by the black curve in Figure A.3), then pooled within each region). The pooled responses in the OFC are visually less compelling (Fig. A.5B, bottom right subplot) due to the heterogeneity in the latency of the individual electrodes but the responses in the OFC were as vigorous as the ones in other areas (e.g. Fig. A.9).

A.2.2 Behavioral relevance of physiological responses

Several lines of evidence demonstrate a link between these neural signals and cognitive control: the neural signals correlated with reaction times, showed behavioral adaptation, and demonstrated error monitoring. As shown in previous studies, there was a wide distribution of behavioral reaction times (Fig. A.1B). Consistent with the example electrode in Fig. A.3, behavioral reaction times across the population correlated with the strength of the physiological signals, even after controlling for trial history (Fig. A.5C, $P < 10^{-5}$, sign-rank test).

The strength of these neural signals also revealed a neural correlate of the behavioral Gratton effect documented in Figure A.1D: gamma power was greater in cI compared to iI trials (Fig A.5D). Using the aforementioned multilevel model, we found a significant interaction between trial history (cI or iI) and task ($\chi^2 = 4.4, P = 0.03$). This Gratton effect was stronger in the Stroop task ($P < 0.001$) than in the Reading task ($P = 0.72$). These differences were not observed for cC versus iC trials, where the interaction was not significant ($\chi^2 = 1.9, P = 0.17$) (Fig A.5E). This analysis was performed after removing stimulus repetition trials. To control for reaction time effects on these comparisons, we ran an analysis of covariance (ANCOVA) to test for a main effect of trial history on the gamma power with the behavioral reaction time as a covariate (Methods). The neural Gratton effect during the Stroop task persisted under these controlled conditions ($P = 0.0002$, multilevel model). We also explicitly ruled out reaction time differences by subsampling to match the reaction

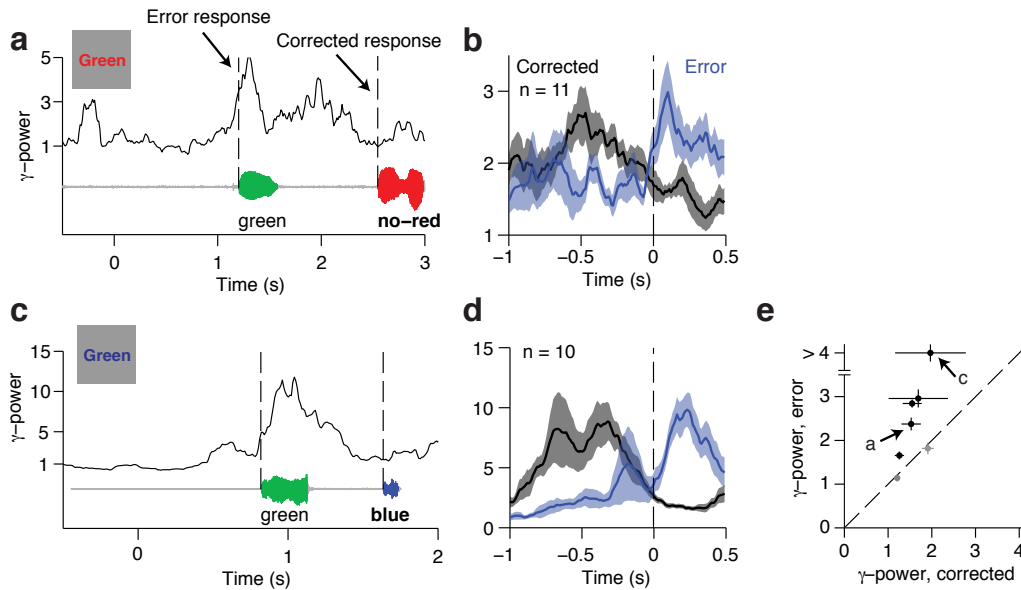


Figure A.6: Responses during self-corrected error trials

(A) An example self-correction trial from the ACC electrode in Fig. A.3 when the word Green colored in red was presented. The single trial gamma power is shown on top, with the speech waveform below. The dashed lines indicate the onset of the initially incorrect response ("green") and the following corrected response in bold ("no - red"). Note the increased gamma power after an error response. (B) Average gamma power aligned to the onset of the initial error response (blue) and the onset of the corrected response (black) for $n=11$ self-correction trials. Shaded areas indicate s.e.m. The post-response power was significantly greater after the error ($P = 0.001$, signed-rank test). (C-D) Same as (A-B) for another example electrode in the dorsolateral prefrontal cortex. The post-response power was significantly greater after the error response ($P = 0.002$, signed-rank test). (E) Across the $n = 7$ electrodes with $n = 10$ or greater self-correction trials, the gamma power during the initial error response was larger than during the corrected response. Electrodes with significant differences ($P < 0.05$, signed-rank test) are colored black. Letters mark the examples in (A) and (C).

time distribution between conditions, with similar results ($P = 0.01$, multilevel model). Together, these results suggest that the neural signals described here code for an internally perceived level of conflict that exhibits conflict adaptation and correlates with the across-trial variability in reaction times.

The conflict responses reported above are based on correct trials only. Yet, error monitoring has also been ascribed to frontal cortical circuits (Bonini et al., 2014; Shenhav et al., 2013; Yeung et al.,

2004). To investigate whether the same electrodes responding to conflict are also involved in successful error monitoring, we analyzed the neural signals during self-corrected trials. In these trials, subjects initially made an erroneous response and rapidly corrected themselves with the right answer. Given the high performance level of all subjects (Fig. A.2), the number of such trials is low. However, these trials are particularly interesting because we can be certain that there was successful error detection (as opposed to error trials without any self-correction). An example self-corrected trial from the ACC electrode shown previously is illustrated in Fig. A.6A. The subject initially made an incorrect response (green), which was rapidly followed with the correct response (red). Increased gamma power was observed after onset of the erroneous response. In contrast, the following corrected behavioral response exhibited no such post-response signal. These error-monitoring signals were also not observed in correct incongruent trials (Fig. A.3D), and were consistent across the $n = 11$ self-corrected trials for this subject (Fig. A.6B, $P = 0.001$, signed rank test). Another example electrode is shown in Fig. A.6C-D. There were only two subjects contributing $n = 7$ conflict-signaling electrodes that had a sufficient number of self-correction trials (greater than five trials) for this analysis. For each electrode, we compared the difference in neural signals during the one-second post-response window between the initial error and the following self-correction. Of those $n=7$ electrodes, $n = 5$ electrodes showed evidence of error monitoring (Figure A.6E, $P < 0.05$, sign-rank test). Although the number of electrodes and trials in this analysis is small, these results provide a direct correlate of error monitoring signals. Furthermore, these results highlight that the same electrodes that respond to conflict leading up to the behavioral response can also show post-response error monitoring.

A.2.3 Regional differences in conflict response latencies

We observed conflict-selective responses in the anterior cingulate cortex, medial frontal cortex, dorsolateral prefrontal cortex and orbitofrontal cortex. To examine the dynamics of cognitive control orchestrating the transformation of conflicting visual signals to motor outputs, we compared, across

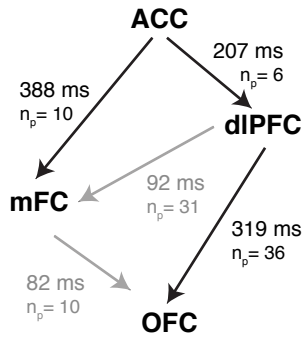


Figure A.7: **Latency comparison across regions**

Latency differences between different regions computed from all pairs of simultaneously recorded electrodes. n_p denotes the number of electrode pairs. Significant latency differences ($P < 0.05$, permutation test, Methods) are shown in black, and non-significant differences in gray. ACC leads both mFC ($P = 0.001$) and dlPFC ($P = 0.02$), with OFC following dlPFC ($P = 0.009$).

those four regional groups, the latencies relative to behavioral response onset at which the congruent and incongruent trials could be discriminated. Comparing latencies across regions is difficult especially across subjects with varying reaction times. For a controlled and direct comparison, we restricted the analysis to compute the latency differences between pairs of simultaneously recorded electrodes. This within-subject pairwise analysis had increased power to examine the relative dynamics between frontal lobe areas (Fig. A.7). The relative latencies were significantly different across the regions ($P = 0.01$, permutation test, post-hoc testing was controlled for multiple comparisons using the Benjamin-Hochberg procedure, Methods). Conflict responses in the ACC preceded those in all the other frontal lobe regions, followed 207 ± 40 ms later by dorsolateral prefrontal cortex and 388 ± 83 ms later by medial frontal cortex. Signals in orbitofrontal cortex emerged 319 ± 78 ms after dlPFC. This entire processing cascade took approximately 500 ms. For comparison, subjects' behavioral reaction times to incongruent trials were $1,105 \pm 49$ ms. These results suggest a temporal hierarchy of cognitive control mechanisms culminating in speech onset.

A.2.4 Conflict responses in other frequency bands

The results presented above focus on the neural signals filtered within the gamma frequency band (70-120 Hz). We also examined the responses elicited in the broadband signals (1 to 100 Hz) as well

as in the theta, (4 to 8 Hz), beta, (9 to 30 Hz), and low gamma (30-70 Hz) bands. No conflict selective responses were observed in the broadband signals or low gamma band. We found conflict-selective responses both in the theta and beta bands (see example in Fig. A.9). Across theta and beta frequency bands, we also observed a significant interaction between Congruency and Task (theta: $P < 10^{-5}$, beta: $P < 10^{-4}$, multilevel model). Consistent with the results reported in the gamma frequency band, conflict responses in the theta and beta bands were more prominent during the Stroop task compared to the Reading task (Fig. A.10). In contrast to the results in the gamma band, power in the theta and beta bands decreased during incongruent trials. Furthermore, power in the theta and beta frequency bands was not correlated with reaction times (theta: $P = 0.43$, beta: $P = 0.09$, sign-rank test).

In addition to separately examining the responses in different frequency bands, an important aspect of encoding of cognitive information is the relationship between signals across frequencies. In particular, several studies have demonstrated that the amplitude of the gamma band is coupled to the phase of slower oscillations in the theta band (Canolty et al., 2006; Oehrn et al., 2014; Tort et al., 2008). We therefore examined the cross-frequency coupling between the signals in the gamma and theta bands (Fig. A.11). Consistent with previous studies, we found that 50% of the electrodes demonstrated significant theta-gamma coupling. However, the strength of this coupling was not different between congruent and incongruent trials for both the ACC electrode shown in Fig. A.3 ($P = 0.61$, permutation test) and across the population of conflict-selective electrodes ($P = 0.52$, sign-rank test).

A.3 Discussion

We used intracranial field potentials to measure the dynamics of conflict responses across frontal cortex leading up to the behavioral response in the Stroop task. Previous physiological and functional

neuroimaging studies have documented the involvement of multiple of these frontal cortex target areas in the Stroop or similar tasks (Botvinick et al., 1999; MacDonald, 2000; Niendam et al., 2012; Oehrns et al., 2014; Sheth et al., 2012). The intracranial field potential recordings reported here show conflict-selective signals in ACC (e.g. Fig. A.3), dlPFC (e.g. Fig. A.8), mFC (e.g. Fig. A.5B) and OFC (e.g. Fig. A.9). The mFC and dlPFC has been previously implicated in cognitive control, and these structures are extensively connected to the rest of frontal cortex areas (Ridderinkhof et al., 2004). The role of the OFC in cognitive control during Stroop-like tasks has not been reported previously, possibly because of technical challenges in neuroimaging near this area (Weiskopf et al., 2006).

We presented several lines of evidence that demonstrate that these conflict-selective physiological signals are relevant for behavior during the Stroop task. Longer behavioral reaction times were correlated with greater gamma power on a trial-by-trial basis during the Stroop task but not during the Reading task, even after accounting for trial history and for differences between congruent and incongruent stimuli (Fig. A.3H, A.5C). The same identical stimuli can elicit a range of behavioral reaction times and this internal degree of conflict can be captured, at least partly, by the strength of gamma power in frontal cortex in each trial.

The neural correlates of behavioral adaptation (Gratton effect) were observed in the ACC, consistent with prior studies based on human single neuron recordings (Sheth et al., 2012), neuroimaging (Botvinick et al., 1999; Kerns, 2006) and also in accordance with the behavioral effects of ACC resection (Sheth et al., 2012). Conflict responses in OFC and mFC also demonstrated the neural Gratton effect, suggesting a more distributed network involved in across-trial adaptation than previously hypothesized. The physiological responses in these areas were stronger in cI trials (incongruent trials that were preceded by congruent trials) than iI trials (Fig. A.5D). While the increased activity in cI trials compared to iI trials is consistent with neuroimaging studies (Botvinick et al., 1999), single neuron recordings in a different Stroop-like task report the opposite relationship (iI > cI) (Sheth et al., 2012). These differences point to potentially interesting distinctions between the activity of

individual neurons and coarser population measures that warrant further investigation.

Another discrepancy between neuroimaging studies and single unit recordings is the presence of conflict responses and error signals. Single unit recording in macaque ACC typically find error monitoring signals but not conflict-selective responses (Cole et al., 2009; Emeric et al., 2010; Ito et al., 2003; Taylor et al., 2006), whereas human neuroimaging studies observe both types of signals in ACC. There has been significant debate concerning whether action monitoring and conflict detection represent distinct processes (Carter et al., 1998; Nee et al., 2011; Swick and Turken, 2002). Because both processes may co-occur on the same trials, high temporal resolution is required to disassociate the two computations. A recent human intracranial study has found error signals in supplementary motor area and medial frontal cortex (Bonini et al., 2014), and a single unit study reported conflict signals in ACC (Sheth et al., 2012), but their co-existence in the same region is unknown. Our analysis of the few self-correction trials in our data suggests that the same areas responsible for pre-behavioral conflict signals can also produce post-behavioral response error-monitoring signals (Fig. A.6). These results are consistent with computational models suggesting that these signals may represent a general error-likelihood prediction, of which conflict and error detection are special cases (Brown and Braver, 2005).

Besides the high gamma band, we also observed conflict responses in the beta and theta bands, but not the low gamma band (e.g. Fig. A.9 and Fig. A.10). Previous work has suggested differential roles for distinct oscillatory components of the local field potential (Cavanagh and Frank, 2014; Kahana et al., 2001; Ullsperger et al., 2014; von Stein and Sarnthein, 2000). There were clear differences in the type of information conveyed by distinct frequencies components. Lack of significant correlations with reaction time in the theta and beta bands suggests that the gamma band better captures the behavior. Additionally, conflict responses were characterized by increased power in the gamma band, but decreased power in the theta and beta bands (Fig. A.10). Previous scalp EEG recordings (Cavanagh and Frank, 2014; Ullsperger et al., 2014; van Driel et al., 2015) have demonstrated that

conflict and/or error trials elicit increased theta power, suggesting potentially interesting differences in how theta is captured across spatial scales. We also observed a decrease in beta power, which is consistent with previous studies that correlate frontal cortex activation with desynchronization in the beta band and increased synchronization in the gamma bands (Crone et al., 1998b,a). Differences across tasks, recording methods, and targeted regions should be interpreted with caution. The roles of different oscillatory components in neocortex are not clearly understood. One possibility is that lower frequency bands reflect the summed dendritic input of the nearby neural population (Logothetis et al., 2001; Mitzdorf, 1987) and can act as channels for communication (Cavanagh and Frank, 2014), whereas higher frequency bands represent the population spiking rate (Buzsaki et al., 2012; Ray and Maunsell, 2011). Along these lines, we speculate that the theta desynchronization we observe could reflect a reduction of inputs, leading to inhibition of the prepotent but erroneous response.

While we observed conflict responses throughout frontal cortex, the spatiotemporal resolution of our intracranial recordings allowed us to separate regions by the latency at which conflict-selective responses emerge with respect to speech onset. By comparing pairs of simultaneously recorded electrodes, we found that conflict responses in the ACC lead the dlPFC by 200 ms. Medial frontal cortex is anatomically close and extensively connected to the ACC, and the two regions are often grouped together (Cavanagh et al., 2009; Ridderinkhof et al., 2004). Yet, conflict responses in the mFC trail the ACC by hundreds of milliseconds, suggesting an important distinction between the two regions (Rushworth et al., 2004). The relative latency measurements place the OFC at the bottom of this cascade.

Since the local field potential pools over many neurons, latency measures can be influenced by a variety of factors, such as the proportion of neurons selective for conflict and their laminar organization. Yet, at least in the ACC, the temporal profile of conflict responses we observed is similar to responses from human single unit recordings (Sheth et al., 2012). The relatively long delays between

regions are also particularly intriguing. There are monosynaptic connections that link these four regions within frontal cortex and yet, it takes 100-200 ms to detect the relative activation in these areas (Fig. A.7).

Daily decisions require integration of different goals, contexts, input signals, and the consequences of the resulting actions. The current study provides initial steps to elucidate not only which brain areas participate in cognitive control on a trial-by-trial basis but also their relative interactions and differential roles. The relative latency measurements and correlations between neural activity and reaction time provide a framework to constrain theories of cognitive control, and propose a plausible flow of conflict responses through frontal cortex.

Subject	ID	Age	Gender	# Electrodes	# Trials	Trial length (s)	Language
1	j00018	42	F	87	202	3	English
2	j00023	27	M	132	414	2	English
3	j00024	30	M	64	337	2	English
4	j00025	32	F	152	439	2	English
5	j00029	50	M	101	470	2	English
6	j00031	41	F	87	213	2	English
7	j00033	39	M	120	470	2	English
8	m00098	45	F	112	375	2	English
9	m00100	10	F	122	461	2	English
10	m00103	10	F	104	431	2	English
11	tw0005	30	M	64	360	2	Mandarin
12	tw0007	34	M	40	364	2	Mandarin
13	tw0009	19	M	64	452	2	Mandarin
14	tw0012	13	M	84	371	2	Mandarin
15	tw0014	24	F	64	470	2	Mandarin

Table A.1: Table of subjects

List of subjects that participated in these studies including their age, gender, number of electrodes implanted, the number of trials completed, the duration of the stimulus and the language in which the experiments were conducted.

	Subject														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ACC	13			14					3						
mFC	22	2	1	26	8	5			23	4				6	16
dIPFC	8	2		24	8	14	7	14	7	13	15	12	9	16	19
OFC	3			23	17	3	11	22	5	16	15	14	16	10	1
Other	41	130	63	65	68	65	102	76	84	71	34	14	39	52	24

Table A.2: Electrode Distribution

Distribution of electrode locations across the n=15 subjects that participated in this study, sorted by regions of interest: anterior cingulate cortex (ACC), medial frontal cortex (mFC), dorsolateral prefrontal cortex (dIPFC), orbitofrontal cortex (OFC), and Other.

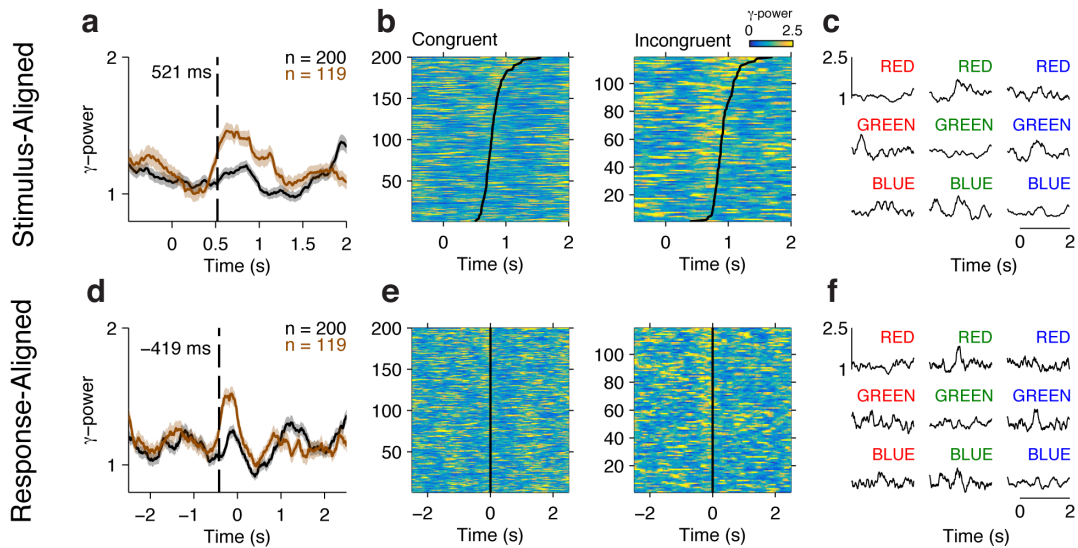


Figure A.8: Example conflict-selective electrode in the right dorsolateral Prefrontal Cortex

(A) Average gamma power signals aligned to the stimulus onset from an electrode during the Stroop task, for congruent (black) or incongruent (brown) stimuli. Shaded areas indicate s.e.m. The total number of trials for each condition is indicated in the upper right. Dashed line indicates the latency at which incongruent and congruent trials can be discriminated (see Methods).

(B) Single-trial data for congruent (left) and incongruent (right) trials. Each row is a trial, and the color indicates the normalized gamma power (color scale on upper right). Trials are sorted by behavioral response time (black line).

(C) Neural responses elicited by each of the 9 possible stimulus combinations. Congruent trials are the diagonal elements.

(D-F) Same as in A-C, but aligning the data to behavioral response time. The conflict responses are particularly strong for specific word combinations (F), such as “RED” (colored green) versus “RED”(colored red) or “RED” (colored blue).

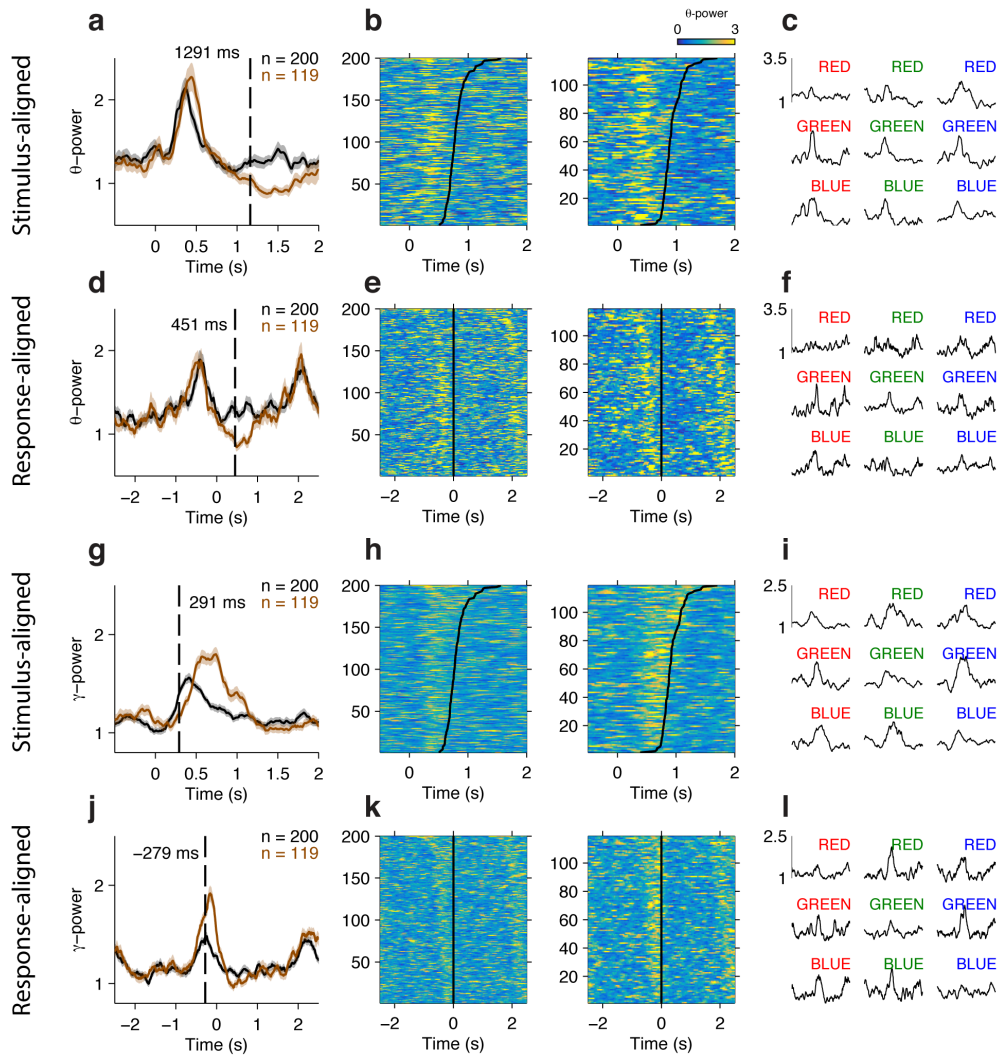


Figure A.9: Example conflict-selective electrode in the Orbitofrontal Cortex comparing responses in the Theta and Gamma Bands

(A-F) Responses in the theta power frequency band. Same format as Fig. A.8.

(G-L) Responses in the gamma power frequency band. Same format as Fig. A.8

This electrode showed an early stimulus-aligned theta power response, but the conflict-selectivity only emerged 450ms after the behavioral response (A). In contrast, the gamma power activity was better aligned to the response, and the conflict responses were present 279 ms before the behavioral response (H).

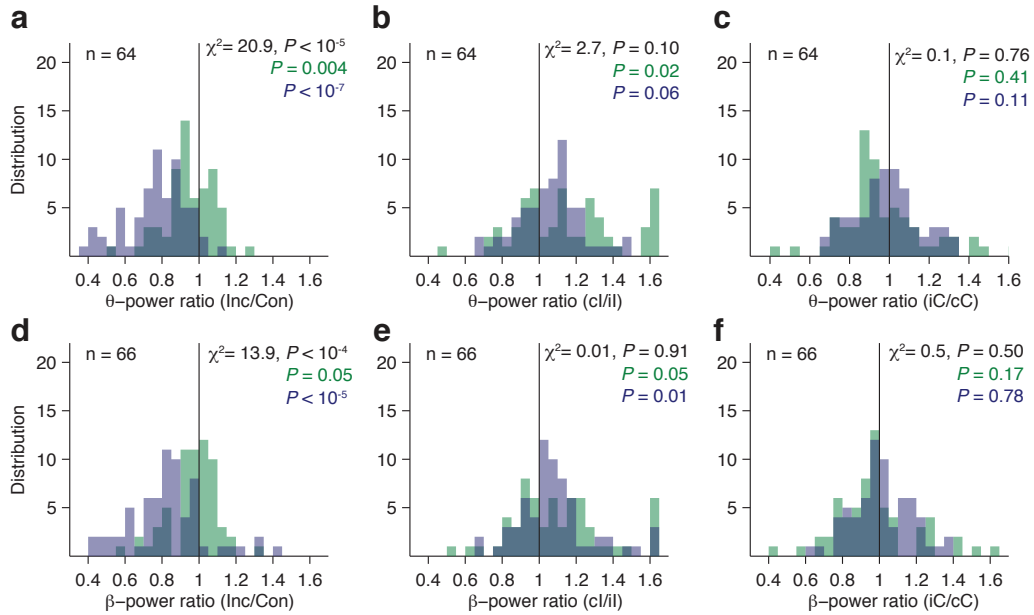


Figure A.10: **Theta and Beta band population results**

(A) Distribution of theta power ratio (Incongruent/Congruent) for the Stroop task (blue) and Reading task (green). Bin size = 0.05.

(B) Distribution of the gamma power ratio between incongruent trials preceded by congruent trials (ci) compared to incongruent trials preceded by incongruent trials (iI).

(C) Distribution of the gamma power ratio between congruent trials preceded by incongruent trials (iC) compared to congruent trials preceded by congruent trials (cC).

(D-F) Same as (A-C), but for power in the beta band.

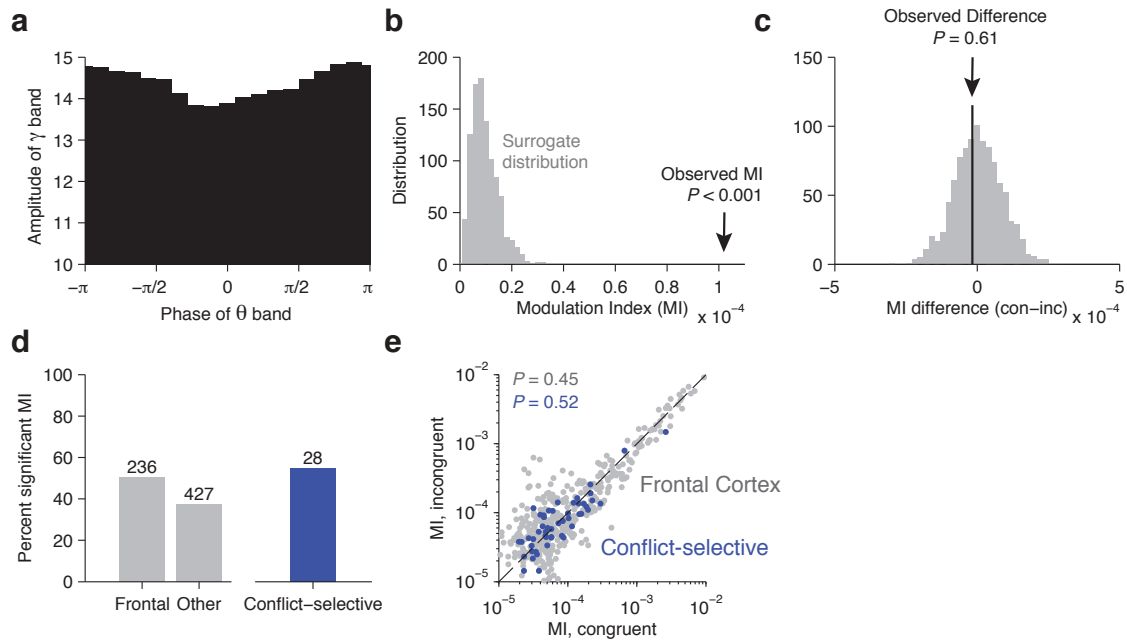


Figure A.11: Cross-frequency coupling analyses

For the anterior cingulate cortex electrode in Figure A.3:

(A) Phase-amplitude distribution during the Stroop task for the anterior cingulate example electrode shown in Figure A.3 (see Methods for calculation of cross-frequency coupling).

(B) The observed Modulation Index (MI, black arrow) is significantly greater than the surrogate distribution generated by adding a lag between the phase and amplitude measurements, demonstrating that the amplitude of the gamma band is strongly coupled to the phase of the theta band.

(C) During the Stroop task, the difference in Modulation Index between congruent and incongruent trials (black arrow) was not significantly different from 0 ($P = 0.61$). The null distribution (gray bars) was generated by randomly permuting the congruent and incongruent labels.

Across the population of electrodes:

(D) The percent of total electrodes in each region (Frontal cortex or non-frontal cortex) that had significant phase-amplitude coupling. Shown on the right is the percentage of the $n = 51$ conflict selective electrodes that showed significant coupling.

(E) The MI of congruent compared to incongruent trials for all Frontal cortex electrodes (gray dots) and the subset that were conflict-selective in the gamma band (blue dots). For both groups, there was no significant difference in the MI between congruent and incongruent trials (Frontal Cortex, $P = 0.45$; Conflict-selective, $P = 0.52$; signed-rank test). For this comparison, the number of congruent and incongruent trials was equalized before computing the MI.

A.4 Methods

A.4.1 Subjects and Recordings

Subjects were 15 patients (10 male, Ages 10-50, Table A.1) with pharmacologically intractable epilepsy treated at Children’s Hospital Boston (CHB), Johns Hopkins Medical Institution (JHMI), Brigham and Women’s Hospital (BWH), or Taipei Veterans General Hospital (TVGH). These subjects were implanted with intracranial electrodes in frontal cortex for clinical purposes. Five other subjects participated in this task but they were excluded from the analyses because they did not have any electrodes in frontal cortex. All studies were approved by each hospital’s institutional review boards and were carried out with the subjects’ informed consent.

Subjects were implanted with 2mm diameter intracranial subdural electrodes (Ad-Tech, Racine, WI, USA) that were arranged into grids or strips with 1 cm separation. Electrode locations were determined by clinical considerations. There were 1,397 electrodes (15 subjects). Sampling rates ranged from 256 Hz to 1000 Hz depending on the equipment at each institution: CHB (XLTEK, Oakville, ON, Canada), BWH (Bio-Logic, Knoxville, TN, USA), JHMI (Nihon Kohden, Tokyo, Japan), and TVGH (Natus, San Carlos, CA). All the data were collected during periods without any seizure events or following any seizures.

Electrodes were localized by co-registering the preoperative magnetic resonance imaging (MRI) with the postoperative computer tomography (CT) (Destrieux et al., 2010; Liu et al., 2009; Tang et al., 2014). In 4 subjects without a postoperative CT, electrodes were localized using intraoperative photographs and preoperative MRI. For each subject, the brain surface was reconstructed from the MRI and then assigned to one of 75 regions by Freesurfer. Depth electrodes were assigned to either a subcortical structure or to gyri/sulci. We focused on those electrodes in visual cortex and in four frontal cortex regions (ACC: anterior and middle-anterior cingulate gyrus, mFC: superior frontal

gyrus, dlPFC: middle frontal gyrus, and OFC: orbitofrontal gyrus).

A.4.2 Task procedures

A schematic of the task is shown in Fig. A.1. After 500 ms of fixation, subjects were presented with a word stimulus for 2 seconds. The stimulus presentation was 3 seconds in two subjects. Stimuli were one of three words (Red, Blue, Green) presented in the subjects' primary language (CHB, BWH, JHMI: English; TVGH: Mandarin) either in red, blue, or green font color. Stimuli subtended approximately 5 degrees of visual angle and were centered on the screen. Trials were either congruent (C), where the font color matched the word, or incongruent (I), where the font color conflicted with the word. The order of congruent and incongruent trials was randomized. Approximately 40% of the trials were incongruent trials. Within congruent trials and within incongruent trials all color-word combinations were counter balanced and randomly interleaved. Subjects were asked to either name the color (Stroop task) or read the word (Reading task) within the time limit imposed by the stimulus presentation time.

Each block contained 18 trials, and the two tasks were completed in separate blocks. Most subjects completed 18 blocks of the Stroop task and 9 blocks of the Reading task (Table A.1). Audio was recorded using a microphone at 8192 Hz sampling rate. No correct/incorrect feedback was provided.

A.4.3 Behavioral Analyses

To determine the behavioral reaction time for each trial, the short-time energy was computed from the audio recordings. For an audio signal $x(t)$, the short-time energy $E(t)$ is defined as:

$$E(t) = \sum_{m=0}^{m=T} (x(m)(t - m))^2 \quad (\text{A.1})$$

where T is the length of the recording and $w(t)$ is a 300-point Hamming window (40 ms). Speech

onset was defined as the first time when the energy crossed a threshold set as 1 standard deviation above the baseline. Only trials where the subject gave a single verbal response and the speech onset could be identified were considered correct trials.

A.4.4 Neural Analyses

Preprocessing. Unless otherwise noted, analyses in this manuscript used correct trials only. Electrodes with significant spectral noise were excluded from analysis ($n=25$ out of 1,397 total electrodes). For each electrode, a notch filter was applied at 60 Hz, and the common average reference computed from all channels was subtracted. Power in the theta (4-8 Hz), beta (9-30 Hz), and high-gamma band (70-120 Hz) was extracted using a moving window multi-taper Fourier transform (Chronux toolbox) with a time-bandwidth product of five and seven tapers. The window size was 100 ms with 10 ms increments. The power was then normalized to the baseline power (500 ms before stimulus onset).

Single electrode analyses

To determine whether and when an electrode responded selectively to conflict, we used a sliding F-statistic procedure (Liu et al., 2009). Electrodes with differential responses between congruent and incongruent trials were selected by computing the F-statistic, for each time bin, comparing the neural responses between congruent and incongruent trials. Electrodes were denoted as ‘conflict selective’ if (1) the F-statistic exceeded a significance threshold for 50 consecutive milliseconds, and (2) the average neural response exceeded one standard deviation above the baseline period at least once during the trial. A null distribution generated by randomly permuting the labels was used to set the significance threshold with $P = 0.001$. The latency at which congruent and incongruent stimuli could be discriminated was defined as the first time of this threshold crossing. For the response-aligned view, only electrodes where the latency preceded the response were included in subsequent

analysis. This selection process was independently performed for each electrode in both stimulus-aligned and response-aligned analyses, and separately for the Stroop and Reading task.

We used a permutation test with 10,000 shuffles to obtain a false discovery rate for our selection process. The congruent/incongruent trial labels were randomized 10,000 times and we measured the average number of electrodes across our population that passed the selection procedure.

For the selected electrodes obtained with the procedure described above, we performed a number of within-electrode analyses. We measured single-trial correlations with behavioral reaction times, assessed the significance of interactions and simple/main effects, and controlled for confounds in measuring the neural Gratton effect.

Single-trial analysis. For single trial comparisons across conditions, signal power for each trial was computed for both response-aligned and stimulus-aligned analyses. For stimulus-aligned data, the signal power was defined as the maximal power from stimulus onset to 1 second after stimulus onset. For response-aligned analyses, the signal power was defined as the maximal power from one second before the response to the response onset. Analyses using the average power within the same window yielded similar results. Single-trial response latency was defined as the time of maximal activation relative to stimulus onset.

Interaction Effects. For conflict-selective electrodes, we measured the significance of task dependence by performing, at each time bin, an ANOVA on the gamma power with the factors Congruency and Task (Nieuwenhuis et al., 2011). The peak F-statistic of the interaction term over the pre-response window was compared against a null distribution generated by randomly shuffling the trial labels. Simple effects were tested using this same approach.

Neural Gratton Effect. We evaluated the neural signal difference between trials with different histories (e.g. cI versus iI), while removing trials with stimulus repetitions. Given that (1) reaction times are different for the cI versus iI trials (Fig. A.5) and (2) gamma power is significantly correlated with reaction time in incongruent trials (Fig. A.5), we would expect differences in gamma power in

cI versus iI trials. To control for this potential confounding effect in our measurements of trial history dependence, we applied two methods. First, for each electrode, we performed an ANCOVA on the gamma power with trial history (cI or iI, for example) as the group and reaction time as a covariate. We computed the regression line, extracted the RT-adjusted gamma power from the y-intercept and used this value in the group analysis. Second, we performed a matched reaction time analysis, where the distribution of reaction times was equalized by subsampling the trials in a histogram-matching procedure with 200ms bins. This resulted in using only 50% of the trials. The same analysis was then applied to this reaction time matched dataset.

A.4.5 Group analyses

To account for both within-subject and across-subject variance, statistical testing of the electrophysiological data was conducted with multilevel models (Aarts et al., 2008; Goldstein, 2011) (also known as random effect models). Random factors included electrodes nested within subjects. Significance of interactions and/or main effects was assessed with a likelihood ratio test against a null model excluding that particular term.

For comparison of latency across regions, we restricted our comparison to measurements made within the same subjects. We computed the latency difference for each pair of simultaneously recorded electrodes from different regions. The F-statistic of this latency difference across the groups was compared against a null distribution generated by shuffling, within each subject, the region labels ($n = 10,000$ shuffles). Post hoc testing used the Benjamin-Hochberg procedure to control for multiple comparisons.

A.4.6 Cross-Frequency Coupling

To measure cross-frequency coupling between the theta and gamma frequency bands, we used the Modulation Index (MI) defined previously (Tort et al., 2008). Activity in the theta (4-8 Hz) and

high gamma (70-120 Hz) bands was obtained with a zero-phase least-squares finite impulse response (FIR) filter. Instantaneous phase and amplitude was extracted with the Hilbert Transform. For the Stroop and Reading Task separately, the MI was computed as the Kullback-Leiber distance between the phase-amplitude histogram and a uniform distribution. For comparison between tasks, the number of trials was equalized. This MI was compared against a surrogate distribution generated by randomly lagging the time series across 1,000 repetitions. Similar results were obtained with the measure defined in Canolty et al. (Canolty et al., 2006). Results were also similar when a surrogate distribution was created by randomly pairing low-frequency phase with high-frequency power from different trials.

To compare the strength of cross-frequency coupling between congruent and incongruent conditions, we computed the difference in MI between the two conditions while equalizing the trial count. This difference was compared against a null distribution generated by randomly shuffling the congruent and incongruent labels.

B

Neural representations of memorability

We are all that we remember, yet we do not remember all that we experience. The selective filtering and interpretation of multi-sensory inputs to form episodic memories is an essential brain function that is poorly understood. Previous studies have typically used words, faces, objects, or static scenes as the individual events with which to examine memory formation (Rubin and Wenzel, 1996; Brady et al., 2008; Bahrack et al., 1975; Vogt and Magnussen, 2007). While instructive, these stimuli are devoid of spatial, temporal, and narrative context, making them very different from memory formation in natural conditions. In this study, we instead use the rich stimuli of commercial movies to probe memorability. Movies contain many elements missing from lists of words or images, such as

This chapter is a product of joint work with Matias Ison, who developed the neurophysiological experiment and collected the neural data. I performed the described analyses and designed the electrical stimulation experiment.

emotive content, narrative structure, and spatiotemporal context. We combine this complex stimuli with neurophysiological recordings and behavioral studies to probe how memorability is represented in human brain. This study is very much a work in progress, so here we briefly present preliminary results and discuss several open questions.

B.1 Methods

In the main experiment, subjects first viewed a 40 minute TV episode of the show "24" (the encoding period). Then, at various time points ranging from 15 minutes afterwards to 1 year away, subjects performed a two-alternative forced choice memory task. In each trial of this task, subjects were presented with shots from either the episode they viewed (target trials) or a separate unseen episode (foil trials), and indicated whether they remember seeing that particular shot (Figure B.1A). The episode "24" was shown mainly because the entire season takes place within a fictional day, so there are very few changes in the environment, scene, characters, clothing, etc. Therefore, both the target episode and the foil episodes are similar in visual statistics and semantic content.

Each episode was divided into shots, based on sharp visual transitions in the movie. Each shot was approximately 1-2 seconds long. Using many volunteers and aided by computer vision, we annotated every shot in the movie. These annotations ranged from visual (e.g. number of objects, character identity, actions) to audio (e.g. sounds, talking) to emotional content. One example of an annotation, character presence, is illustrated in Figure B.1B.

Two groups of subjects participated in these memory experiments. One group consisted of $n = 80$ volunteers, mostly college-aged. In the main experiment, these subjects viewed episode 1 and were repeatedly tested at several time intervals (immediately, one day, one week, one month, three months, and one year after encoding) with different target and foil shots. As expected, performance decreased over time (Figure B.2A, black line). A repeated measures ANOVA on performance with time as a

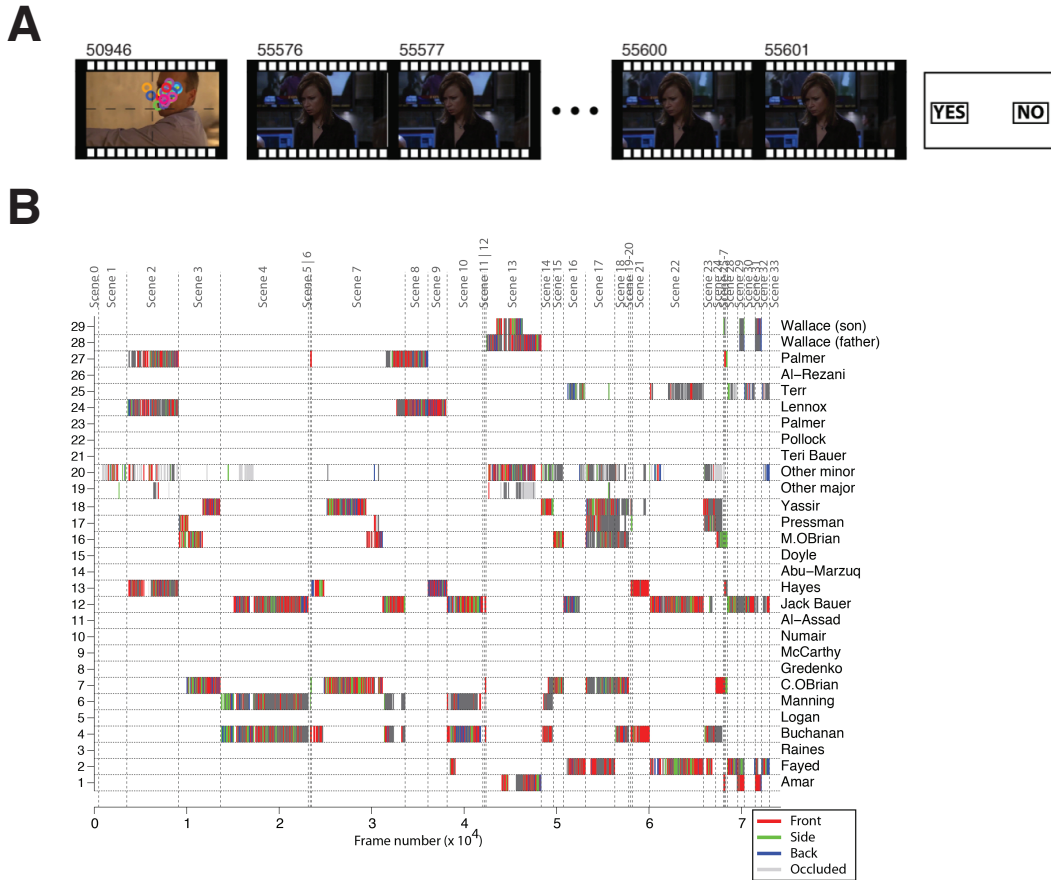


Figure B.1: Memorability experiment

(A) During the encoding phase, subjects viewed a 40 minute TV episode of the show "24". During each memory trial, subjects were presented with a single shot, either from the episode they viewed or a different foil episode. Subjects indicated whether or not they remember that particular shot in a two-alternative forced choice task.

(B) Annotations indicating the presence of various characters for episode 1 of "24" over the time of the movie (x-axis, frames). Each row denotes a different character, and color indicates the viewpoint (red = front, green = side, blue = back, gray = occluded). Vertical dashed lines indicate scene changes. Note that there are multiple shots within each scene. Other movie content was also annotated.

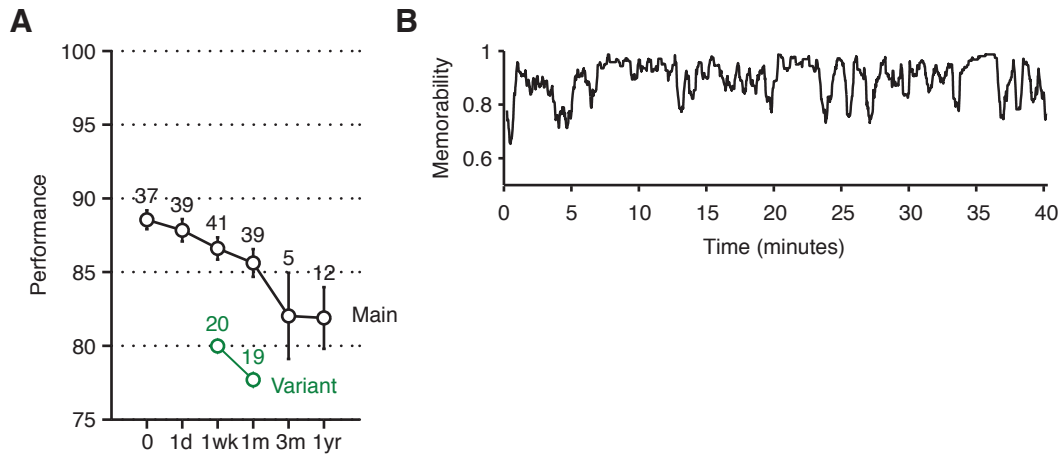


Figure B.2: **Behavioral performance**

(A) Performance for each session for subjects in the main experiment (black lines), and the variant experiment (green lines). Each session occurs at a different time since encoding (0 = immediately, d = day, wk = week, m = month, yr = year). Numbers of subjects that participated in each subject is printed above each datum. Error bars indicate s.e.m. Chance is 50% (not shown).

(B) Memorability over the time course of the 40 minute TV episode. For each shot, we computed the memorability as the average performance of the ~ 40 subjects that participated in our main experiment. Here, data is collapsed over the first four sessions to increase coverage, and the memorability is smoothed with a 15-second gaussian kernel.

factor revealed a significant effect of time ($F = 4.03, P = 0.001$). One concern in this design is that subjects are repeatedly tested on this task. While subjects are not given feedback, and different shots are always tested, we cannot rule out possible training effects of repeated testing. We therefore ran a variant experiment ($n = 39$ subjects) where subjects were only brought in for testing once -- either at one week or at one month post-encoding. Performance was significantly decreased in this task without repeated training (Figure B.2A, green line). With this large-scale psychophysics experiment, we were able to densely sample each shot of the episode about 40 times with the main experiment ($1000 \text{ shots} \times 40 \text{ trials} = 40,000 \text{ trials}$). We define the memorability of each shot as the average correct rate across subjects, generating a measure of memorability over time (Figure B.2B). While overall performance was high, there are significant and consistent reductions in memorability,

supporting the finding that our memory is consistently selective.

A second group consisted of $n = 11$ epilepsy patients implanted with intracranial electrodes (see Chapter 2). These patients differ from the psychophysics subjects described above in several important ways. Because of time limitations in a clinical setting, we were only able to obtain a small sampling of 100 target trials per subject, as opposed to approximately 1000 target trials per subject in the previous group. In addition, we tested the patients with only one session, immediately after encoding. Importantly however, in these patients the neurosurgeon implants penetrating depth electrodes (1.25 millimeter in diameter) to sample directly from medial temporal lobe. These depth electrodes, named Behnke-Fried electrodes (Fried et al., 1997), are specially designed for semi-chronic unit activity recordings. In addition to the low impedance electrodes, a bundle of 9 microwires (40 micron diameter, 300-500 k Ω impedance) are passed through the lumen of the depth electrode and splay into the target region. These microwires allow the recording of single and multi-unit neural activity. The local field potential was acquired at 30kHz (Blackrock Systems, Salt Lake City, Utah) and filtered between 300-5000Hz. Units were detected by setting a noise threshold on the high frequency activity, and the resulting waveforms were sorted using Waveclus, a semi-supervised clustering algorithm (Quiroga et al., 2004). In total, we recorded $N = 761$ single and multi-units from 17 sessions in 11 subjects, mostly from the medial temporal lobe, cingulate, and superior temporal gyrus (Figure B.3).

B.2 Correlations with memorability

We began by examining correlations between the firing rate during encoding and subjects' subsequent memorability. Although we have recorded the firing rate over the course of the entire episode, for each physiology subject only 10% of the cuts were tested for memorability, which is not enough power for analysis. We therefore correlated the firing rate against the memorability measured from psychophysics subjects, which contains a dense sampling of every shot. An example unit in the

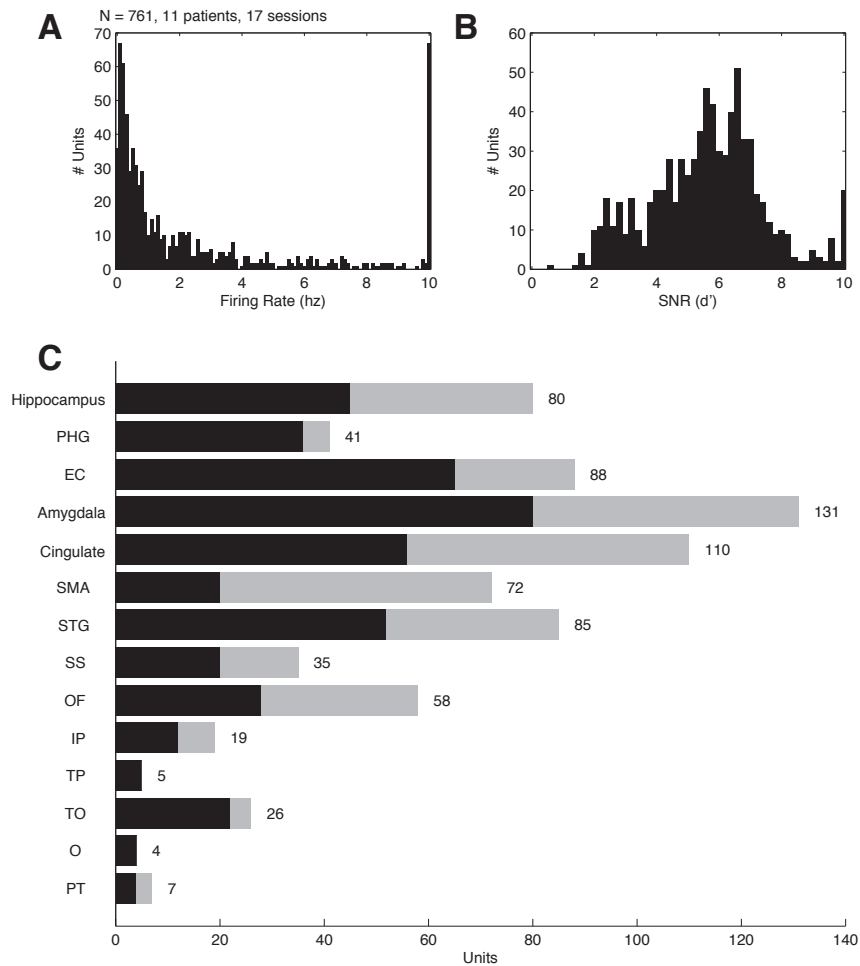


Figure B.3: Summary of recorded units

(A) We recorded $N = 761$ single and multi-units from 17 sessions in 11 subjects. The distribution of firing rate across the recorded units demonstrates a long-tailed distribution typically observed in neural recordings (Buzsaki and Mizuseki, 2014). On the log scale, this distribution is log-normal (not shown).

(B) The signal-to-noise ratio (SNR) is defined as the d-prime between the amplitude of the spike waveform and the baseline period.

(C) Distribution of recorded units over different brain regions for Episode 1 (black bar) and Episode 2 (gray bar) sessions. Abbreviations: PHG = parahippocampal gyrus, EC = Entorhinal Cortex, SMA = Supplementary Motor Area, STG = Superior Temporal Gyrus, SS = Sylvian Sulcus, OF = Orbitofrontal, IP = Inferior Parietal, TP = Temporal Pole, TO = Temporal Occipital, O = Orbital, PT = Posterior Temporal.

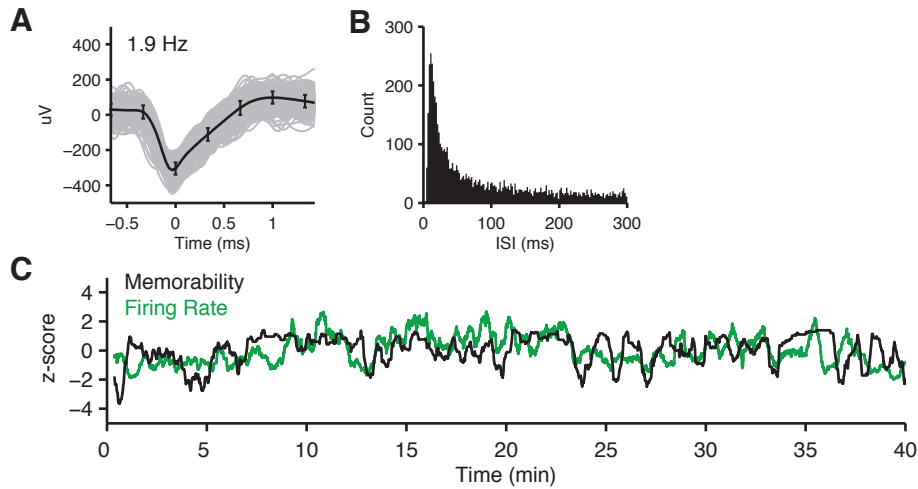


Figure B.4: Example unit in the right parahippocampal gyrus

(A) Example single waveforms (gray traces) from this unit, along with the average waveform (blackline). Error bars indicate standard deviation. The average firing rate of this unit was 1.9 Hz.

(B) Distribution of the inter-spike interval (ISI) for this unit. Note the lack of units in the refractory period (<3 ms ISI), suggesting a well isolated recording.

(C) The firing rate (green) of this unit was well correlated with the memorability (black line) throughout the episode (pearson's correlation = 0.31). Both measures were z-scored and smoothed with a 15-second Gaussian kernel.

parahippocampal gyrus which showed a correlation between memorability and firing rate is illustrated in Figure B.4. Note that the reported correlation coefficient of 0.31 was computed on the smoothed measures, and therefore inflated; correlation on the raw data was computed as 0.17.

We quantified the ability of the population neural firing rates to predict memorability with two main metrics. The first metric used a linear regression on memorability with firing rate as predictors and measured the correlation coefficient of the resulting predictions. For each shot, we computed the average firing rate during the first second of the shot, leading to 1000 features (one for each shot). We used a 95% training and 5% test set split (20 cross-validation folds). Within each training set, we greedily selected neurons based on their individual correlations with memorability and concatenated their feature vectors together. These features were then used to fit a linear model on

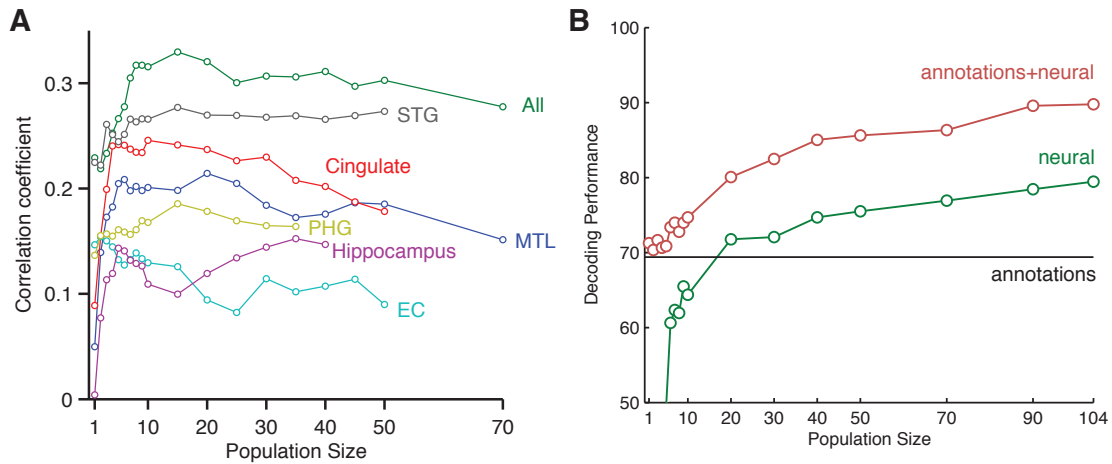


Figure B.5: **Predicting memorability from neural data**

(A) The correlation coefficient of a linear regression on memorability with neural firing rate as predictors. This measure is shown for different population sizes (x-axis) and regions (colors). The model was trained with 20 cross-validation folds, with units greedily selected based on their individual correlations on the training data. Note here that MTL (medial temporal lobe, blue) uses all the neurons in the hippocampus, entorhinal cortex (EC), parahippocampal gyrus (PHG), and amygdala.

(B) Decoding performance with increasing population sizes. Features used could be the neural data (green), annotations only (black), or a combination (red). Performance was computed using a support vector machine algorithm. Annotations were 100 features, including character presence, action, object and emotional content.

the training data and validated on the test set using the correlation coefficient. More sophisticated methods with a general linear model and/or L1 or L2 regularization did not significantly improve predictions. The predictive power for different regions is illustrated in Figure B.5A. Note that a linear regression without cross-validation folds that measures correlation, but not prediction, yields a correlation coefficient of 0.50 with $n = 40$ neurons (data not shown).

Surprisingly, the regions with the highest predictive power were the superior temporal gyrus (STG), which responds to audio stimuli, and the cingulate, followed by the medial temporal lobe. We know that the presence of sound, and especially language, in a shot is highly correlated with memorability. Therefore, we can expect STG activity to also have such correlations. The cingulate has been implicated in many functions, including attentional control and expectation prediction (Shen-

hav et al., 2013). Since our memorability measure is a catch-all that could represent everything from audio to character presence to attention, is it challenging to interpret these results with our current analysis. A future path would be to use more sophisticated methods to regress out basal attentional effects or low-level stimulus/audio effects to clarify the neural representation of memorability.

A second approach measured decoding performance on single trials with firing rates as features. We binarized memorability by setting a 0.5 threshold. We used a support vector machine algorithm (radial basis function kernel), and the greedy selection process was similar to what is described above. Importantly, the labels were imbalanced, so we weighted the cost function such that misclassifications in the minority class had greater cost compared to the majority class. Decoding performance was robust, and the neural data provided additional information beyond the semantic content, as captured by the annotations (Figure B.5).

B.3 Time scales of memorability

The advantage of using continuous stimuli such as movies is that one can measure which time scales optimally code for memorability. We had previously used the average firing rate during the shot as the feature predictors, but in this analysis we expanded the window in which we computed the firing rate. For each shot that lasted from t_0 to t_1 , and a given time scale T , we computed the average firing rate over the window $w(T)$ given by:

$$w(T) = \begin{cases} [t_0, t_1 + T] & \text{if } T > 0, \\ [t_0 + T, t_1] & \text{if } T < 0. \end{cases} \quad (\text{B.1})$$

Note that T can be negative, meaning that we are asking whether firing rate before the onset of the shot correlates with the memorability of the shot. $T = 0$ corresponds to analysis presented previously. For reference, memorability itself has a correlation length of ~ 1 second. We examined a range

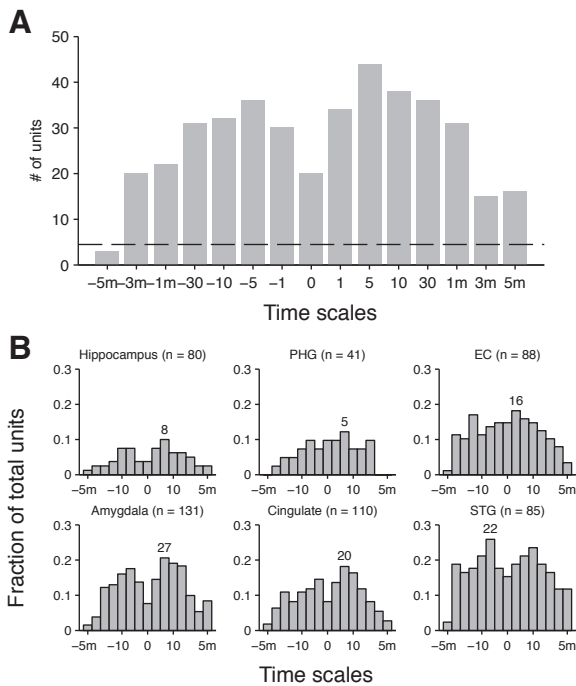


Figure B.6: Time scales of memorability

(A) For each time scale (see text for definition), the number of units (out of 449 units in episode 1 recordings) with significant correlations with memorability, as assessed with a non-parametric permutation test with a $P < 0.01$ threshold. Dashed line indicates chance level ($n = 4.5$ units).

(B) For each brain region, the fraction of total units that have significant correlations with memorability at various time scales. Abbreviations: PHG = Parahippocampal Gyrus, EC = Entorhinal Cortex, STG = Superior Temporal Gyrus. The number of units for the time scale with the highest fraction is located for each region above the appropriate time scale.

of time scales, from -5 minutes to 5 minutes. For each unit, we computed the correlation coefficient between the firing rate and memorability at these various time scales. Significance was determined for each unit by shuffling the memorability to generate null distributions for each time scale, and using a $P < 0.01$ threshold. The total number of significant units is shown in Figure B.6A for units from episode 1 sessions ($N = 449$ total units). Notably, across the overall population, neural activity on the time scale of $T \sim \pm 5$ seconds yielded the most number of significant units. However, this is not true in all brain regions (Figure B.6B); neurons in the entorhinal cortex were most significant at $T = 1$ second. The amygdala showed the largest differential between $T = 0$ and $T = \pm 5$ seconds.

This is a potentially interesting, yet puzzling, finding. Why would activity averaged over five seconds around a shot better correlate with memorability of that shot than the firing rate during the shot itself? We tried two control analyses to reject several hypothesis. One possible explanation is that the larger window is required to obtain a better estimate of the firing rate, and the memora-

bility is sufficiently smooth the support this averaging. To address this possibility, we simulated a non-homogeneous Poisson process using thinning (Lewis and Shedler, 1978), where the rate $\lambda(t)$ depends on a baseline rate λ_o modulated by the memorability $m(t)$ with a strength $\Delta\lambda$:

$$\lambda(t) = \lambda_o + 2\Delta\lambda(m(t) - 0.5) \quad (\text{B.2})$$

The ratio of λ_o to $\Delta\lambda$ determines the fidelity of the Poisson neuron's representation of the underlying memorability of the movie. We then computed the correlation coefficient with memorability at various time scales using this synthetic neuron at various noise levels ($\lambda_o = 3\text{Hz}, \Delta\lambda = 0.1, 0.5, 1, 2, 3\text{Hz}$). Even in this case, however, correlations were still maximal at $T = 0$, and approached zero as $T \rightarrow 5\text{minutes}$, suggesting that noise averaging cannot explain the phenomenon observed in our data.

Another hypothesis is that our measure of correlations become biased at longer time scales. Through a temporal shuffling procedure however, we are able to eliminate these long time scale correlations. We computed a temporal shuffle, where we shuffled the ordering of the shots. Importantly, we preserved the pairing of memorability to shots; this method only removes the temporal context. Therefore, we would expect the correlations at $T = 0$ (i.e. the duration of the shot itself) to remain the same in this shuffle. Indeed, our shuffle conserved the correlations at $T = 0$, but affected the correlations at longer time scales ($T > 1$). This confirms that temporal context is important, and that our measure is not biased for correlations at these time scales.

The correlations between memorability and firing rate on the time scale of seconds could be explained by arousal, attentional, or emotional factors that may fluctuate at this scale. However, a convincing future explanation would include a method to, given the underlying memorability, construct a Poisson neuron that exhibits these time scale effects. Such a generative model could provide a mechanistic explanation of these results.

B.4 Adventures in electrical stimulation

While electrodes are typically used to record neural activity, they can also deliver electrical current to the neural tissue. The physiological and behavioral effects of electrical stimulation vary depending on a number of factors, including the electrode properties, behavioral task, stimulation location, and stimulation parameters (amplitude, number of pulses, pulse rate, etc.). This leads to a variety of reported effects of electrical stimulation (For a review, see (Cohen and Newsome, 2004)). The most common clinical use of stimulation through macroelectrodes is part of the treatment protocol for surgical epilepsy patients. During these stimulation sessions, the neurologists will use electrical stimulation to temporarily disrupt certain brain regions and assess any effects on important cognitive abilities (reading, naming, memory, etc.). This approach, called cortical mapping, is instrumental to identifying parts of cortex to avoid during the subsequent resection (Crone et al., 1998a; Cervenka et al., 2011). Consistent with this approach, previous research studies have used macroelectrode stimulation to disrupt, for example, face processing in the fusiform face area (Parvizi et al., 2012), or memory retrieval processes (Halgren et al., 1985; Lacruz et al., 2010).

However, electrical stimulation has also been used with non-disruptive effects. In one study, Fried and colleagues report that stimulation of the entorhinal region during learning improved subsequent recognition (Suthana et al., 2012). Macaque studies have shown that microstimulation of the relevant regions can bias the monkey's decisions in various cognitive tasks such as face recognition (Afraz et al., 2006), or motion discrimination (Hanks et al., 2006). Microstimulation of rodent hippocampus has also been associated with improved memory (Bliss and Lomo, 1973; Williams and Givens, 2003). Microstimulation of visual cortex has been used to speed reaction times in a perceptual decision-making task (Ditterich et al., 2003). In a particularly interesting study, microstimulation of frontal eye fields was found to modulate V4 activity in a way qualitatively similar to visual attention modulation (Moore and Armstrong, 2003). Taken together, these studies highlight that electrical stimulation can

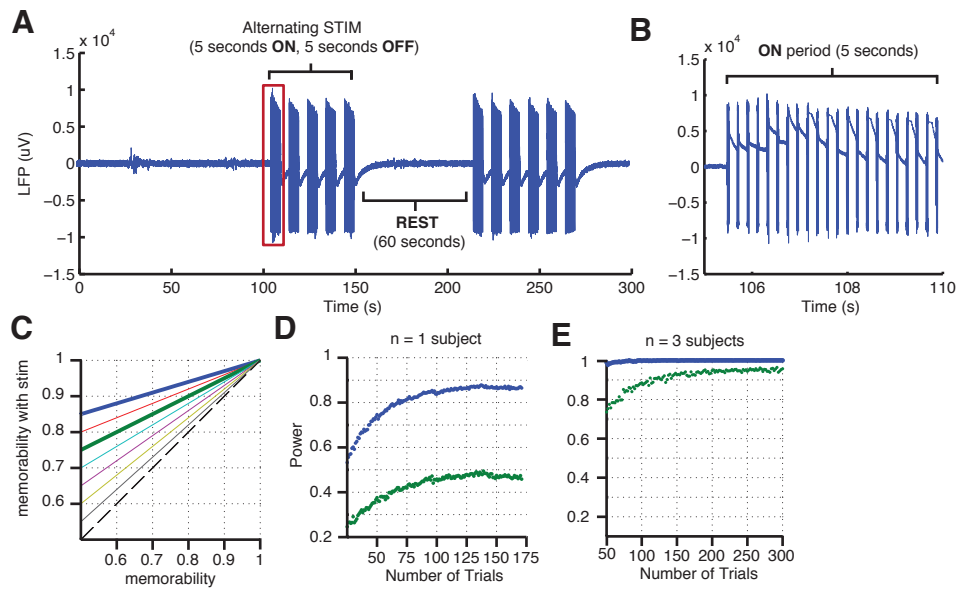


Figure B.7: Stimulation paradigm

(A) Recording from right entorhinal cortex during microwire stimulation during movie viewing showing the stimulation pattern. Shots were categorized into three period: ON, OFF, and REST.

(B) Higher temporal resolution plot of the five-second stimulation event marked in the red box in (A). For microwire stimulation, the pulses were sent at 4Hz (theta rhythm). For macroelectrode stimulation, the pulses were sent at a rate of 60Hz (not shown).

(C) Power calculations based on a range of hypothesized effects of stimulation on memorability.

(D-E) For the two hypothesized stimulation effects (see green and blue lines in (C)), the estimated power as a function of the number of trials. These simulation took into account the underlying memorability distribution of the shots. Power was computed for one subject (D) and three subjects (E).

often be a persuasive proxy for a naturally occurring signal.

To ask whether electrical stimulation can affect the formation of episodic memories during naturalistic movie viewing, we performed electrical stimulation experiments during encoding. In the macro-stimulation sessions, we sent electrical current across a pair of macroelectrode contacts. Stimulation amplitudes were determined by an on-site neurologist and ranged from 0.8-6mA. The stimulation pattern is shown in Figure B.7A-B. The stimulation period consisted of alternating stimulation with 5 seconds on (ON period) and 5 seconds off (OFF). After five cycles, there was no stimulation

for 60 seconds (REST). For analysis, shots were assigned into one of these three categories. Because we stimulated far more shots during encoding than we could test during the memory task, we carefully selected a subset of $n = 150$ shots from the three periods for subsequent testing such that the memorability distribution (as determined from the psychophysics data) is balanced. This balances the inherent difficulty level across the three periods. In this selection process, shots with lower memorability were selected first to increase the power of observing an improvement in memorability.

Given this clip selection, we used a numerical simulation to compute the power for different hypothesized effect sizes of stimulation (Figure B.7C-E). Across many repetitions ($n = 1000$), we simulated subject responses based on the assumed effect size, and then asked if we could distinguish the effect of stimulation with a threshold of $P < 0.05$. The power then represents, across these repetitions, the probability of obtaining a significant result. As shown in the figure, high power can be obtained with $n \approx 3$ subjects and approximately 150 trials. Note that this calculation assumed that each subject identically experiences the hypothesized stimulation effect. In reality, the stimulation amplitude, electrode placement, and underlying behavior may be variable across subjects.

For macroelectrode stimulation, the pulses during the ON period were sent at a rate of 60 Hz. For microwires, the stimulation amplitude is typically $150\mu\text{A}$, with pulses sent at the theta rhythm (4 Hz), motivated by studies suggesting that increased theta power is associated with improved memory (Kahana et al., 2001; Sederberg et al., 2003; Osipova et al., 2006). Stimulation targeted the right entorhinal cortex based on prior findings in the literature (Suthana et al., 2012).

The results are shown in Table B.1 for $n = 4$ subjects. Some subjects performed the memory task on two episodes, and one subject (460) was tested twice on the same episode at two time intervals (immediately afterwards, and the next day). Because of clinical considerations, the amplitude varied widely across the subjects. While there are some positive effects, there are also some negative effects of stimulation. Overall, there was no trend in the effect of stimulation. Note that here we have computed performance over all shots in each stimulation category. However, the effect of stimulation

Patient	Type	Amplitude	Sess.	Perf.	d'	SILENT	ON	OFF	
457	Macro	6mA	E01	1	0.59/0.91	1.85	62%	0%	-10%
458	Macro	2mA	E01	1	0.70/0.76	1.27	72%	-6%	0%
458	Macro	3mA	E02	1	0.49/0.91	1.73	52%	-2%	-6%
460	Micro	150uA	E01	1	0.83/0.88	2.15	80%	+5%	+5%
460				2	0.72/0.89	1.96	68%	+2%	+10%
461	Macro	0.8mA	E01	1	0.57/0.81	1.28	56%	0%	+2%
461	Macro	1mA	E02	1	0.82/0.78	1.66	80%	+6%	0%

Table B.1: **Stimulation results in right entorhinal cortex**

For each memory test session, the stimulation parameters, episode viewed, and testing session (1 = 15 minutes, 2 = 1 day). Then, the overall performance (hit rate / correct rejection), and d-prime. The last set of columns describes performance during the SILENT trials, then the differential in percentage points of performance during the ON and OFF trials.

could depend on the time from stimulation onset (e.g. a shot immediately after stimulation onset may not be affected, where a shot that occurs a few seconds into the stimulation period may yield altered memorability). Additionally, one recent study determined that the phase of theta at which stimulation occurs is critical for improving memory (Siegle and Wilson, 2014). With more data, we can take a finer look at how the shots interact with the stimulation, and the resulting impact on memorability.

B.5 Conclusion

The neural representations of memorability are under active debate, partly because electrophysiological studies of memory in humans are rare. One previous study with single items have found that spike-field coherence, but not firing rate, during the encoding period predict subsequent memorability (Rutishauser et al., 2010). In contrast, our preliminary results suggest that, with rich continuous stimuli such as commercial movies, the firing rate of neurons in medial temporal lobe and other structures can predict memorability. Of course, using richer stimuli is a double-edged sword*;

*'Richer' stimuli may also be interpreted by some scientists as 'uncontrolled' stimuli.

movies also bring to bear complex emotions, attentional effects, characters, and other narrative content that could confound our measured correlations with memorability. Single items may also carry this confound (for example, faces may be more easily remembered than non-face objects, and therefore one might find a spurious correlation between neural activity in the face area and memorability), but perhaps to a lesser extent than with these commercial movies. These effects may also be reflected in our intriguing finding that neural activity on the scale of ~ 5 seconds best correlates with memorability. The challenge moving forward will be exploring ways to tease apart these effects while still taking advantage of this rich stimuli that better approximates memory formation in natural conditions.

C

Supplemental Information

This chapter contains supplemental figures and tables to the neurophysiological recordings discussed in Chapter 3.

Subject	Age	Gender	H	# Electrodes	# Trials	% OV	Perf. (W)	Perf. (P)
1	25	M	R	72	2760	11	96	82
2	18	M	R	64	2840	18	98	92
3	12	F	R	144	1000	19	81	80
4	21	M	R	72	3680	9	99	70
5	9	F	R	104	720	9	99	77
6	27	M	R	64	3760	9	97	79
*7	17	F	R	120	760	17	98	84
8	15	M	R	109	200	11	95	73
9	40	M	R	44	1640	15	96	78
10	8	F	R	108	240	25	96	88
11	16	M	R	124	600	25	85	92
*12	16	F	R	104	4440	13	98	82
13	23	F	R	89	4000	21	99	73
14	25	M	R	105	920	27	91	69
15	16	M	R	80	865	21	97	68
16	16	F	R	108	400	27	98	76
17	12	M	L	92	1185	21	94	75
18	22	M	R	96	1120	24	99	80

Table C.1: Table of neurophysiology subjects

Each of the neurophysiology subjects that participated in the occlusion experiment. Several abbreviations were used (H = Handedness, OV = Object Visible, Perf. = Performance, W = Whole, P = Partial). Asterisk indicates subjects with simultaneous eye-tracking data.

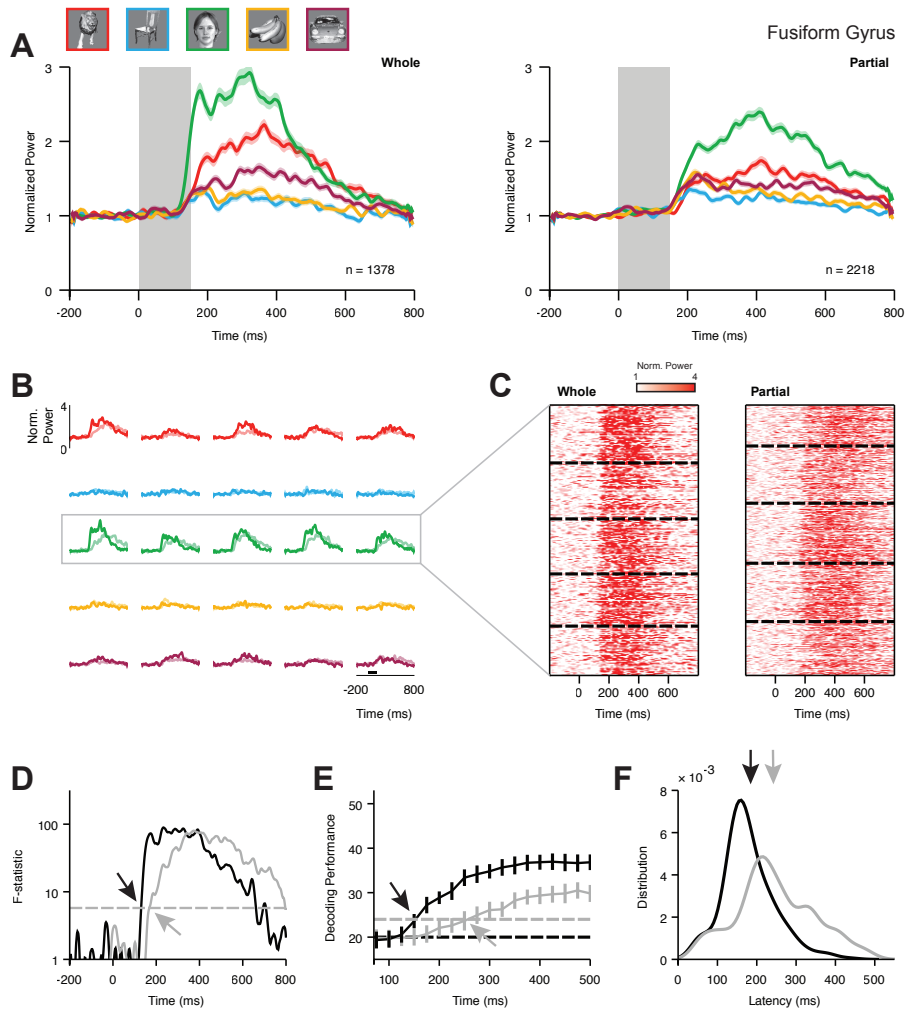


Figure C.1: Example responses in the Gamma band

Responses in the 70-100 Hz (Gamma band) for an example electrode in the left Fusiform Gyrus. The amplitude is measured as power in the Gamma band normalized against the pre-stimulus baseline.

- (A) Average response to Whole (left) and Partial (right) objects belonging to five different categories. Shaded areas indicate s.e.m. The gray rectangle denotes the image presentation time (150 ms). The total number of trials is indicated on the bottom right of each subplot.
- (B) Average responses to each of the exemplar objects (dark lines = Whole, light lines = Partial).
- (C) Raster of the neural responses for Whole (left) and Partial (right) objects for the category that elicited the strongest responses (human faces). Rows represent individual trials. The color indicates the normalized power at each time point (bin size = 2 ms, see scale on top).
- (D-F) Various measures for latency: F-statistic (D), Decoding performance (E), and response latency (F). For more detail, see Figure 3.3.

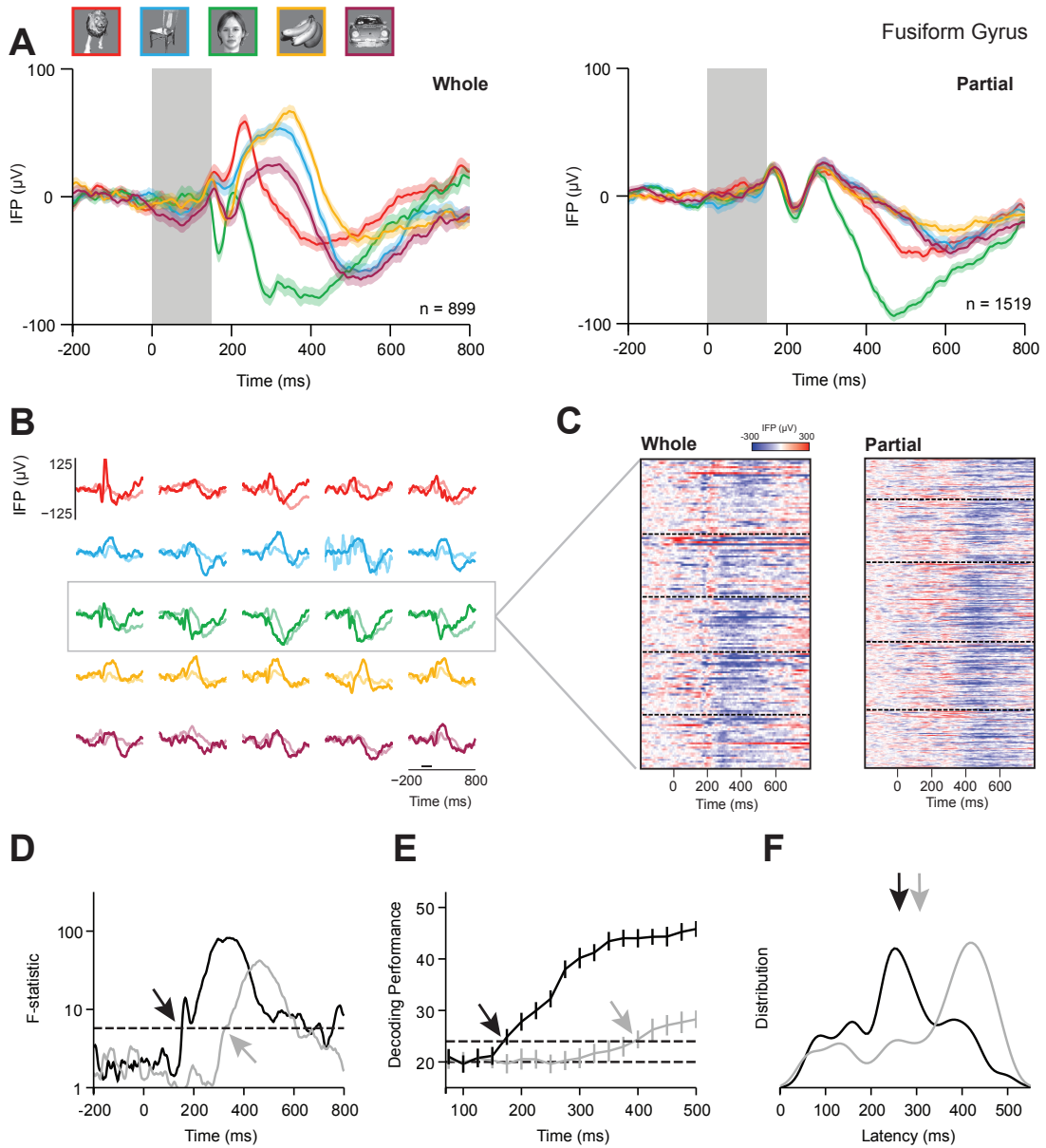


Figure C.2: Example responses in inferior temporal gyrus

Example responses from an electrode in the left Inferior Temporal Gyrus (Main experiment). The format and conventions are as in Figure C.1, except the units reflect the broadband field potential (μV). Note that the responses during the Partial condition in this example are consistent from trial to trial and still show a delay with respect to the Whole condition.

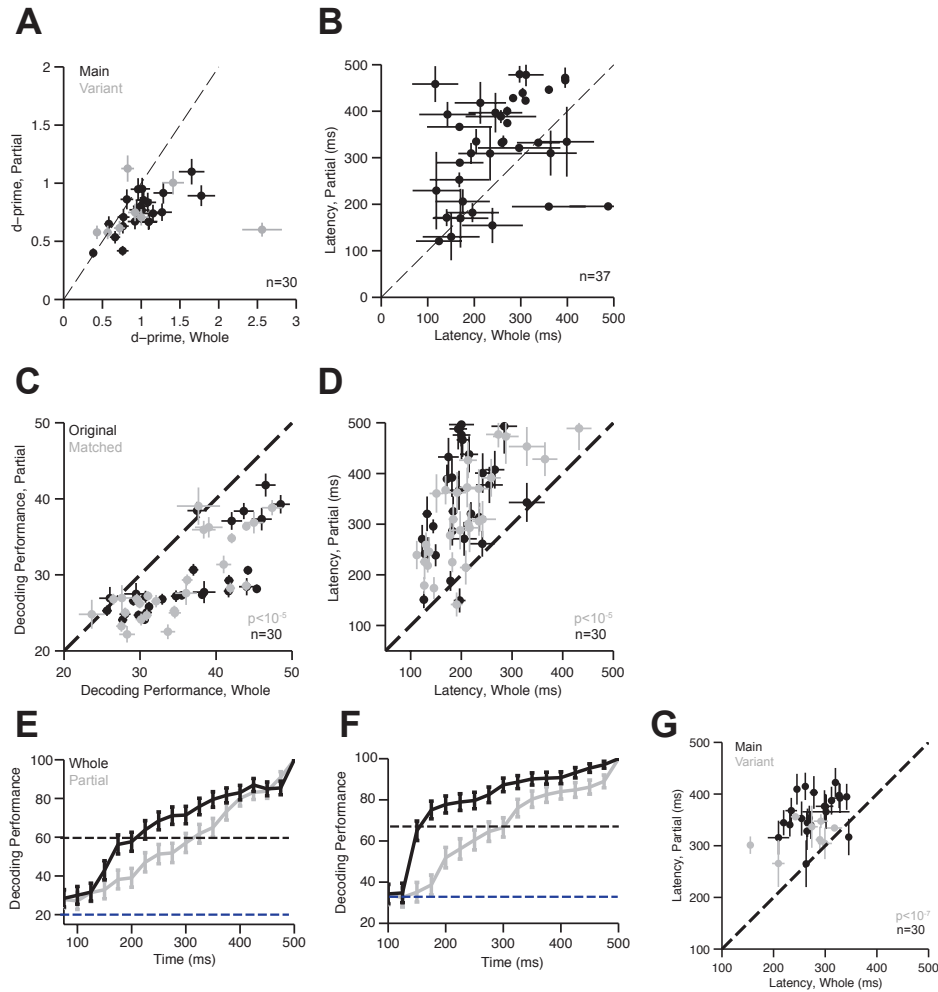


Figure C.3: d' metric, matched amplitude, and matched decoding comparisons

- (A) Comparison of d' for Partial versus Whole conditions for the $n = 30$ electrodes described in the text. d' was computed for each electrode by comparing the best versus the worst category.
- (B) Comparison of selectivity latency based on the d' metric. Shown here are the $n = 37$ electrodes that were selective in both Whole and Partial trials, as measured with d' .
- (C-D) Even after matching the distribution of the IFP amplitudes between conditions, the differences in decoding performance (C) and latencies (D) were preserved.
- (E-F) For the example electrodes in Figures 3.3 and 3.4, we computed the decoding performance over time considering only trials that were correctly decoded at 500ms. Even after matching decoding performance, latencies were delayed in the Partial condition. The latency was defined as the point where 60% (E) and 67% (F) of those trials were correctly decoded (black dashed lines). The thresholds are different because the Main and Variant experiment have different chance levels (blue dashed lines).
- (G) Even after matching the decoding performance at 500 ms, the latency difference between Whole and Partial conditions was statistically significant (rank-sum test, $p < 10^{-7}$).

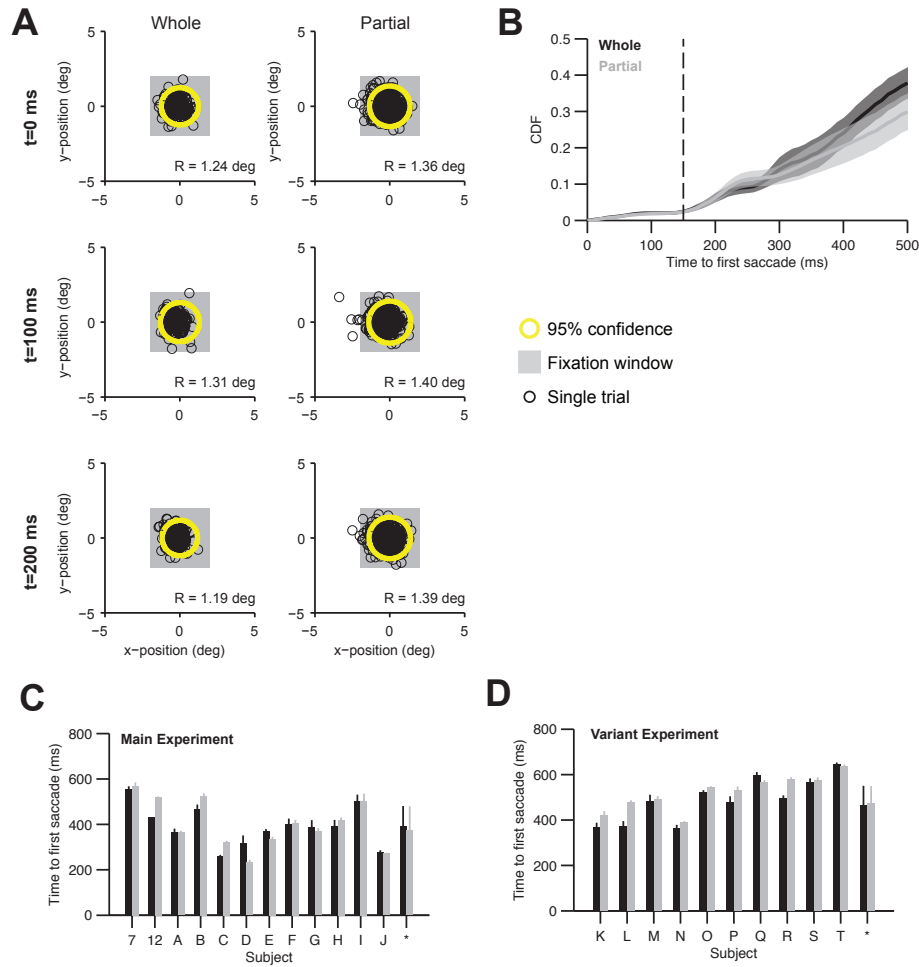


Figure C.4: Eye-tracking data and analyses

- (A) Data for one of the two subjects where we recorded eye movements simultaneously with the physiological data. Eye position for individual trials (black circles) in either the Whole (left) or Partial (right) condition, at $t = 0$ ms, $t = 100$ ms, and $t = 200$ ms from stimulus onset. The stimulus lasted 150 ms, and was approximately 5 degrees in size (gray box). The yellow circle represents 95% confidence across trials for the eye position. The radius of the confidence intervals was similar between Whole and Partial conditions.
- (B) Distribution of the time to first saccade, averaged over 22 subjects (2 subjects with concomitant physiology recordings, 20 psychophysics subjects) for the Whole (black) and Partial (gray) conditions. There was no significant difference between the distributions for the Whole and Partial conditions. Error bars denote s.e.m.
- (C) Average time to first saccade for each of the 12 subjects in the Main experiment (2 physiology subjects, 10 psychophysics subjects), as well as the group average (marked as *). Error bars denote s.e.m.
- (D) Average time to first saccade for each of the 10 psychophysics subjects in the Variant experiment, as well as the group average (marked as *). Error bars denote s.e.m.

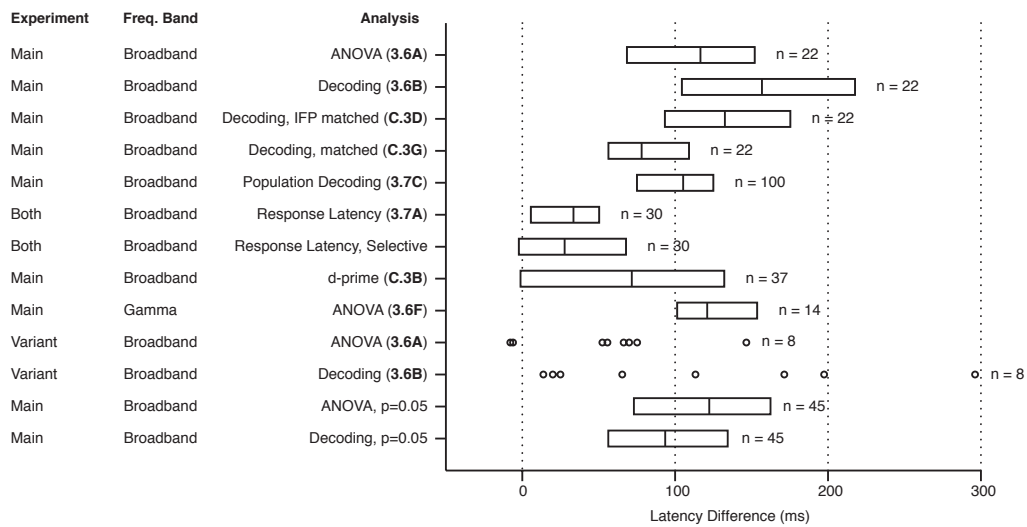


Figure C.5: Detailed summary of latency measurements

Summary of latency difference between Partial and Whole conditions for multiple combinations of experiments, frequency bands, and analyses methods (the relevant figure is indicated in parenthesis). Positive values mean increased latency in the Partial condition. Box plots represent the median and quartile across the selective electrodes. The n indicates the number of electrodes, except for the Population Decoding, where the $n = 100$ refers to the number of repetitions.

Subject	Location	Talairach	Latency (ms)		Experiment
			Whole	Partial	
1	Fusiform	41.8, -42.6, -20.7	100	320	Main
2	Inferior Occipital	42.5, -86.0, -5.3	108	229	Main
2	Middle Temporal	53.1, -73.7, 7.3	143	300	Main
2	Occipital Pole	26.3, -84.7, -16.9	167	384	Main
2	Inferior Occipital	35.9, -82.8, -14.5	154	349	Main
2	Inferior Occipital	47.7, -80.4, -10.4	175	249	Main
2	Inferior Occipital	48.7, -74.5, -5.1	139	198	Main
2	Middle Temporal	56.2, -69.0, 2.1	118	272	Main
2	Parahippocampal	30.5, -33.8, -19.4	185	244	Main
2	Fusiform	39.2, -40.0, -19.8	209	331	Main
2	Inferior Temporal	50.1 - 45.3 - 20.5	191	314	Main
2	Inferior Temporal	55.6, -48.4, -14.1	215	283	Main
2	Parahippocampal	35.7, -26.6, -23.0	138	268	Main
4	Fusiform	-32.2, -45.6, -16.1	83	219	Main
4	Fusiform	-30.2, -40.7, -32.7	132	257	Main
4	Fusiform	-21.5, -58.6, -12.6	163	303	Main
4	Fusiform	-30.7, -53.1, -16.4	160	211	Main
7	Parahippocampal	-26.2 - 23.0 - 27.5	162	236	Main
7	Temporal Pole	-35.4, -18.3, -34.8	160	201	Main
7	Middle Occipital	-45.2 - 87.9 - 6.9	259	306	Main
12	Middle Temporal	-58.7 - 67.015.5	136	288	Main
12	Fusiform	-42.0 - 42.3 - 23.2	122	222	Main
13	Fusiform	-33.7 - 43.3 - 19.0	185	255	Variant
13	Inferior Temporal	-49.0 - 46.6 - 18.7	132	279	Variant
15	Inferior Occipital	-45.6, -73.0, -10.1	241	235	Variant
15	Fusiform	-41.9, -62.6, -15.8	106	181	Variant
15	Fusiform	-26.3, -44.7, -15.8	327	379	Variant
17	Fusiform	-27.2 - 74.1 - 17.3	166	221	Variant
18	Fusiform	-40.2 - 40.7 - 26.9	264	238	Variant
18	Inferior Occipital	-43.7 - 80.5 - 0.7	108	174	Variant

Table C.2: Table of selective electrodes

Description of electrodes selective in both Whole and Partial conditions, including Talairach Coordinates. The latency value reported here is based on ANOVA.

References

- E. Aarts, M. Verhage, J. V. Veenliet, C. V. Dolan, and S. van der Sluis. A solution to dependency: using multilevel analysis to accommodate nested data. *Nat Neurosci*, 17(4):491--6, 2014.
- Esther Aarts, Ardi Roelofs, and Miranda van Turenout. Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *J Neurosci*, 28(18):4671--8, 2008.
- Seyed-Reza R. Afraz, Roozbeh Kiani, and Hossein Esteky. Microstimulation of inferotemporal cortex influences face categorization. *Nature*, 442(7103):692--5, 8 2006.
- M. Ahissar and S. Hochstein. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci*, 8(10):457--64, 2004.
- T. Allison, A. Puce, DD. Spencer, and G. McCarthy. Electrophysiological studies of human face perception. i: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex*, 9(5):415--430, 1999.
- Harry P Bahrick, Phyllis O Bahrick, and Roy P Wittlinger. Fifty years of memory for names and faces: A cross-sectional approach. *Journal of experimental psychology: General*, 104(1):54, 1975.
- J S Bakin, K. Nakayama, and C D Gilbert. Visual responses in monkey areas v1 and v2 to three-dimensional surface configurations. *J Neurosci*, 20(21):8188--98, 2000.
- R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky. Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. U.S.A.*, 92(9):3844--3848, Apr 1995.
- I. Biederman and E. E. Cooper. Priming contour-deleted images: evidence for intermediate representations in visual object recognition. *Cognit Psychol*, 23(3):393--419, 1991.
- C.M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.
- Tim VP Bliss and Terje Lomo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, 232(2):331--356, 1973.

- F. Bonini, B. Burle, C. Liegeois-Chauvel, J. Regis, P. Chauvel, and F. Vidal. Action monitoring and medial frontal cortex: leading role of supplementary motor area. *Science*, 343(6173):888--91, 2014.
- M. Botvinick, L E Nystrom, K. Fissell, C S Carter, and J D Cohen. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, 402(6758):179--81, 1999.
- Matthew M Botvinick, Todd S Braver, Deanna M Barch, Cameron S Carter, and Jonathan D Cohen. Conflict monitoring and cognitive control. *Psychological review*, 108(3):624, 2001.
- Kristofer E Bouchard, Nima Mesgarani, Keith Johnson, and Edward F Chang. Functional organization of human sensorimotor cortex for speech articulation. *Nature*, 495(7441):327--32, 2013.
- Timothy F Brady, Talia Konkle, George A Alvarez, and Aude Oliva. Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38):14325--14329, 2008.
- Albert S Bregman. Asking the 'what for' question in auditory perception. *Perceptual organization*, pages 99--118, 1981.
- B. G. Breitmeyer and H. Ogmen. Recent models and findings in visual backward masking: a comparison, review, and update. *Percept Psychophys*, 62(8):1572--1595, Nov 2000.
- Joshua W Brown and Todd S Braver. Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, 307(5712):1118--21, 2005.
- J. M. Bugg, L. L. Jacoby, and J. P. Toth. Multiple levels of control in the stroop task. *Mem Cognit*, 36(8):1484--94, 2008.
- T. J. Buschman and E. K. Miller. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, 315(5820):1860--2, 2007.
- G. Buzsaki and K. Mizuseki. The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.*, 15(4):264--278, Apr 2014.
- G. Buzsaki, C. A. Anastassiou, and C. Koch. The origin of extracellular fields and currents--eeg, ecog, lfp and spikes. *Nat Rev Neurosci*, 13(6):407--20, 2012.
- Charles F Cadieu, Ha Hong, Daniel L K Yamins, Nicolas Pinto, Diego Ardila, Ethan A Solomon, Najib J Majaj, and James J DiCarlo. Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS Comput Biol*, 10(12):e1003963, 2014.

- E. M. Callaway. Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Netw*, 17(5-6):625--32, 2004.
- R. T. Canolty, E. Edwards, S. S. Dalal, M. Soltani, S. S. Nagarajan, H. E. Kirsch, M. S. Berger, N. M. Barbaro, and R. T. Knight. High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, 313(5793):1626--8, 2006.
- M. Carandini and D. L. Ringach. Predictions of a recurrent model of orientation selectivity. *Vision Res.*, 37(21):3061--3071, Nov 1997.
- GA Carpenter and S Grossberg. *Adaptive Resonance Theory*. MIT Press, Cambridge, 2nd edition, 2002.
- C. S. Carter, T. S. Braver, D. M. Barch, M. M. Botvinick, D. Noll, and J. D. Cohen. Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280(5364):747--9, 1998.
- J. F. Cavanagh and M. J. Frank. Frontal theta as a mechanism for cognitive control. *Trends Cogn Sci*, 18(8):414--21, 2014.
- J. F. Cavanagh, M. X. Cohen, and J. J. Allen. Prelude to and resolution of an error: Eeg phase synchrony reveals cognitive control dynamics during action monitoring. *J Neurosci*, 29(1):98--105, 2009.
- Mackenzie C Cervenka, Dana F Boatman-Reich, Julianna Ward, Piotr J Franaszczuk, and Nathan E Crone. Language mapping in multilingual patients: electrocorticography and cortical stimulation during naming. *Front Hum Neurosci*, 5:13, 2011.
- J. Chen, T. Zhou, H. Yang, and F. Fang. Cortical dynamics underlying face completion in human visual system. *J Neurosci*, 30(49):16692--8, 2010.
- Juan Chen, Bingyun Liu, Bing Chen, and Fang Fang. Time course of amodal completion in face perception. *Vision research*, 49(7):752--758, 2009.
- Jonathan D Cohen and David Servan-Schreiber. Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychological review*, 99(1):45, 1992.
- Marlene R Cohen and William T Newsome. What electrical microstimulation has revealed about the neural basis of cognition. *Curr Opin Neurobiol*, 14(2):169--77, Apr 2004.

- Michael W Cole, Nick Yeung, Winrich A Freiwald, and Matthew Botvinick. Cingulate cortex: diverging data from humans and monkeys. *Trends Neurosci*, 32(11):566--74, 2009.
- C. E. Connor, S. L. Brincat, and A. Pasupathy. Transformation of shape information in the ventral pathway. *Curr Opin Neurobiol*, 17(2):140--7, 2007.
- N E Crone, D L Miglioretti, B. Gordon, and R P Lesser. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. ii. event-related synchronization in the gamma band. *Brain*, 121 (Pt 12):2301--15, 1998a.
- N. E. Crone, D. L. Miglioretti, B. Gordon, J. M. Sieracki, M. T. Wilson, S. Uematsu, and R. P. Lesser. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. i. alpha and beta event-related desynchronization. *Brain*, 121 (Pt 12):2271--99, 1998b.
- I. Davidesco, M. Harel, M. Ramot, U. Kramer, S. Kipervasser, F. Andelman, M. Y. Neufeld, G. Goelman, I. Fried, and R. Malach. Spatial and object-based attention modulates broadband high-frequency responses across the human visual cortical hierarchy. *J Neurosci*, 33(3):1228--40, 2013.
- G. Deco and E. T. Rolls. A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res*, 44(6):621--42, 2004.
- R Desimone, TD Albright, CG Gross, and C Bruce. Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, 4(8):2051--2062, 1984.
- C. Destrieux, B. Fischl, A. Dale, and E. Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1--15, 2010.
- J. J. DiCarlo, D. Zoccolan, and N. C. Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415--34, 2012.
- Jochen Ditterich, Mark E Mazurek, and Michael N Shadlen. Microstimulation of visual cortex affects the speed of perceptual decisions. *Nat Neurosci*, 6(8):891--8, Aug 2003.
- G. M. Doniger, J. J. Foxe, M. M. Murray, B. A. Higgins, J. G. Snodgrass, C. E. Schroeder, and D. C. Javitt. Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes. *J Cogn Neurosci*, 12(4):615--21, 2000.
- R. J. Douglas and K. A. Martin. Neuronal circuits of the neocortex. *Annu Rev Neurosci*, 27:419--51, 2004.

- T. Egner and J. Hirsch. The neural correlates and functional integration of cognitive control in a Stroop task. *Neuroimage*, 24(2):539--47, 2005.
- E. E. Emeric, M. Leslie, P. Pouget, and J. D. Schall. Performance monitoring local field potentials in the medial frontal cortex of primates: supplementary eye field. *J Neurophysiol*, 104(3):1523--37, 2010.
- D.J. Felleman and D.C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1--47, 1991.
- Lesley K Fellows and Martha J Farah. Is anterior cingulate cortex necessary for cognitive control? *Brain*, 128(Pt 4):788--96, 2005.
- L. Fisch, E. Privman, M. Ramot, M. Harel, Y. Nir, S. Kipervasser, F. Andelman, M. Y. Neufeld, U. Kramer, I. Fried, and R. Malach. Neural ignition: enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron*, 64(4):562--574, Nov 2009.
- I. Fried, K. A. MacDonald, and C. L. Wilson. Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron*, 18(5):753--765, May 1997.
- K. Fukushima. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193--202, 1980.
- Harvey Goldstein. *Multilevel statistical models*. Wiley series in probability and statistics. Wiley, Chichester, West Sussex, 4th edition, 2011. ISBN 9780470748657 (print) 9780470973400 (ePDF) 9780470973394 (oBook).
- F. Gosselin and P. G. Schyns. Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Res*, 41(17):2261--71, 2001.
- Gabriele Gratton, Michael G H Coles, and Emanuel Donchin. Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121(4): 480, 1992.
- D. Green and J. Swets. *Signal detection theory and psychophysics*. Wiley, New York, 1966.
- K. Grill-Spector, Z. Kourtzi, and N. Kanwisher. The lateral occipital complex and its role in object recognition. *Vision Res*, 41(10-11):1409--22, 2001.

CG Gross, DB Bender, and CE Rocha-Miranda. Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science*, 166(3910):1303--1306, 1969.

Eric Halgren, Charles L. Wilson, and J. M. Stapleton. Human medial temporal-lobe stimulation disrupts both formation and retrieval of recent memories. *Brain and cognition*, 4(3):287--295, 1985.

Timothy D. Hanks, Jochen Ditterich, and Michael N. Shadlen. Microstimulation of macaque area lip affects decision-making in a motion discrimination task. *Nat Neurosci*, 9(5):682--9, 5 2006.

J. Hegde, F. Fang, S. Murray, and D. Kersten. Preferential responses to occluded objects in the human visual cortex. *Journal of Vision*, 8(4):1--16, 2008.

G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504--7, 2006.

J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *PNAS*, 79:2554--2558, 1982.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160:106--54, 1962.

DH Hubel and TN Wiesel. Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology (London)*, 148:574--591, 1959.

CP Hung, G Kreiman, T Poggio, and JJ DiCarlo. Fast read-out of object identity from macaque inferior temporal cortex. *Science*, 310:863--866, 2005.

E. B. Issa and J. J. Dicarlo. Precedence of the eye region in neural processing of faces. *J Neurosci*, 32(47):16666--82, 2012.

M. Ito, H. Tamura, I. Fujita, and K. Tanaka. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol*, 73(1):218--26, 1995.

Shigehiko Ito, Veit Stuphorn, Joshua W Brown, and Jeffrey D Schall. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science*, 302(5642):120--2, 2003.

Hongjun Jia and Aleix M Martinez. Face recognition with occlusions in the training and testing sets. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1--6. IEEE, 2008.

- J. S. Johnson and B. A. Olshausen. The recognition of partially visible natural objects in the presence and absence of their occluders. *Vision Res*, 45(25-26):3262--76, 2005.
- Kevin Johnston, Helen M Levin, Michael J Koval, and Stefan Everling. Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. *Neuron*, 53(3):453--62, 2007.
- M. J. Kahana, D. Seelig, and J. R. Madsen. Theta returns. *Curr Opin Neurobiol*, 11(6):739--44, 2001.
- Gaetano Kanizsa. *Organization in vision: Essays on Gestalt perception*. Praeger Publishers, 1979.
- E.S. Keeping. *Introduction to Statistical Inference*. Dover, New York, 1995.
- Philip J Kellman, Sharon E Guttman, and Thomas D Wickens. Geometric and neural models of object. *From fragments to objects: Segmentation and grouping in vision*, 130:183, 2001.
- John G Kerns. Anterior cingulate and prefrontal cortex activity in an fmri study of trial-to-trial adjustments on the simon task. *Neuroimage*, 33(1):399--405, 2006.
- John G Kerns, Jonathan D Cohen, Angus W MacDonald, Raymond Y Cho, V Andrew Stenger, and Cameron S Carter. Anterior cingulate conflict monitoring and adjustments in control. *Science*, 303(5660):1023--6, 2004.
- C. Keysers, D.K. Xiao, P. Foldiak, and D.I. Perret. The speed of sight. *Journal of Cognitive Neuroscience*, 13(1):90--101, 2001.
- H. Komatsu. The neural mechanisms of perceptual filling-in. *Nat Rev Neurosci*, 7(3):220--31, 2006.
- Yoshito Kosai, Yasmine El-Shamayleh, Amber M Fyall, and Anitha Pasupathy. The role of visual area v4 in the discrimination of partially occluded shapes. *J Neurosci*, 34(25):8570--84, 2014.
- G. Kovacs, R. Vogels, and G. A. Orban. Selectivity of macaque inferior temporal neurons for partially occluded shapes. *J Neurosci*, 15(3 Pt 1):1984--97, 1995a.
- G. Kovacs, R. Vogels, and G. A. Orban. Cortical correlate of pattern backward masking. *Proc. Natl. Acad. Sci. U.S.A.*, 92(12):5587--5591, Jun 1995b.
- G Kreiman. *Computational models of visual object recognition*. Taylor and Fracis Group, 2013.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097--1105, 2012.

M. E. Lacruz, A. Valentín, J. J. García Seoane, R. G. Morris, R. P. Selway, and G. Alarcón. Single pulse electrical stimulation of the hippocampus is sufficient to impair human episodic memory. *Neuroscience*, 170(2):623--32, 10 2010.

V. A. Lamme and P. R. Roelfsema. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci*, 23(11):571--9, 2000.

Y LeCun, L Bottou, Y Bengio, and P Haffner. Gradient-based learning applied to document recognition. *Proc of the IEEE*, 86(11):2278--2324, 1998.

T. S. Lee and D. Mumford. Hierarchical bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis*, 20(7):1434--48, 2003.

T S Lee and M. Nguyen. Dynamics of subjective contour formation in the early visual cortex. *Proc Natl Acad Sci U S A*, 98(4):1907--11, 2001.

Y. Lerner, T. Hendler, and R. Malach. Object-completion effects in the human lateral occipital complex. *Cereb Cortex*, 12(2):163--77, 2002.

Y. Lerner, M. Harel, and R. Malach. Rapid completion effects in human high-order visual areas. *NEuroimage*, 21:516--526, 2004.

Peter A Lewis and Gerald S Shedler. Simulation of nonhomogeneous poisson processes by thinning. Technical report, DTIC Document, 1978.

Ming Liang and Xiaolin Hu. Recurrent convolutional neural network for object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3367--3375, 2015.

H Liu, Y Agam, J.R. Madsen, and G Kreiman. Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, 62(2):281--290, 2009.

N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fmri signal. *Nature*, 412(6843):150--7, 2001.

N.K. Logothetis and D.L. Sheinberg. Visual object recognition. *Annual Review of Neuroscience*, 19: 577--621, 1996.

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

A W MacDonald. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, 288(5472):1835--1838, 2000.

J. R. Manning, J. Jacobs, I. Fried, and M. J. Kahana. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J. Neurosci.*, 29(43):13613--13620, Oct 2009.

F. A. Mansouri, M. J. Buckley, and K. Tanaka. Mnemonic function of the dorsolateral prefrontal cortex in conflict-induced behavioral adjustment. *Science*, 318(5852):987--990, 2007.

C Mehring, J Rickert, E Vaadia, S Cardosa de Oliveira, A Aertsen, and S. Rotter. Inference of hand movements from local field potentials in monkey motor cortex. *Nature Neuroscience*, 6(12):1253--1254, 2003.

B Mel. Seemore: Combining color, shape and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, 9:777, 1997.

E.M. Meyers and G Kreiman. *Tutorial on Pattern Classification in Cell Recordings*, book section 19. MIT Press, Boston, 2011.

M. P. Milham, M. T. Banich, E. D. Claus, and N. J. Cohen. Practice-related effects demonstrate complementary roles of anterior cingulate and prefrontal cortices in attentional control. *Neuroimage*, 18(2):483--93, 2003.

E K Miller and J D Cohen. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*, 24:167--202, 2001.

Earl K Miller. The prefrontal cortex and cognitive control. *Nature reviews neuroscience*, 1(1):59--65, 2000.

M. Missal, R. Vogels, and G. A. Orban. Responses of macaque inferior temporal neurons to overlapping shapes. *Cereb Cortex*, 7(8):758--67, 1997.

- U. Mitzdorf. Properties of the evoked potential generators: current source-density analysis of visually evoked potentials in the cat cortex. *Int J Neurosci*, 33(1-2):33--59, 1987.
- Tirin Moore and Katherine M Armstrong. Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, 421(6921):370--3, Jan 2003.
- D. Mumford. On the computational architecture of the neocortex. ii. the role of cortico-cortical loops. *Biol Cybern*, 66(3):241--51, 1992.
- M M Murray. Setting boundaries: Brain dynamics of modal and amodal illusory shape completion in humans. *Journal of Neuroscience*, 24(31):6898--6903, 2004.
- R. F. Murray, A. B. Sekuler, and P. J. Bennett. Time course of amodal completion revealed by a shape discrimination task. *Psychon Bull Rev*, 8(4):713--20, 2001.
- Kae Nakamura, Matthew R Roesch, and Carl R Olson. Neuronal activity in macaque set and acc during performance of tasks involving conflict. *J Neurophysiol*, 93(2):884--908, 2005.
- K Nakayama, ZJ He, and S Shimojo. *Visual surface representation: a critical link between lower-level and higher-level vision*, volume 2. The MIT press, Cambridge, 1995.
- D. E. Nee, S. Kastner, and J. W. Brown. Functional heterogeneity of conflict, error, task-switching, and unexpectedness effects within medial prefrontal cortex. *Neuroimage*, 54(1):528--40, 2011.
- Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *arXiv preprint arXiv:1412.1897*, 2014.
- K Nielsen, NK Logothetis, and G Rainer. Dissociation between lfp and spiking activity in macaque inferior temporal cortex reveals diagnostic parts-based encoding of complex objects. *Journal of Neuroscience*, 26(38):9639--9645, 2006a.
- K. J. Nielsen, N. K. Logothetis, and G. Rainer. Discrimination strategies of humans and rhesus monkeys for complex visual displays. *Curr Biol*, 16(8):814--20, 2006b.
- T. A. Niendam, A. R. Laird, K. L. Ray, Y. M. Dean, D. C. Glahn, and C. S. Carter. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn Affect Behav Neurosci*, 12(2):241--68, 2012.
- S. Nieuwenhuis, B. U. Forstmann, and E. J. Wagenmakers. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat Neurosci*, 14(9):1105--7, 2011.

- Y. Nir, L. Fisch, R. Mukamel, H. Gelbard-Sagiv, A. Arieli, I. Fried, and R. Malach. Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. *Curr. Biol.*, 17(15):1275--1285, Aug 2007.
- C. R. Oehrn, S. Hanslmayr, J. Fell, L. Deuker, N. A. Kremers, A. T. Do Lam, C. E. Elger, and N. Axmacher. Neural communication patterns underlying conflict detection, resolution, and adaptation. *J Neurosci*, 34(31):10438--52, 2014.
- G. A. Ojemann. Treatment of temporal lobe epilepsy. *Annu. Rev. Med.*, 48:317--328, 1997.
- B. A. Olshausen, C. H. Anderson, and D. C. Van Essen. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J Neurosci*, 13(11):4700--19, 1993.
- I. R. Olson, J. C. Gatenby, H. C. Leung, P. Skudlarski, and J. C. Gore. Neuronal representation of occluded objects in the human brain. *Neuropsychologia*, 42(1):95--104, 2004.
- H. P. Op de Beeck, J. Wagemans, and R. Vogels. Effects of perceptual learning in visual backward masking on the responses of macaque inferior temporal neurons. *Neuroscience*, 145(2):775--789, Mar 2007.
- L. M. Optican and B. J. Richmond. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. iii. information theoretic analysis. *Journal of Neurophysiology*, 57(1):162--78, 1987.
- R. C. O'Reilly, D. Wyatte, S. Herd, B. Mingus, and D. J. Jilk. Recurrent processing during object recognition. *Front Psychol*, 4:124, 2013.
- D. Osipova, A. Takashima, R. Oostenveld, G. Fernandez, E. Maris, and O. Jensen. Theta and gamma oscillations predict encoding and retrieval of declarative memory. *J. Neurosci.*, 26(28):7523--7531, Jul 2006.
- Josef Parvizi, Corentin Jacques, Brett L. Foster, Nathan Witthoft, Nathan Withoft, Vinitha Rangarajan, Kevin S. Weiner, and Kalanit Grill-Spector. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J Neurosci*, 32(43):14915--20, 10 2012.
- A. Pasupathy and C. E. Connor. Population coding of shape in area v4. *Nat Neurosci*, 5(12):1332--8, 2002.

- W. Penfield and H.H. Jasper. *Epilepsy and the Functional Anatomy of the Human Brain*. Little, Brown, 1954.
- Bojan Pepik, Rodrigo Benenson, Tobias Ritschel, and Bernt Schiele. What is holding back convnets for detection? *arXiv preprint arXiv:1508.02844*, 2015.
- E. Perrett. The left frontal lobe of man and the suppression of habitual responses in verbal categorical behavior. *Neuropsychologia*, 12(3):323--330, 1974.
- E. Peterhans and R. von der Heydt. Subjective contours - bridging the gap between psychophysics and physiology. *Trends in Neuroscience*, 14:112--119, 1991.
- MC Potter and E Levy. Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 81(1):10--15, 1969.
- R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput*, 16(8):1661--1687, Aug 2004.
- R. P. Rao and D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*, 2(1):79--87, 1999.
- Robert Rauschenberger, Mary A Peterson, Fauzia Mosca, and Nicola Bruno. Amodal completion in visual search: preemption or context effects? *Psychological science*, 15(5):351--355, 2004.
- S. Ray and J. H. Maunsell. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol*, 9(4):e1000610, 2011.
- D. S. Reich, F. Mechler, and J. D. Victor. Temporal coding of contrast in primary visual cortex: when, what, and why. *J Neurophysiol*, 85(3):1039--50, 2001.
- M. W. Reimann, C. A. Anastassiou, R. Perin, S. L. Hill, H. Markram, and C. Koch. A biophysically detailed model of neocortical local field potentials predicts the critical role of active membrane currents. *Neuron*, 79(2):375--390, Jul 2013.
- J. H. Reynolds and L. Chelazzi. Attentional modulation of visual processing. *Annu Rev Neurosci*, 27:611--47, 2004.
- BJ Richmond, RH Wurtz, and T Sato. Visual responses in inferior temporal neurons in awake rhesus monkey. *Journal of Neurophysiology*, 50(6):1415--1432, 1983.

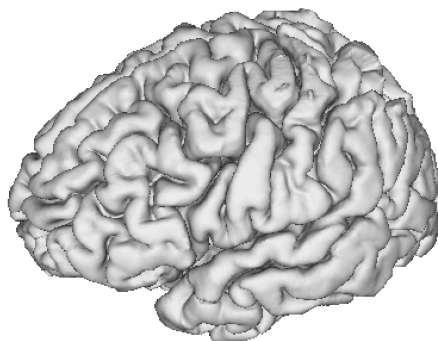
- K Richard Ridderinkhof, Markus Ullsperger, Eveline A Crone, and Sander Nieuwenhuis. The role of the medial frontal cortex in cognitive control. *Science*, 306(5695):443--7, 2004.
- M Riesenhuber and T Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019--1025, 1999.
- E. T. Rolls, M. J. Tovee, and S. Panzeri. The neurophysiology of backward visual masking: information analysis. *J Cogn Neurosci*, 11(3):300--311, May 1999.
- Edmund T Rolls. An attractor network in the hippocampus: theory and neurophysiology. *Learning & Memory*, 14(11):714--731, 2007.
- ET Rolls. Neural organization of higher visual functions. *Current Opinion in Neurobiology*, 1:274--278, 1991.
- David C Rubin and Amy E Wenzel. One hundred years of forgetting: a quantitative description of retention. *Psychological review*, 103(4):734, 1996.
- M. F. Rushworth, M. E. Walton, S. W. Kennerley, and D. M. Bannerman. Action sets and decisions in the medial frontal cortex. *Trends Cogn Sci*, 8(9):410--7, 2004.
- Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015.
- U. Rutishauser, I. B. Ross, A. N. Mamelak, and E. M. Schuman. Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature*, 464(7290):903--907, Apr 2010.
- MT Schmolesky, YC Wang, DP Hanes, KG Thompson, S Leutgeb, JD Schall, and AG Leventhal. Signal timing across the macaque visual system. *Journal of Neurophysiology*, 79(6):3272--3278, 1998.
- Philippe G Schyns, Lucy S Petro, and Marie L Smith. Dynamics of visual information integration in the brain for categorizing facial expressions. *Current Biology*, 17(18):1580--1585, 2007.
- P. B. Sederberg, M. J. Kahana, M. W. Howard, E. J. Donner, and J. R. Madsen. Theta and gamma oscillations during encoding predict subsequent recall. *J. Neurosci.*, 23(34):10809--10814, Nov 2003.
- P. Sehatpour, S. Molholm, T. H. Schwartz, J. R. Mahoney, A. D. Mehta, D. C. Javitt, P. K. Stanton, and J. J. Foxe. A human intracranial study of long-range oscillatory coherence across a frontal-occipital-hippocampal brain network during visual object processing. *Proc Natl Acad Sci U S A*, 105(11):4399--404, 2008.

- A. B. Sekuler and S.E. Palmer. Perception of partly occluded objects: a microgenetic analysis. *Journal of Experimental Psychology: General*, 121(1):95--111, 1992.
- Allison B Sekuler and Richard F Murray. Amodal completion: A case study in grouping. *Advances in Psychology*, 130:265--293, 2001.
- Allison B Sekuler, Stephen E Palmer, and Carol Flynn. Local and global processes in visual completion. *Psychological science*, 5(5):260--267, 1994.
- T Serre, G Kreiman, M Kouh, C Cadieu, U Knoblich, and T Poggio. A quantitative theory of immediate visual recognition. *Progress In Brain Research*, 165C:33--56, 2007a.
- T Serre, A Oliva, and T Poggio. Feedforward theories of visual cortex account for human performance in rapid categorization. *PNAS*, 104(15):6424--6429, 2007b.
- H Sebastian Seung. Learning continuous attractors in recurrent networks. In *NIPS*, volume 97, pages 654--660. Citeseer, 1997.
- R. M. Shapley and J. D. Victor. The effect of contrast on the transfer properties of cat retinal ganglion cells. *J Physiol*, 285:275--98, 1978.
- Amitai Shenhav, Matthew M Botvinick, and Jonathan D Cohen. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, 79(2):217--40, 2013.
- Gordon M Shepherd. *The synaptic organization of the brain*. 1998.
- Sameer A Sheth, Matthew K Mian, Shaun R Patel, Wael F Asaad, Ziv M Williams, Darin D Dougherty, George Bush, and Emad N Eskandar. Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature*, 488(7410):218--21, 2012.
- S. Shimojo and K. Nakayama. Amodal representation of occluded surfaces: role of invisible stimuli in apparent motion correspondence. *Perception*, 19(3):285--99, 1990a.
- S. Shimojo and K. Nakayama. Real world occlusion constraints and binocular rivalry. *Vision Research*, 30(1):69--80, 1990b.
- D I Shore and J T Enns. Shape completion time depends on the size of the occluded region. *J Exp Psychol Hum Percept Perform*, 23(4):980--98, 1997.
- J. H. Siegle and M. A. Wilson. Enhancement of encoding and retrieval functions through theta phase-specific manipulation of hippocampus. *Elife*, 3:e03061, 2014.

- K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- Jedediah M Singer, Joseph R Madsen, William S Anderson, and Gabriel Kreiman. Sensitivity to timing and order in human visual cortex. *J Neurophysiol*, 113(5):1656--69, 2015.
- M. Singh. Modal and amodal completion generate different shapes. *Psychol Sci*, 15(7):454--9, 2004.
- John R Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18:643--662, 1935.
- Y. Sugita. Grouping of image fragments in primary visual cortex. *Nature*, 401(6750):269--72, 1999.
- Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deeply learned face representations are sparse, selective, and robust. *arXiv preprint arXiv:1412.1265*, 2014.
- Nanthia Suthana, Zulfi Haneef, John Stern, Roy Mukamel, Eric Behnke, Barbara Knowlton, and Itzhak Fried. Memory enhancement and deep-brain stimulation of the entorhinal area. *New England Journal of Medicine*, 366(6):502--510, 2012.
- D. Swick and A. U. Turken. Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex. *Proc Natl Acad Sci U S A*, 99(25):16354--9, 2002.
- Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lars Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1701--1708. IEEE.
- K. Tanaka. Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19:109--139, 1996.
- Hanlin Tang, Calin Buia, Radhika Madhavan, Nathan E Crone, Joseph R Madsen, William S Anderson, and Gabriel Kreiman. Spatiotemporal dynamics underlying object completion in human ventral visual cortex. *Neuron*, 83(3):736--48, 2014.
- Yichuan Tang, Ruslan Salakhutdinov, and Geoffrey Hinton. Robust boltzmann machines for recognition and denoising. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2264--2271. IEEE, 2012.

- S. F. Taylor, B. Martis, K. D. Fitzgerald, R. C. Welsh, J. L. Abelson, I. Liberzon, J. A. Himle, and W. J. Gehring. Medial frontal cortex activity and loss-related responses to errors. *J Neurosci*, 26(15):4063--70, 2006.
- S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520--522, 1996.
- A. B. Tort, M. A. Kramer, C. Thorn, D. J. Gibson, Y. Kubota, A. M. Graybiel, and N. J. Kopell. Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a t-maze task. *Proc Natl Acad Sci U S A*, 105(51):20517--22, 2008.
- Shimon Ullman. Filling-in the gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics*, 25(1):1--6, 1976.
- M. Ullsperger, C. Danielmeier, and G. Jocham. Neurophysiology of performance monitoring and adaptive behavior. *Physiol Rev*, 94(1):35--79, 2014.
- Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.
- J. van Driel, J. C. Swart, T. Egner, K. R. Ridderinkhof, and M. X. Cohen. (no) time for control: Frontal theta dynamics reveal the cost of temporally guided conflict anticipation. *Cogn Affect Behav Neurosci*, 2015.
- J. R. Vidal, T. Ossandon, K. Jerbi, S. Dalal, L. Minotti, P. Ryvlin, P. Kahan, and J. P. Lachaux. Category-specific visual responses: an intracranial study comparing gamma, beta, alpha, and erp response selectivity. *Frontiers in Human Neuroscience*, 4:1--23, 2010.
- Stine Vogt and Svein Magnussen. Long-term memory for 400 pictures on a common theme. *Experimental psychology*, 54(4):298--303, 2007.
- R von der Heydt, E Peterhans, and G Baumgartner. Illusory contours and cortical neuron responses. *Science*, 224:1260--1262, 1984.
- A. von Stein and J. Sarnthein. Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *Int J Psychophysiol*, 38(3):301--13, 2000.
- G. Wallis and E. T. Rolls. Invariant face and object recognition in the visual system. *Progress in Neurobiology*, 51(2):167--94, 1997.

- N. Weiskopf, C. Hutton, O. Josephs, and R. Deichmann. Optimal epi parameters for reduction of susceptibility-induced bold sensitivity losses: a whole-brain analysis at 3 t and 1.5 t. *Neuroimage*, 33(2):493--504, 2006.
- Paul J Werbos. Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4):339--356, 1988.
- J. M. Williams and B. Givens. Stimulation-induced reset of hippocampal theta in the freely performing rat. *Hippocampus*, 13(1):109--16, 2003.
- D. Wyatte, T. Curran, and R. O'Reilly. The limits of feedforward vision: recurrent processing promotes robust object recognition when objects are degraded. *J Cogn Neurosci*, 24(11):2248--61, 2012.
- Daniel L K Yamins, Ha Hong, Charles F Cadieu, Ethan A Solomon, Darren Seibert, and James J DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci U S A*, 111(23):8619--24, 2014.
- N. Yeung, M. M. Botvinick, and J. D. Cohen. The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol Rev*, 111(4):931--59, 2004.
- A. Yuille and D. Kersten. Vision as bayesian inference: analysis by synthesis? *Trends Cogn Sci*, 10(7):301--8, 2006.
- Zihan Zhou, Andrew Wagner, Hossein Mobahi, John Wright, and Yi Ma. Face recognition with contiguous occlusion using markov random fields. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1050--1057. IEEE, 2009.
- K. Zipser, V.A. Lamme, and P.H. Schiller. Contextual modulation in primary visual cortex. *Journal of neuroscience*, 16(22):7376--7389, 1996.



THIS THESIS WAS TYPESET using \LaTeX , originally developed by Leslie Lamport and based on Donald Knuth's \TeX . The body text is set in Minion Pro, developed by Robert Slimbach, with mathematical content set in MnSymbol. A customized version of the Dissertate template, found online at github.com/asm-products/Dissertate, was used to format the look & feel of this dissertation.