2013

Neural models of inter-cortical networks in the primate visual system for navigation, attention, path perception, and static and kinetic figure-ground perception

https://hdl.handle.net/2144/15224

Boston University
BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

NEURAL MODELS OF INTER-CORTICAL NETWORKS
IN THE PRIMATE VISUAL SYSTEM FOR
NAVIGATION, ATTENTION, PATH PERCEPTION, AND
STATIC AND KINETIC FIGURE-GROUND
PERCEPTION

by

OLIVER W. LAYTON

BA, Skidmore College, 2009

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2013
Approved by

First Reader
Ennio Mingolla, PhD
Adjunct Professor of Psychology

Second Reader
N. Andrew Browning, PhD
Adjunct Research Assistant Professor

Third Reader
Arash Yazdanbakhsh, MD, PhD
Research Assistant Professor
The rules that govern behavior are not like laws enforced by an authority or decisions made by a commander; behavior is regular without being regulated.

James J. Gibson
Acknowledgments

I would like to foremost thank Sarah Whateley whom I love with all my heart. Her energy and love drives me to do great work. Words cannot describe my appreciation for her immeasurable patience and support during the past four years. I would also like to thank Brad, Liz, and everyone in Sarah’s family for their support and making me feel like family.

Next, I would like to thank my advisors Arash Yazdanbakhsh, Andrew Browning, and Ennio Mingolla. Arash always went above and beyond the call of duty as a scientist and friend. I will always treasure our 3+ hour meetings at the coffee house working out with pencil and paper the next creative breakthrough. He was always there for me, no matter what time of night I had eureka moments and needed to talk. I am incredibly grateful to Andrew for teaching me how to transform vague ideas into scientific projects with specific objectives and goals. Without Andrew, I probably would not have studied or entered the field of visually-guided navigation. His brutal honesty taught me how to refine and be intensely critical of my own work. I thank Ennio for the intellectual freedom to pursue my interests, no matter how much they wandered and evolved over time. I remember our first meeting, when I presented on my vague interest in multi-modal receptive fields. His guidance was essential in differentiating between good ideas and those without direction. His scientific scrutiny and relentless counter-examples made my work so much better, and the endless rounds of copy editing taught me how to construct good scientific manuscripts.

I would like to thank Flip Phillips for advising and mentoring me as an undergraduate at Skidmore College. Without him, I would have never heard of computational modeling in neuroscience/psychology or vision science. I am forever indebted to him for grounding me in ecological psychology, which has proven incredibly useful in evaluating brain theories and models, not falling prey in my work to homunculi, and
understanding that perception is an active dynamical system. I can never forget the summer day in the lab when the shipment arrived with Gibson’s book inside—I was instantly hooked and read the whole thing in a single sitting.

I would like to thank Tim Barnes, Cloud Tsao, Florian Raudies, Praveen Pilly, Jasmin Leveille, Gennady Livitz, and everyone else in the CNS Vision Lab who mentored and inspired me.

I thank Bill Stitson and Brad Nesbitt for being my friends no matter how stressed I got or little time I had.
Vision provides the primary means by which many animals distinguish foreground objects from their background and coordinate locomotion through complex environments. The present thesis focuses on mechanisms within the visual system that afford figure-ground segregation and self-motion perception. These processes are modeled as emergent outcomes of dynamical interactions among neural populations in several brain areas. This dissertation specifies and simulates how border-ownership signals emerge in cortex, and how the medial superior temporal area (MSTd) represents path of travel and heading, in the presence of independently moving objects (IMOs).

Neurons in visual cortex that signal border-ownership, the perception that a border belongs to a figure and not its background, have been identified but the under-
lying mechanisms have been unclear. A model is presented that demonstrates that inter-areal interactions across model visual areas V1–V2–V4 afford border-ownership signals similar to those reported in electrophysiology for visual displays containing figures defined by luminance contrast. Competition between model neurons with different receptive field sizes is crucial for reconciling the occlusion of one object by another. The model is extended to determine border-ownership when object borders are kinetically-defined, and to detect the location and size of shapes, despite the curvature of their boundary contours.

Navigation in the real world requires humans to travel along curved paths. Many perceptual models have been proposed that focus on heading, which specifies the direction of travel along straight paths, but not on path curvature. In primates, MSTd has been implicated in heading perception. A model of V1, medial temporal area (MT), and MSTd is developed herein that demonstrates how MSTd neurons can simultaneously encode path curvature and heading. Human judgments of heading are accurate in rigid environments, but are biased in the presence of IMOs. The model presented here explains the bias through recurrent connectivity in MSTd and avoids the use of differential motion detectors which, although used in existing models to discount the motion of an IMO relative to its background, is not biologically plausible. Reported modulation of the MSTd population due to attention is explained through competitive dynamics between subpopulations responding to bottom-up and top-down signals.
## Contents

1 Introduction ......................................................... 1

1.1 Figure-ground segregation ................................. 3

1.1.1 Visual borders ......................................... 8

1.1.2 Perceptual grouping ................................. 11

1.1.3 Local and global interactions in figure-ground segregation ... 13

1.2 Visually-guided navigation ................................. 15

1.2.1 Rotation in optic flow ................................ 21

1.2.2 Physiology of navigation ................................. 23

1.3 Outline of the thesis ......................................... 26

2 Border-ownership: A vector code for figure-ground segregation 31

2.1 Border-ownership ........................................... 31

2.2 The RGB model of border-ownership ......................... 35

2.3 Contributions .............................................. 36

3 Dynamic coding of border-ownership in visual cortex 37

4 Border-ownership is a general-purpose signal for figure-ground segregation 84

4.1 Functional segregation in the visual system ..................... 84

4.2 Model of kinetic border-ownership ............................. 86

5 The border-ownership of kinetic edges: a neural model 88

5.1 Introduction ..................................................... 88

5.1.1 Kinetic random dot displays .............................. 92
5.2 Methods ......................................................... 100
  5.2.1 Magnocellular and parvocellular pathways ................. 100
  5.2.2 Spatio-temporal correlation detection ...................... 104
  5.2.3 Border-ownership cells .................................. 107
  5.2.4 Magnocellular and parvocellular interactions: feedback from V4
           and MT to V2 ........................................... 108
  5.2.5 Visual Displays .......................................... 113
  5.2.6 Model Equations ......................................... 115
5.3 Results ....................................................... 128
  5.3.1 Stationary condition .................................... 128
  5.3.2 Encroach condition ...................................... 131
  5.3.3 Shear condition ......................................... 131
  5.3.4 Strip condition ......................................... 134
  5.3.5 Object & Window conditions .............................. 135
  5.3.6 Gap condition .......................................... 138
  5.3.7 Temporal Dynamics ...................................... 139
5.4 Discussion ................................................... 141
  5.4.1 The role of higher visual areas in figure-ground segregation .... 142
  5.4.2 Border-ownership cells and kinetic edges ................... 143
  5.4.3 Motion mechanisms ...................................... 145
  5.4.4 The representation of relative depth ....................... 146
5.5 Conclusion .................................................... 148

6 Shape detection ................................................. 150

7 Dynamics of multi-scale shape integration and detection in the pri-
    mate ventral stream ........................................... 155
  7.1 Introduction ................................................ 155
8 Independently moving objects and heading perception 213
  8.1 Independently moving objects .............................. 213
  8.2 Detection of independently moving objects ................. 213
  8.3 Observer- versus world-centered coordinates .............. 216
  8.4 Model of heading perception in the presence of independently moving objects .............................. 220
  8.5 Contributions ............................................. 222

9 A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects 223
  9.1 Introduction ................................................ 223
    9.1.1 Theories of heading estimation ....................... 224
    9.1.2 Neurophysiological background ....................... 227
    9.1.3 Human heading estimation models .................... 229
    9.1.4 Heading estimation in the presence of an approaching moving object ....................................... 234
    9.1.5 Heading estimation in the presence of a moving object maintaining a fixed distance ................... 237
  9.2 The Model .................................................. 239
  9.3 Results ..................................................... 251
  9.4 Discussion ................................................ 253
    9.4.1 Approaching IMO ..................................... 254
    9.4.2 Non-approaching IMO .................................. 256
    9.4.3 Optic flow illusion .................................... 258
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.4.4 Timing</td>
<td>259</td>
</tr>
<tr>
<td>9.4.5 MSTd cell types</td>
<td>260</td>
</tr>
<tr>
<td><strong>10 Heading perception and attention</strong></td>
<td>261</td>
</tr>
<tr>
<td>10.1 Attentional signals in MSTd</td>
<td>261</td>
</tr>
<tr>
<td>10.2 Contributions</td>
<td>262</td>
</tr>
<tr>
<td><strong>11 Recurrent Competition Explains Temporal Effects of Attention in MSTd</strong></td>
<td>264</td>
</tr>
<tr>
<td>11.1 Introduction</td>
<td>264</td>
</tr>
<tr>
<td>11.2 Materials and Methods</td>
<td>269</td>
</tr>
<tr>
<td>11.2.1 The Model</td>
<td>274</td>
</tr>
<tr>
<td>11.3 Results</td>
<td>280</td>
</tr>
<tr>
<td>11.4 Discussion</td>
<td>287</td>
</tr>
<tr>
<td><strong>12 A neural theory of path perception</strong></td>
<td>297</td>
</tr>
<tr>
<td><strong>13 A unified model of heading and path perception in primate MSTd</strong></td>
<td>301</td>
</tr>
<tr>
<td>13.1 Introduction</td>
<td>301</td>
</tr>
<tr>
<td>13.1.1 Theories of Path Perception</td>
<td>304</td>
</tr>
<tr>
<td>13.1.2 Neurophysiology of Path Perception</td>
<td>307</td>
</tr>
<tr>
<td>13.1.3 Spiral-selective MSTd cells dynamically encode path and heading direction</td>
<td>311</td>
</tr>
<tr>
<td>13.2 Materials and Methods</td>
<td>318</td>
</tr>
<tr>
<td>13.2.1 Model Area Descriptions</td>
<td>318</td>
</tr>
<tr>
<td>13.2.2 Experimental Descriptions</td>
<td>323</td>
</tr>
<tr>
<td>13.3 Results</td>
<td>326</td>
</tr>
<tr>
<td>13.3.1 Path Perception &amp; Gaze</td>
<td>326</td>
</tr>
<tr>
<td>13.3.2 Different dot densities</td>
<td>331</td>
</tr>
<tr>
<td>13.3.3 Heading</td>
<td>333</td>
</tr>
</tbody>
</table>
List of Tables

5.1 Summary of feedforward and feedback connections in kinetic border-ownership model ........................................... 113

7.1 Parameter settings used in shape detection model ............... 212

11.1 Parameter settings used in the model of heading estimation with attention .......................................................... 274

13.1 Parameter settings used in the model of path perception ........ 347

15.1 Parameter settings used in analysis of heading and path model ..... 362
# List of Figures

1-1  Figure-ground displays .................................................. 5

3-1  Bistable border-ownership examples ................................ 47

3-2  Visual displays simulated .................................................. 47

3-3  Schematic model diagram ..................................................... 50

3-4  G and B cells ................................................................. 52

3-5  R cells ................................................................. 54

3-6  Simulation results: T-junction simulation ................................. 64

3-7  Simulation results: L- & X-junctions .................................... 68

3-8  Simulation results: C-shape .................................................. 70

3-9  Simulation results: Convexity ................................................. 72

3-10  Simulation results: B cell firing rate ..................................... 73

3-11  Simulation results: Kanizsa Square ..................................... 74

5-1  Camouflaged mossy frog (*Theodorina corticale*). ....................... 89

5-2  Kinetic random dot displays & experimental conditions. ............... 94

5-3  Kinetic border-ownership model schematic diagram .................... 101

5-4  Spatio-temporal correlation detection circuit ........................... 105

5-5  Feedforward and feedback connectivity between model V2 and V4 .. 110

5-6  Model V4 Receptive Fields ................................................. 124

5-7  Simulation results: Stationary, Encroach, shear, and degenerate conditions. ......................................................... 129

5-8  Simulation results: Strip condition ........................................ 133
5.9 Simulation results: Object and window conditions
5.10 Simulation results: Gap condition
5.11 Simulation results: Temporal dynamics

7.1 Model Diagram
7.2 Curved contour units
7.3 Single-scale units
7.4 Multi-scale units
7.5 Multi-scale radial feedback
7.6 Single-scale feedback
7.7 Visual displays simulated
7.8 Simulation results: Exemplars
7.9 Simulation results: Natural scan and thin bar
7.10 Single-scale unit subunit orientation permutation test
7.11 Feedback and shape selectivity analysis
7.12 Feedback and figure-ground percent correct classification
7.13 Feedback and single-scale unit response gain
7.14 Proposed experiment to test model and neural circuit implementation

9.1 Approaching and non-approaching IMOs
9.2 MT+/MSTd and MT−/MSTv motion pathways
9.3 Visual displays of Warren & Saunders
9.4 Visual displays of Royden & Hildreth
9.5 Model diagram
9.6 Experimental setup
9.7 MST heading templates
9.8 Simulation results: MT representation
9.9 Simulation results: MSTd activation

xvi
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-10</td>
<td>Simulation results: Non-approaching IMO human heading bias</td>
<td>250</td>
</tr>
<tr>
<td>9-11</td>
<td>Simulation results: Approaching IMO human heading bias</td>
<td>252</td>
</tr>
<tr>
<td>9-12</td>
<td>Simulation results: Receding IMO</td>
<td>255</td>
</tr>
<tr>
<td>11-1</td>
<td>Mean MSTd population firing rates in experimental conditions</td>
<td>267</td>
</tr>
<tr>
<td>11-2</td>
<td>Sample optic flow field used in the model simulations</td>
<td>271</td>
</tr>
<tr>
<td>11-3</td>
<td>Simulation results: MSTd population activation</td>
<td>278</td>
</tr>
<tr>
<td>11-4</td>
<td>Simulation results: Sigmoid signal function</td>
<td>282</td>
</tr>
<tr>
<td>11-5</td>
<td>Simulation results: Spatial dynamics in MSTd</td>
<td>285</td>
</tr>
<tr>
<td>11-6</td>
<td>Simulation results: Attention tuning curves</td>
<td>288</td>
</tr>
<tr>
<td>13-1</td>
<td>Exemplar first-order optic flow fields</td>
<td>305</td>
</tr>
<tr>
<td>13-2</td>
<td>Experimental paradigm of Froehler and Duffy</td>
<td>309</td>
</tr>
<tr>
<td>13-3</td>
<td>Model and electrophysiological spiral space</td>
<td>311</td>
</tr>
<tr>
<td>13-4</td>
<td>Schematic depiction of the model coding of path curvature and heading in MSTd</td>
<td>314</td>
</tr>
<tr>
<td>13-5</td>
<td>Diagram of model V1-MT-MSTd</td>
<td>317</td>
</tr>
<tr>
<td>13-6</td>
<td>Gaze conditions during observer curvilinear locomotion</td>
<td>325</td>
</tr>
<tr>
<td>13-7</td>
<td>Simulation Results: Path errors in the five gaze conditions</td>
<td>328</td>
</tr>
<tr>
<td>13-8</td>
<td>Simulation results: Model robustness to dot density</td>
<td>332</td>
</tr>
<tr>
<td>13-9</td>
<td>Simulation results: Model heading bias</td>
<td>334</td>
</tr>
<tr>
<td>13-10</td>
<td>Simulation results: Path errors with eye movements</td>
<td>335</td>
</tr>
<tr>
<td>13-11</td>
<td>Simulation Results: Model lesions</td>
<td>338</td>
</tr>
<tr>
<td>13-12</td>
<td>Simulation results: Heading bias with simulated rotation</td>
<td>338</td>
</tr>
<tr>
<td>13-13</td>
<td>Simulation results: Experiment of Froehler and Duffy</td>
<td>341</td>
</tr>
<tr>
<td>15-1</td>
<td>Model MSTd coding schematic</td>
<td>352</td>
</tr>
<tr>
<td>15-2</td>
<td>Model diagram</td>
<td>353</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>15.3 Summary of simulated experimental conditions</td>
<td>354</td>
<td></td>
</tr>
<tr>
<td>15.4 Geometry of observer gaze conditions</td>
<td>358</td>
<td></td>
</tr>
<tr>
<td>15.5 Simulation results: Template match with and without competition</td>
<td>365</td>
<td></td>
</tr>
<tr>
<td>15.6 Simulation results: Activation in MSTd spiral space</td>
<td>366</td>
<td></td>
</tr>
<tr>
<td>15.7 Simulation results: Heading bias over time</td>
<td>369</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------</td>
<td></td>
</tr>
<tr>
<td>FEF</td>
<td>Frontal eye fields</td>
<td></td>
</tr>
<tr>
<td>FoC</td>
<td>Focus of contraction</td>
<td></td>
</tr>
<tr>
<td>FoE</td>
<td>Focus of expansion</td>
<td></td>
</tr>
<tr>
<td>IMO</td>
<td>Independently moving object</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>Inferotemporal cortex</td>
<td></td>
</tr>
<tr>
<td>LGN</td>
<td>Lateral geniculate nucleus</td>
<td></td>
</tr>
<tr>
<td>MDS</td>
<td>Multi-dimensional scaling</td>
<td></td>
</tr>
<tr>
<td>MSTd</td>
<td>Dorsal medial superior temporal area</td>
<td></td>
</tr>
<tr>
<td>MSTv</td>
<td>Ventral medial superior temporal area</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>Medial temporal area</td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>Receptive field</td>
<td></td>
</tr>
<tr>
<td>RGB Model</td>
<td>R-, G-, B-cell model</td>
<td></td>
</tr>
<tr>
<td>TTC</td>
<td>Time to contact</td>
<td></td>
</tr>
<tr>
<td>VIP</td>
<td>Ventral Intraparietal Area</td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>Primary visual cortex</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Detecting the presence of predators and coordinating retreat or avoidance maneuvers are crucial to the survival of many animals. Prey that use vision to search for predators face the challenge that predators may be identifiable in a variety of ways, including differences in luminance contrast and a breaking in camouflage (motion) from their surroundings. For survival, the visual system must therefore integrate multiple types of visual information to identify the predator, despite the varied conditions under which it may appear. A predator in pursuit is an independently moving object (IMO), and the prey’s visual system must be capable of guiding locomotion through complex environments to escape. Because the environment often contains obstacles, the animal, with limited attentional resources, must be able to negotiate travel along curved paths. Guiding locomotion about the environment and distinguishing foreground objects from their background (e.g. identifying the presence of a predator) therefore play crucial roles in survival and depend on visual perception.

Humans encounter, perceive, and act on a diverse array of objects within a threedimensional world. The human visual system must parse and make sense of a large number of object surface properties, such as texture, color, and surface transparency, in the absence of physical contact with the objects. Vision also guides human locomotion through the perception of self-motion through the environment, circumventing obstacles, and approaching or avoiding IMOs. The information conferred by vision allows humans to perform actions that are critical to survival and normal functioning
in everyday life, such as steering, traveling along curved paths, chasing or retreating from IMOs, and pursuing stationary or moving targets.

Visual perception occurs in animals with eyes through their interaction with light, which specifies the intricately structured layout of the world. Gibson termed the light available to animals at any point of observation the *ambient optic array* (Gibson, 1968; Gibson, 1979). Light that is focused on the mammalian retina’s two-dimensional surface contains rich optical structure about the world. We call this 2D projection the retinal image. The light that reaches the eye has reflected off many surfaces in the world and specifies salient information about the form of objects in the visual scene. Matte (lambertian) and specular surfaces are perceived on the basis of the diffuse and asymmetric patterns of light, respectively, that reflect off the object. The ambient optic array remains invariant at any one point of observation over changes in ambient illumination because the visual angle between the observer and objects in the world are fixed. The array differs at different points of observation (Gibson, 1961). When an animal moves through space, the point of observation changes and transformations in the ambient optic array occur over time that are specific to the pattern of locomotion and environmental structure. Gibson called this cascade of transformations of the ambient optic array *optic flow* (Gibson, 1979). Whether stationary or in motion, light stimulates photoreceptors on the retinal surface, which transmits signals to visual cortex.

Over the past half century, neurophysiological studies have identified two main pathways within the primate visual system: dorsal and ventral (Hubel and Wiesel, 1962; Hubel, 1982). Individual areas within the dorsal stream have been shown to contain neurons sensitive to patterns of motion, such as those produced when an animal navigates in the environment. Conversely, the ventral stream contains neurons selective for visual figures and their backgrounds. Neurophysiology and computa-
tional modeling of the visual system have largely been concerned with the role of neurons in single cortical areas have on visual perception. By contrast, the models developed herein make progress toward understanding how neuronal populations spanning multiple visual areas across the dorsal and ventral streams organize to afford figure-ground segregation and navigation.

In what follows, we review figure-ground segregation and visually-guided navigation. The chapter concludes with goals that are addressed in the thesis as well as an outline.

1.1 Figure-ground segregation

Figure-ground segregation refers to the process by which the visual system parses the complex array of luminance that appears on the retina into perceptually grouped foreground objects (figures) and backgrounds (ground). At minimum, adjacent regions in the visual field must share a border for figure-ground segregation to occur. As discussed below, the constituent properties of the border between the regions may be rather complex and need not be defined by sharp luminance transitions. When one of the regions appears closer in depth than the other, the region is called the *figure*, and the other the *ground*. The figure is said to *occlude* the background. The abutting border is grouped with and seen as “belonging to” the figure rather than the ground. The ground is generally seen to continue underneath the occluding border between the two regions. Despite the apparent automaticity with which the visual system parcellates the visual scene into coherent, holistic objects, figure-ground segregation is an incredibly challenging problem.

The perceptual organization of a visual scene is highly context-dependent, and the configuration of elements have been shown to impact how the visual system performs figure-ground segregation. Certain types of visual information influence
Figure 1.1: (a) Convexity display of Kanizsa and Gerbino. Human subjects are more likely to judge the convex regions as figure than the concave regions. (b) Bertamini showed that subject reaction times and errors decreased when a triangular vertex was part of the convex region (left) compared to when it was part of the concave part (right), demonstrating a convexity advantage. (c) Symmetry display used by Driver and colleagues and Peterson. Subjects were more likely to judge the symmetric regions as figure than the asymmetric regions. (d–h) Hole displays used by Nelson and Palmer. A gray region was ordered by an outer white region and an inner white region. (d) A shadow placed on the side of the gray region biases subjects to judge the interior region as a figure. (e) A shadow placed on the side of the inner region biases subjects to judge the interior region as a hole. (f) Continuity of the black region from the inner to outer region of biases subjects to judge the interior region as a hole, but discontinuity of the black does not (g). (h) When the inner texture matches that of the outer region, subjects judge the inner region as a hole. (i–j) Entropy contrast displays of Gilliam and Gove. (i) and (j) show low and high entropy line terminators, respectively. The greater the entropy, the stronger the occlusion percept. (k–l) An extremal edge (boxed) in a natural scene (k), a configural cue discovered by Palmer and Ghose. The shadow projected from the partial self-occlusion of one zucchini gives the impression of occlusion (l). (m–n) Vecera and colleagues found that in horizontally-divided bipartite displays, such as the shown wave (m) and Kanizsa (n) displays, subjects demonstrate a reliable bias in judging the bottom region as figure.
to which side of an edge a shape is perceived and are called ‘configural cues’ by Gestalt psychologists. These properties include convexity and symmetry. Peterson and Salvagio presented human subjects with the classical visual displayed first used by Kanizsa and Gerbino, and demonstrated that all other factors remaining equal, convex regions are far more likely to be classified as figure than concave regions, even when the viewing duration was only 100 msec (Kanizsa and Gerbino, 1976; Kanizsa, 1979; Peterson and Salvagio, 2008). Convex regions were more likely to be perceived as figure even when a single convex-concave region pair was presented (Figure 1·1a). The original displays used by Kanizsa and Gerbino are composed of periodic black-white shaded regions with equal areas that are separated by contours that undulated in the sign of curvature. The contours curved in such a way to make one adjacent region have convex protrusions and consequently the other region appears to have concave indentations. Interestingly, the convex figure bias persisted despite whether the concave regions were filled with homogenous or inhomogenous color or textures. When subjects are asked to judge the position of the vertex of a triangular extension from the border between two adjacent regions that differ in luminance, reaction times and errors were lower when the vertex was part of the convex region, indicating a convex advantage (Bertamini, 2001). Hence, convexity represents a powerful piece of visual information that modulates how the visual system performs figure-ground segregation (Figure 1·1b).

Symmetry has also been shown to represent an important property that influences figure-ground segregation. In displays that contain between two and eight panels alternating in symmetry, the symmetric regions are more likely judged as figures by human subjects (Driver et al., 1992). Even with brief exposure times, human were more likely to judge the symmetric regions as figure than the asymmetric regions on ~75% of trials (Peterson, 1994). When a convex figure was challenged by competing
global symmetry, the bias to classify the convex region figure persisted (Kanizsa and Gerbino, 1976). Therefore, the visual system may weight convexity higher than symmetry in the presence of both sources of information (Figure 1.1c).

Recent studies have identified other sources of visual information that affects figure-ground percepts. Enclosed regions may appear ambiguous with respect to figure-ground—the enclosed region may appear as the figure, ground, or a hole “punched” through the surrounding surface (Bertamini and Hulleman, 2006; Bertamini, 2006; Bertamini et al., 2008). Nelson and Palmer conducted psychophysical experiments in which a gray square that contained an enclosed region defined by thin curved lines had the same luminance as the background. When a shadow flanked the curved interior region on the side of the square, subjects indicated that the region appeared as a figure (Figure 1.1d). When the shadow appeared on the inside of the enclosed region, subjects reported that the region looked like a hole (Figure 1.1e). Subject percepts of the hole were modulated in an augmented version of the display by the occlusion and continuity of a dark square (Figure 1.1f–g). Continuity of the dark region inside the enclosed region and outside the gray square was essential for subjects to indicate the presence of a hole (Nelson and Palmer, 2001). Common motion, textural similarity (Figure 1.1h), and synchronous flashing with the region surrounding the square also resulted in the curved central region to be perceived as a hole. Gillam and Grove further developed conditions under which a region may be more robustly perceived as ground (Figure 1.1i–j). The authors found that occlusion percepts are enhanced as an increasing function of the irregularity of the line terminations (“entropy contrast”) that define the occluded region (Gillam and Grove, 2011).

Palmer and colleagues recently discovered ‘extremal edges’ that occur in images of natural scenes, and provide information that biases the region perception to one side of the edge as figure (Figure 1.1k–l). Extremal edges occur due to the view-
point projection of objects by the camera onto the image plane. The projection of an object may result in self-occlusion of one of its smooth convex edges, which yields a characteristic luminance gradient pattern that promotes that side to figure. When displays that are constructed that contain extremal edges in conflict with other figure-ground “configural cues”, such as relative size, convexity, and surroundedness, extremal edges best predict human figure-ground judgments (Palmer and Ghose, 2008; Ghose and Palmer, 2010).

Other studies have identified a figure-ground bias for lower regions of bipartite displays in which the two regions are horizontally separated by symmetric articulations (Vecera et al., 2002; Vecera et al., 2004). This figural bias for the bottom region occurred in “stalactite/stalagmite” displays with square articulations, as well as wave and “city-scape” images that had smoother variations around the horizontal border (Figure 1·1m–n). This unusual bias has been interpreted with respect to ecological statistics and gravity (Vecera and Palmer, 2006).

In sum, contextual information such as convexity, symmetry, shadow, occlusion, textural continuity, and region placement highly impacts human figure-ground perception. The underlying mechanisms in the visual system that integrate scene context are complex. The visual system must perform two complementary processes to distinguish between foreground objects from their background—detecting defining borders and integrating parts into wholes (Grossberg and Mingolla, 1985a; Braddick, 1993; Singh and Hoffman, 1998; Kogo et al., 2010). We first focus the discussion on borders and then turn to the integration of parts into wholes.

1.1.1 Visual borders

Borders perceptually partition the visual scene into regions at different depths. Borders occur where an object occludes other elements in the visual scene behind it. Identifying the location of perceptual borders is challenging as individual local vi-
sual attributes, such as sharp changes in luminance contrast (Kanizsa, 1979), color, and motion speed and direction (Kaplan, 1969; Gibson et al., 1969), or combinations thereof, may be sufficient to detect borders between foreground and background objects, but they are not necessary. The Kanizsa square represents a classical example in figure-ground segregation whereby a salient square is perceived that is brighter than the surrounding space, yet the contours that separate it from the background are illusory, not defined by local image properties, and require long-range spatial interactions to be perceived (Lesher and Mingolla, 1993). In general, foreground objects may be perceived as distinct from their backgrounds on the basis of smooth luminance gradients (Mingolla and Todd, 1986; Todd, 2004), differences in texture (Malik and Rosenholtz, 1997; Rosenholtz, 1999), transparency (Beck et al., 1984; Beck and Ivry, 1988), and many other nonlocal properties of the visual scene that require integration over an extended area. Nevertheless, figures are perceptually different and “pop out” from their background (Grossberg and Mingolla, 1985a; Wolfe and Horowitz, 2004). These differences need not readily follow from local, or often even global, luminance properties in the 2D projection of the visual scene on the retina.

Borders, edges, and contours are not the same. Edges refer to physical parts of objects, and their presence implies the existence of a figure. Contours, on the other hand, define the outline of shapes, which may be schematically depicted by line drawings of the boundary of shapes, but they are not physically present. Abrupt changes in image properties, such as luminance, motion, or color, can be represented as contours. Edges encompass sudden (e.g. luminance steps) and graded (e.g. luminance gradients) boundaries between figures and their background, while contours and borders only describe sharp discontinuities between the regions that can be represented by lines (Kogo et al., 2010). Borders are defined herein as perceptual borderlines that separate a figural region from its background, and may appear smooth, abrupt,
or discontinuous (e.g. be “incomplete” with gaps). The distinction between borders, edges, and contours is important for the present discussion because the concept of border-ownership will be introduced. When figure-ground segregation occurs, borders between figure and ground regions perceptually “belong to” or are part of the figure, not the ground. An explicit contour may or may not exist, which is a critical point developed in Chapter 5.

Given an arbitrary 2D image of a visual scene, detecting all the borders between objects and their background is nontrivial. The presence of occlusion, clutter, low contrast image patches, and noise obscures the how contours should be grouped to form smooth object contours. How detected contours should be associated is known as the ‘contour integration’ problem, and comes up during object segmentation in computer vision (Elder and Zucker, 1993). Canny edge detection is a prevalent technique that linearly filters the image using simple cell-like operators to obtain a gradient map (Canny, 1986). Visual cortex also confronts the contour integration problem as it performs figure-ground segregation. Although neurons in V1 respond to luminance steps (i.e. contours), our perceptual experience includes edges (i.e. separations that are physically attached the figure), but not isolated contours. In other words, we see objects but not “line-like” outlines that appear to be recovered in V1. In fact, boundaries have been said to be “perceptually invisible” in the sense that visual perception contains grouped, Gestalt representations, rather than constellations of discontinuous contours (Grossberg and Mingolla, 1985a). A wealth of psychophysical studies support the notion that the visual system solves the contour integration problem by employing classical Gestalt grouping principles. For example, proximity predicts that contours are more likely to be grouped the closer they are to one another. Psychophysical studies suggest that the association strength may follow a power law (Kubovy and Wagemans, 1995; Kubovy et al., 1998; Claessens
and Wagemans, 2008). Other Gestalt principles that garner psychophysical support include good continuation (Geisler et al., 2001), similarity (Gilchrist et al., 1997), closure (Elder and Zucker, 1993), and parallelism (Feldman, 2007). Roelfsema and colleagues have shown using behavioral and single-cell electrophysiological investigations that V1 neurons may use synchrony to associate contours (Roelfsema et al., 2004; Roelfsema, 2006).

Even when borders are detected, they are not per se sufficient to perform figure-ground segregation. Determining figure-ground from borders alone is an underconstrained and ill-posed problem—the principles by which the visual system groups regions separated by the borders into figures must be considered.

1.1.2 Perceptual grouping

Figure-ground segregation critically depends on spatial integration because neurons in the visual system inherently suffer from the aperture problem (Pack and Born, 2001). Any one neuron is limited by the spatial resolution and size of its receptive field. The consequence is that visual properties extracted at any one particular spatial scale of analysis may not correlate with what is actually perceived. The visual system groups a visual scene using emergent perceptual units, rather than image-based local “features”, such as color, texture density, and orientation (Grossberg and Mingolla, 1985b; Grossberg and Mingolla, 1985a). These emergent perceptual units refer to higher-order, meaningful, context-specific entities extracted from a visual image that are used by the visual system to group the global scene. Identifying these units represents a fundamental issue in vision science that is far from being resolved. However, there have been a number of proposals for candidate units in the literature.

The local confluence of different regions of uniform luminance is called a junction, which may be used by the visual system to construct unitary figure representations (McDermott, 2004). A number of junction geometries have been studied along
with related predictions concerning the grouping and relative depth assignment of proximal regions. T-junctions define the confluence of three luminance values at a point. One region of luminance locally has greater area at the intersection (“hat”) than the other two (“stem”). Psychophysical studies suggest that humans tend to perceive the region attached to the T-junction hat as occluding (i.e. figure) the regions attached to the stem (i.e. ground) (Anderson, 2003). While it may be tempting to consider T-junctions perceptual units that specify the occlusion of one region by another, occlusion is perceived in the absence of T-junctions (Yazdanbakhsh et al., 2002; Zaidi et al., 1997). Perhaps more troublesome, neurons in primate visual cortex have not been found that selectively respond to luminance junctions.

The confluence of four luminance values at an ‘X-junction’ has been extensively studied as a signature for the perception of transparency. That is, the percept that an image region adjacent to the X-junction appears a transparent filter over another region, as if light were permitted to partially pass through a translucent medium. Since the time of Metelli’s epicoaster transparency demonstration and influential model (Metelli, 1974), a number of heuristics have been proposed that make predictions about percepts of images containing X-junctions (Adelson et al., 1990; Anderson, 1997; Anderson, 2003; Beck and Ivry, 1988; Beck et al., 1984). For a review and a computational generalization, please see (Grossberg and Yazdanbakhsh, 2005). A number of other proposals for perceptual units have been made, including textons, which consist of line terminators, junctions, and other textural grouping features (Caelli and Julesz, 1979).

Gestalt psychologists proposed a parsimonious set of principles by which grouping may occur (Wertheimer, 1923). These classical principles include symmetry, parallelism, good continuation, closure, and common fate. Common fate is typically considered applicable to motion phenomena whereby elements that move together are
grouped together. However, Sekuler and Bennett showed that as elements became brighter or darker in tandem, despite global variations in luminance, subjects tended to group elements together (Sekuler and Bennett, 2001). Therefore, common fate may not be perceptually restricted to the motion domain. In a field of random black and white dots that reverse in their luminance polarity sign at random temporal frequencies, humans will perceptually group dots that flash synchronously (Lee and Blake, 1999). This grouping could be considered Gestalt common fate, with zero velocity element motion. Palmer and colleagues have proposed another Gestalt grouping principle, which may represent an asymptotic case of proximity, whereby elements that share a common border are grouped (Palmer and Rock, 1994). It has been argued that this principle makes ecological sense, since many objects have multiple connected parts, such as a hammer (Palmer and Beck, 2007).

Whether any of the aforementioned proposals represent perceptual units is unclear, but the mechanisms by which the visual system may transform image properties into the units have not been specified. While they have appeal in their simplicity and psychophysical support, little progress has been made to show how neurons in the visual system organize to implement Gestalt principles. The R cell, G cell, B cell (RGB) model presented in the next chapter predicts that cells called ‘G cells’ have annulus-shaped receptive fields and directly implement the symmetry and closure Gestalt principles. That is, the model builds on known physiological data and makes the prediction that Gestalt principles have been mapped onto cortex.

1.1.3 Local and global interactions in figure-ground segregation

Traditionally, perceptual grouping has been considered a bottom-up, preattentive process that constructs progressively higher-order representations from image properties through a cascade of stages (Marr, 1982). However, feedback is likely involved, as global representations have been shown to bias figure-ground segregation (Peter-
son et al., 1991). Despite their diminutive receptive field sizes compared to figures in the visual scene, V1 neurons demonstrate enhanced responses to figure regions compared to ground regions (Lamme, 1995). Receptive fields in V1 typically extend no farther than 1–2°, yet the figure-ground modulation occurs due to textural properties in the visual scene at least 8° away. When activated by a stimulus, V1 neurons exhibit modulation in their firing rate from signals outside their classical receptive field. In V1 neurons, studies have identified the existence of “near surround” and “far surround” regions that concentrically border the classical receptive field. Anatomical investigations suggest that modulation in the near surround occurs due to horizontal connections between neurons in the same visual area, and modulation in the far surround occurs due to feedback from neurons in higher visual areas (Levitt and Lund, 1997; Schwabe et al., 2010). Extra-classical modulation has been shown to facilitate or suppress the V1 neuron when contrast in the classical receptive field is low or high, respectively (Gilbert and Wiesel, 1990). The prevalence of feedback and horizontal connections early in visual cortex suggests the feedforward, preattentive account of visual processing is incomplete at best. Alternate proposals advocate the influence grouped properties in higher visual areas with larger receptive field sizes and coarse spatial representations have via feedback on earlier visual areas. Reverse Hierarchy Theory proposes a rapid implicit feedforward stage, followed by extensive feedback to focus attention on particular low-level units and properties of the visual scene in what is called “vision with scrutiny” (Hochstein and Ahissar, 2002).

Neuron response latencies across different areas in the visual system provide evidence that processing does not occur in a linear fashion, building more complex representations out of successively simpler ones, as predicted by preattentive, feedforward theories. Areas MT, MST, and FEF do not greatly differ in their response latencies from V1. Neurons located toward the end of the ventral stream in inferotem-
poral (IT) cortex differ in latencies with respect to V1 by only $\sim 40$ msec (Lamme and Roelfsema, 2000; Nowak and Bullier, 1997; Bullier, 2001). Even within an area, such as V1, neural activation due to brief visual display presentations persists for hundreds of milliseconds, and scatter in the onset responses within a visual area can reach 60–80 msec (Zipser et al., 1996). Instead of serving as a early step in feedforward processing cascade, perhaps V1 is intimately involved in all stages of visual perception as an active “blackboard” and nexus for feedforward, horizontal, and feedback signals (Bullier, 2001). Perhaps V1 participates in selective processing rather than in fixed feedforward operations (Paradiso, 2002).

The extensive evidence for global-to-local interactions by way of feedback in the visual system echoes the difficulty in modeling the visual system using feedforward, local operators (Marr and Hildreth, 1980; Canny, 1986). Poor performance with local operators motivated nonlocal post-processing of edge maps (Zucker et al., 1989; Iverson and Zucker, 1995). The use of visual junctions, such as T- and X-junctions, in the detection of occlusion and transparency, respectively, are appealing due to explicit rules that can be constructed to improve model performance. However, these approaches have been limited by the inherent noise in local processing (Freeman et al., 2000). The presence of visual junctions are not always discernible on the basis of local information, which implicates at least some feedback and nonlocal processing (Lee and Mumford, 2003).

1.2 Visually-guided navigation

How do animals coordinate self-motion and action about their environments? Because vision represents a dominant sense for many animals, research has examined how vision contributes to navigation. The spatio-temporal deformation of the retinal image that occurs due to movement of an animal, movement of objects in the
environment, or both is known as *optic flow*. Visually-guided navigation refers to an animal’s use of optic flow to coordinate self-motion through the environment.

Gibson suggested that sighted animals use the rich information contained within the optic flow field to guide their self-motion (Gibson, 1950). When an animal moves, a large field pattern of motion is produced on the retina. The motion pattern induced by observer movement is relative to that of the environment, and therefore does not necessarily indicate the absolute speed. All other factors remaining equal, closer objects result in faster motion and far away objects result in slower motion (*motion parallax*) (Longuet-Higgins and Prazdny, 1980). Objects that get closer or farther from the observer increase and decrease their retinal speeds compared to if they were stationary. Forward movement of an observer along a straight path without eye, head, or other rotation results in a characteristically radial optic flow field. Visual elements that are displaced on the retina are often depicted as vectors, corresponding to the first-order approximation to the optic flow. Note that optic flow generally contains far richer high-order spatio-temporal information than the first-order representation affords. The optic flow during forward self-motion radiates from a singular point of outflow, where the velocity of retinal motion is zero. This singularity is termed the *focus of expansion* (FoE), and there is a corresponding *focus of contraction* (FoC) when the animal moves backward. We define *heading* as the instantaneous direction of travel, which is specified by the FoE/FoC in the simple case of travel along a straight path without rotation.

Gibson proposed a strategy whereby animals control locomotion using optic flow by aligning the FoE with their desired destination (Gibson, 1950; Gibson, 1979). To move forward, the animal should “contract the muscles so as to make the optic array flow outward,” and to stop the animal should “make the flow cease” (Gibson, 1958). Gibson’s strategy requires information about heading or the FoE in the case of
straight paths of travel. Unfortunately, the optic flow patterns routinely experienced by animals are far more complex than the simple radial case, and Gibson’s simple approach is insufficient to describe the control of self-motion. For instance, eye, head, and body rotations that frequently occur during locomotion deform the optic flow field such that singularities may no longer specify the heading direction. Moreover, animals travel along curvilinear paths, trajectories that may be straight or curved, which also may introduce rotation into the optic flow field. For Gibson’s proposal to apply under general locomotion, at least along straight paths, heading must be recovered from the optic flow field. The rotation problem is introduced and developed in the next section of this chapter.

Assuming heading is recovered, studies provide support for Gibson’s strategy. Both human infants and adults stabilize their standing posture by zeroing out the optic flow (Lee and Aronson, 1974; Lee and Lishman, 1975). The canceling of optic flow has been shown to support the approach of a target and brake without collision. Lee defined an optical variable \( \tau \), the ratio between the visual angle subtended by the target on the retina and its rate of change, that is specified by optic flow and provides information about the observer’s time-to-contact (ToC) (Lee, 1976). The derivative, \( \dot{\tau} \), specifies whether the observer’s deceleration is sufficient to avoid a collision. For example if \( \dot{\tau} = -1 \), the observer moves at a constant velocity and if the present course is maintained, a collision will occur. Psychophysical experiments have shown that humans can steer toward a target using optic flow by equalizing the time-to-closure between the target and heading directions and the observer’s time-to-passage with the target, which depends on \( \tau \) (Fajen, 2001). Experiments estimate \( \dot{\tau} \) to approximately equal -0.5, which supports an adequate rate of deceleration to stop in front of an object (Kim et al., 1993; Yilmaz and Warren, 1995). However, human subjects have been shown to use the global optic flow rate, rather than \( \dot{\tau} \), to control
of braking (Fajen, 2005; Fajen, 2008).

To pursue or intercept a moving target, research suggests that observers maintain a fixed bearing angle between their trajectory and the target (constant bearing angle strategy). Bearing defines the angular difference between the observer and the target in a world coordinate system. A trajectory with a constant bearing angle to a pursued target implicates that an interception will occur. If pursuit or avoidance is desired the course should be maintained or abandoned, respectively. The constant bearing angle strategy best fits empirical data on bats using echolocation to pursue prey (Ghose et al., 2006), human pursuit and interception of stationary and moving targets (Fajen and Warren, 2003; Fajen and Warren, 2007), baseball outfielders catching fly balls (Chardenon et al., 2005; Fink et al., 2009), and human interception on foot (Lenoir et al., 2002). Fajen and Warren created a behavioral dynamical systems model that simulates the trajectories of human subjects performing interception tasks of stationary and moving targets while avoiding obstacles (Fajen and Warren, 2003; Fajen, 2008). Targets and obstacles are considered point attractors and repellers, respectively. However, the model assumes a priori knowledge of heading and models the observer in a third-person perspective as a particle moving according to a physical mass-spring second order differential equation system. In other words, it does not obtain visual information from optic flow, as humans do. Elder, Grossberg, and Mingolla showed that trajectories produced by a computational model of primate visual cortex that processes the optic flow experienced by an observer pursuing a target are consistent with those in human psychophysical experiments (Fajen and Warren, 2003) and the constant bearing angle strategy (Elder et al., 2009).

Steering represents another navigation behavior that appears to be controlled at least in part through visual information in optic flow. Gibson proposed that observers steer by adjusting heading to maintain the FoE centered on the target to
magnify it, and position the FoE outside objects to avoid them (Gibson, 1958). In the psychophysical experiments of Fajen and Warren, humans used optic flow to steer in cluttered, complex visual scenes toward targets and away from obstacles (Fajen and Warren, 2003). Land and Horwood showed that humans, when driving cars, use visual information about the distant parts of the road to gauge road curvature and near parts to correct in-lane steering deviations (Land and Horwood, 1995). Hildreth and colleagues demonstrated that humans initiated corrective steering maneuvers with deviations in heading, even at high speeds (Hildreth et al., 2000). When steering down a narrow corridor, humans, like bees, appear to equalize the optic flow produced from the left and right walls (Duchon and Warren, 2002; Srinivasan et al., 1996). Therefore, evidence suggests that optic flow guides steering through the perception of heading.

Unlike Gibson’s theory, other proposals for steering do not require the recovery of heading. Wilkie and Wann performed a series of experiments that showed humans directed gaze to regions on a curved roadway proximal to where they would subsequently steer (Wilkie and Wann, 2002; Wilkie and Wann, 2003b; Wilkie and Wann, 2003a; Wilkie and Wann, 2005; Wilkie et al., 2008; Robertshaw and Wilkie, 2008). Moreover, steering bias toward the inside or outside roadway edges is consistent with the direction of gaze (Wilkie et al., 2010). Subjects have also been shown to gaze toward the tangent point of a curved bend while driving (Land and Lee, 1994). Land and Lee demonstrated that the angle of gaze while fixating the tangent point of a roadway bend is related to the road curvature and in principle, observers could steer proportional to the gaze angle to negotiate the constant curvature of the bend. Therefore, active gaze to cancel out target drift, not the FoE in the optic flow, may guide steering. Wann and Swapp proposed another theory whereby humans steer toward a target using optic flow, but do not need to perceive their heading or integrate extra-
retinal signals (Wann and Swapp, 2000). The authors showed that when humans maintain gaze on a stationary target that is to be intercepted, a sequence of vertical streamlines in the optic flow emerge that connect the observer and the target. Steering according to the vertical flowlines results in successful target interception (Wann, 2000).

However, other studies suggest humans may not necessarily depend on optic flow to steer. Rushton and colleagues studied locomotion on foot toward a target by equipping humans with prism goggles, which manipulates optic flow by shifting the target position on the retina (Rushton et al., 1998). An optic flow-based strategy, as proposed by Gibson, predicts that the relative position of the FoE and the target is important for guiding steering. Because prism goggles shift the entire optic flow field, the FoE and target relative positions do not change but the absolute position on the retina does. With or without prism goggles, an optic flow mediated strategy predicts that subjects will walk along straight paths to the target. However, Rushton and colleagues found that humans walked along characteristically curved paths toward the target while wearing goggles, which is consistent with the perceived egocentric location of the target (Rushton et al., 1998). That is, perceived position in space appeared to guide steering, not optic flow. Experiments performed by Harris and Bonas replicated the finding that humans appear to ignore optic flow and use perceived egocentric direction to steer toward targets when wearing prism goggles (Harris and Bonas, 2002). In more structured environments with rich texture, studies show an increased reliance on optic flow to steer toward stationary targets (Warren et al., 2001; Bruggeman et al., 2007). By manipulating the amount of texture present in an immersive virtual environment, Warren and colleagues found that humans used an egocentric direction strategy as the amount of texture in the environment decreased. The authors concluded that in general humans use a combination of optic
flow and egocentric direction, depending on the availability of rich environmental structure (Warren et al., 2001). Interestingly, humans appear to rely on egocentric direction, not optic flow, when steering to intercept moving targets (Fajen and Warren, 2003).

1.2.1 Rotation in optic flow

The fact that primates actively change their gaze complicates theories of visually-guided navigation. Gibson suggested that locomotion is controlled by using the FoE to steer toward a desired destination. The theory assumes that heading must be detected to guide locomotion. However, fixating targets (Wann, 2000; Wann and Swapp, 2000), the tangent point of a bend in the road (Land and Lee, 1994), or other objects in the environment introduces rotation into the pattern of optic flow that appears on the retina (retinal flow). Eye, heading, or body rotations drastically change the appearance of the retinal flow. A stationary human performing a smooth-pursuit eye movement results in a uniform motion pattern whereby elements are shifted on the retina by a constant speed. As first-order optic flow, motion vectors with the same length span the entire optic flow field, pointing in the direction opposite that of the eye movement. This is called the rotational component of optic flow (Longuet-Higgins and Prazdny, 1980). Recall that travel along a straight path without rotation induces a radially expanding optic flow field, which is known as the translational component of optic flow. A moving observer who executes a smooth pursuit eye movement generates the vector sum of the translational and rotational component fields. The uniform shift from the rotational component moves the location of the singularity such that it no longer coincides with the heading direction. This poses a challenge to theories of visually-guided navigation that depend on heading, such as Gibson’s. However, psychophysical experiments demonstrate that humans accurately judge heading while performing eye movements (Warren and Hannon, 1990; Royden...
et al., 1992; Royden et al., 1994), which suggests that the visual system recovers the translational component of optic flow, which specifies heading.

Two main classes of computational models propose solutions to the rotation problem. The first class of models exploits differential motion to recover heading. The motion vector magnitudes in the translational component of optic flow have an inverse dependence on depth in the scene. Motion far from the observer results in shorter motion vectors and closer motion results in longer motion vectors. As in the smooth pursuit eye movement example in the absence of self-motion, motion in the rotational component is independent of depth. Differential motion models subtract spatially proximal flow vectors, which in principle cancels out the rotation and results in a scaled version of the translational component. Heading, by way of locating the FoE, can be recovered by triangulating the scaled translational vectors (Rieger and Lawton, 1985; Heeger, 1992; Hildreth, 1992).

The second approach employs extra-retinal signals to cancel out the rotational component in the retinal flow to recover heading, which assumes the rotation is introduced by muscular contractions performed by the observer (Lappe et al., 1996; van den Berg and Beintema, 1997). Vestibular signals from the contraction serve as an efference copy that may be subtracted from the retinal optic flow signal. The extra-retinal approach is supported by the fact that MSTd neurons demonstrate tuning to vestibular eye signals (Bradley et al., 1996; Ben Hamed, 2003). Human psychophysical research also supports the use of extra-retinal signals to recover heading. Warren and colleagues presented human subjects with two types of optic flow displays to determine whether extra-retinal signals are used (Warren and Hannon, 1990). The first type of display simulated travel along a straight path while performing a smooth-pursuit eye movement along a ground plain or through a cloud defined by dots. The human subjects would fixate during each trial, but the displays con-
tained a constant amount of simulated rotation. The second type of display showed a simulated sequence of travel along a straight path without rotation, but subjects were instructed to perform a smooth-pursuit eye movement. Thus, there is rotation in both displays, but only the latter contains extra-retinal signals. Warren found that at walking speeds (\(~1-2^\circ/\text{sec}\)), heading judgment accuracy was comparable with or without extra-retinal signals. Royden and colleagues performed similar experiments, replicating the findings of Warren and Hannon, but found heading judgments markedly decreased in accuracy at higher speeds when extra-retinal signals were not present (Royden et al., 1992; Royden et al., 1994; Crowell et al., 1998; Ehrlich et al., 1998). Van den Berg and colleagues tested heading judgments in displays that introduced noise, defined by dot motion that randomly deviated in direction from those produced by simulated self-motion. Heading judgment accuracy rapidly degraded in the simulated rotation condition. Performance decreased to a much weaker degree when extra-retinal signals were available, particularly in the 3D dot cloud environment (van den Berg, 1992; van den Berg, 1996).

1.2.2 Physiology of navigation

Electrophysiological studies show that areas in the primate dorsal stream respond to the patterns of motion experienced during locomotion, and the representation of motion patterns becomes progressively more complex along the visual hierarchy. Areas in the dorsal stream that demonstrate sensitivity to optic flow patterns include V1, medial temporal area (MT), MSTd, ventral interparietal area (VIP), and area 7a. V1 neurons have small receptive field sizes (\(~1^\circ\)) and respond to the speed and direction of local motion patterns (Bair, 2004; Livingstone and Conway, 2003). Therefore, V1 neurons are not well suited for the analysis of the complex motion patterns experienced during locomotion. V1 projects to MT\(^+\), which contains neurons with roughly 2–5 times larger receptive fields at a given eccentricity (Gattass and
Gross, 1981; van Essen and Maunsell, 1983). MT$^+$ neurons preferentially respond to large uniform motion patterns, and responses markedly decrease when multiple motion directions are present (motion transparency) (Snowden et al., 1991). Because MT$^+$ receives direct projections from V1, MT$^+$ neurons are thought to spatially pool over V1 neurons with common motion direction and speed tunings (Born and Bradley, 2005). Like V1, MT$^+$ neurons do not respond to complex motion patterns that are experienced during locomotion.

MST represents the first dorsal area in which neurons were discovered that selectively respond to complex large-field motion patterns. The motion pattern sensitivity encompasses up to 100° of the bilateral visual field (Duffy and Wurtz, 1991a; Duffy and Wurtz, 1991b; Tanaka et al., 1986). MST contains two functionally and anatomically distinct subregions: dorsal MST (MSTd) and ventral MST (MSTv). MSTd contains neurons tuned to radial expansion and contraction, similar to the pattern of optic flow experienced while traveling along a straight path, and therefore has been the focus of most investigations of MST. The few studies of MSTv have identified neurons tuned to object motion trajectories, particularly at faster speeds (Duffy, 1998; Born and Tootell, 1992; Tanaka et al., 1993). Neurons in MSTd tuned to radial motion patterns are biased for expansion, which is suggests the neurons are involved in heading detection, since animals more frequently move forward than backwards. Indeed, MSTd neurons exhibit tuning to the FoE location within the optic flow field (Duffy and Wurtz, 1995). DeAngelis, Angelaki, and colleagues showed that MSTd neurons exhibit sensitivity to 3D heading direction and more neurons across the MSTd population exhibit tuning to headings in the left and right regions of the visual field than the center (Gu, 2006). A population decoding analysis suggests this bias may afford greater sensitivity to headings in the central region of the visual field (Gu et al., 2010). Many electrophysiological studies probe MSTd tuning properties using dots on a 2D
plane, but heading tuning improved when dots were distributed across multiple depth planes (Upadhyay et al., 2000). MSTd neurons exhibit tuning to global speed changes in radial optic flow patterns, suggesting that the neurons are sensitive to the speed of locomotion (Duffy and Wurtz, 1997).

Consistent with extra-retinal theories of heading recovery in the presence of rotation, MSTd neurons demonstrate multi-modal tuning to both visual and vestibular signals. Although MSTd neurons exhibit tuning to vestibular signals alone (i.e. in the absence of visual stimulation), the combination of visual and vestibular signals generally yields the greatest response (Page, 2003; Fetsch et al., 2013). Curiously, vestibular and visual tuning occur in opposite directions (~180°). For example, a MSTd neuron tuned to a heading direction on the right side of the visual field might show a vestibular sensitivity along leftward linear trajectories (Gu et al., 2012). In the context of extra-retinal signals during smooth-pursuit, Bradley and colleagues demonstrated that MSTd neurons compensate for rotation in the retinal flow signal by activating neurons that are tuned to the veridical heading direction (Bradley et al., 1996; Shenoy et al., 1999). MSTd neurons on average undercompensate by shifting the retinal flow ~ 70% the necessary amount to bring the optic flow in head-centered coordinates. The degree of compensation scales with the speed of pursuit and the monkey’s translation (Shenoy, 2002). Together, these results provide robust support for extra-retinal theories for heading recovery.

Comparatively less attention has been allocated to other areas in primate cortex sensitive to optic flow. The ventral parietal area (VIP), like MSTd, receives feed-forward input from MT (Felleman and van Essen, 1991) and contains neurons sensitive to radial expansion, contraction, spiral patterns, and centers (Bremmer et al., 2002; Schaafsma and Duysens, 1996; Schaafsma et al., 1997; Bremmer, 2005). VIP neurons exhibit multi-sensory tuning to visual, auditory, vestibular, and somatosen-
sory signals. The visual optic flow tuning is similar to that of MSTd, but receptive fields are comparatively smaller and neurons respond vigorously to directional planar motion (Colby et al., 1993; Zhang et al., 2004). The parallel projections from MT to VIP and MSTd and the similar complex visual motion pattern sensitivity suggests redundancy in the detection of heading. Somatosensory tuning in VIP is often consistent with the visual tuning. For example, a visual tuning to the periphery of the visual field might accompany a somatosensory tuning to the shoulder or arm (Duhamel et al., 1998). Interestingly, VIP neurons exhibit depth-dependent visual responses, which may implicate the area in the processing of object time-to-contact, approach, and avoidance. For example, a VIP neuron may stop responding when an approaching object comes within a certain distance of the monkey (Colby et al., 1993). Microstimulation of VIP results in defensive/evasive shoulder, face, and arm movements (Graziano and Cooke, 2006; Cooke et al., 2003). VIP and MSTd both send feedforward projections to area 7a and the superior temporal polysensory area (STP), which also contain neurons tuned to optic flow patterns (Merchant et al., 2001; Merchant, 2004; Siegel and Read, 1997; Anderson and Siegel, 1999; Anderson and Siegel, 2005). However, the visual responses tend to be more complex. In area 7a, the response of neurons depends on the speed of IMOs (Phinney and Siegel, 2000). Although STP neurons respond to observer translation, they tend to exhibit stronger tuning to rotating objects in depth (Anderson and Siegel, 2005).

1.3 Outline of the thesis

Because figure-ground segregation and visually-guided navigation underlie so many important behaviors of humans and other sighted animals, the present thesis aims to elaborate their neural mechanisms. Gibson theorized that visual perception arises in the “retino-neuro-muscular system” (Gibson, 1973). In this vein, the research pre-
sented herein assumes that behavior and perception are intimately coupled to the activity of neurons in the brain. We seek the physiological basis for figure-ground segregation and visually-guided navigation. The approach taken in the present thesis models how these perceptual phenomena in vision arise emergently through the dynamical interactions among networks of neurons distributed in different areas of visual cortices. Constraints are imposed by known primate anatomy and neurophysiology and emergent neural dynamics are related to the results of human psychophysical experiments.

The thesis is divided into two parts, corresponding to figure-ground segregation and visually-guided navigation. Even numbered chapters focus on descriptions and the context of the problems models aim to address, while odd numbered chapters introduce models, analysis, and simulation results of neurophysiological and psychophysical data. Chapter 3 presents computational neural models of border-ownership perception in statically scenes, Chapter 5 introduces an extension for kinetic shapes, and Chapter 7 presents a model of shape detection. Chapter 9 presents models of heading perception in the presence of IMOs, Chapter 11 introduces a model of the impact attentional signals have on MSTd during locomotion, Chapter 13 develops a neural theory of curvilinear path perception, and Chapter 15 presents an analysis of the path perception model. Chapter 16 concludes the thesis with descriptions of model predictions and results in the context of the current state of psychophysics and electrophysiology of visual cortex.

Chapter 3 develops a model of how inter-cortical networks of neurons organize to give rise to the perception of border-ownership, a crucial aspect of figure-ground segregation. The R Cell, G cell, B cell (RGB) neural network model demonstrates how fast inter-areal connections between areas V1, V2, and V4 result in border-ownership assignment. Existing models of border-ownership assume the visual system contains
units that detect luminance junctions to resolve the impact occlusion has on border-ownership. However, little evidence exists to support neural junction detectors. The detection of occlusion occurs in the RGB model as an emergent outcome of interareal competitive dynamics. Feedback from neurons in higher visual areas with large receptive fields propagate contextual information about figure-ground segregation to neurons with side-of-figure selectivities. The circuit biases border-ownership toward region interiors that likely represent figures, and suppresses border-ownership toward the ground. Model neurons in V4 have receptive fields that implement the Gestalt grouping principles of convexity and closure.

Consistent with human perception of border-ownership, corresponding neural signals seem to develop whenever a figure border is perceived, independent of its low-level visual appearance. Border-ownership signals are produced despite whether a figural edge is defined by luminance contrast, motion, disparity, color, and so forth. A goal of the thesis is to develop the computational principles by which border-ownership signals emerge in a broad range of visual contexts. In Chapter 5, a neural model is developed that demonstrates how border-ownership signals emerge in visual displays that contain figures defined only by kinetically moving dots. The research extends the RGB model, which focused on circuits concerned with edges defined by luminance contrast, to handle kinetic edge border-ownership. Neurons have been discovered in V2 that are sensitive to kinetically defined edges, both due to the orthogonal (i.e. accretion/deletion) and parallel (e.g. shearing) motion of dots (Marcar et al., 2000), and in V4 that respond to kinetically-defined figures (Mysore et al., 2006). The model proposes the a subpopulation of kinetic-edge cells in V2 are border-ownership selective and V4 neurons play a crucial role in grouping kinetically-defined shapes, enhancing border-ownership signals in the presence of texture accretion/deletion. The model yields emergent border-ownership around kinetic edges, without explicit kinetic edge
detection and region segmentation, and relative depth is coded in the differential border-ownership response magnitudes.

A central problem that is solved by the primate visual system is the detection of shapes. Neurons in ventral visual areas demonstrate an increased response gain when a region appears within the receptive field that represents an interior portion of a shape. This is a challenging problem because the convexity of local shape boundary patches is ambiguous: it is not clear within a limited-size aperture whether a set of local boundary contours define a convex (i.e. inside) or concave (i.e. outside) part of a shape. How do neurons, with limited receptive field sizes, exhibit enhanced activity to figures? A computational model is introduced in Chapter 7 that proposes that scatter in neuronal receptive field size and visuotopic position within the visual system serve a crucial role in detecting the inside versus outside of shapes. The core model circuit shows how the fusion of shape representations at multiple spatial scales may be used to detect the interior of shapes. Feedback is responsible for enhancement of activity to the interior and suppression of activity to the exterior of shapes.

This thesis also aims to clarify the mechanisms underlying visually-guided navigation, in particular in the presence of IMOs. Chapter 9 develops a computational to show that motion pooling can explain human heading perception in the presence of approaching and non-approaching independently moving objects by biasing the peak model MST activity in the direction of the reported human bias. Existing models of human heading perception rely on differential motion operators, which are not consistent with known neurophysiology. Special focus is dedicated to how heading bias from human psychophysical experiments can be explained using known physiological properties of neurons in MSTd and those that project to MSTd.

Humans frequently shift their focus of attention while navigating, for example, depending on the favorable or threatening context of approaching independently moving
objects. Recent neurophysiological studies show that the spatial tuning curves of primate dorsal MST (MSTd) neurons change based on the difference in visual angle between an attentional prime and the FoE (Dubin and Duffy, 2007; Dubin and Duffy, 2009). The proposed dynamical model of MSTd in Chapter 11 explains the electrophysiological finding that the mean MSTd population response peak recedes in time as the visuotopic distance traveled by attention increases due to competition between top-down attentional and bottom-up sensory neural signals. This result has important implications for the dynamics of human locomotion while attention is paid to different regions of visual space.

Finally, a goal of the present thesis is to understand navigation along trajectories that may be curved. Perceiving future curvilinear paths may be useful for gauging whether a traversed path may lead to an undesired collision. Chapter 13 develops a model that shows how MSTd simultaneously represents heading and path curvature in the same neural population. No existing model addresses the neural representation underlying the perception of future path. The model reproduces human path estimation bias and proposes that MSTd neurons sensitive to radial and spiral motion patterns encode path curvature. Chapter 15 presents mathematical analysis of the model, as well as the optic flow experienced by subjects as a function of where they look while traveling along curvilinear paths.
Chapter 2

Border-ownership: A vector code for figure-ground segregation

2.1 Border-ownership

Border-ownership has been an integral part of the Gestalt tradition in psychology, since the time of Wertheimer and Rubin in the early part of the twentieth century (Wertheimer, 1923; Rubin, 1921). In the famous vase display of Rubin, shapes of a vase and face share common borders and perceptually compete for a figural interpretation. Humans perceive a reversal in the figure-ground assignment between the regions corresponding to the vase and the face—the display is perceptually bistable. Nakayama and colleagues demonstrated in a psychophysical experiment of stereoscopically defined occlusion that the shared contour between the occluding (closer in depth) and the occluded object (farther in depth) always grouped with the closer object (Nakayama et al., 1989). Because the boundary between the vase and face can only be grouped with one figure at a time, the border-ownership of the boundary reverses with the figure-ground interpretation of the scene.

Until von der Heydt and colleagues demonstrated the existence of neurons in primate visual cortex that signal border-ownership when the boundaries of figures appear within the receptive field, border-ownership was thought to only represent a perceptual phenomenon. The firing rate of a border-ownership neuron depends on where the figure is located relative to a boundary that enters the neuron’s receptive
field (Zhou et al., 2000). This finding is important because it reveals the outcome of the neural coding of a phenomenon directly available in human conscious experience. As stated previously, the responses of other neurons early in visual cortex, for example to edges, do not isomorphistically map so directly onto human visual perception. Each border-ownership neuron has a single side-of-figure preference, and as a population, all side-of-figure directions are equally represented. For example, a rightward-selective border-ownership neuron yields a higher firing rate when the left boundary of the square is in the classical receptive field than when the right boundary is, even when the local contrast is the same. In principle, the border-ownership of boundaries between foreground and background regions may be coded through opponent pairs of border-ownership neurons. Neurons selective to border-ownership also demonstrate tuning to edge orientation, but activation to isolated edges are weaker than when the edge is attached to a figure on the preferred side. Border-ownership neurons have been found in V1, V2, and V4, which is surprising because the neurons have small receptive field sizes, yet demonstrate large-range context integration. In the sample, the median receptive field size was less than 3.5° (Zhou et al., 2000) and figures spanned 10° or more. Moreover, the differential side-of-figure response (i.e. figure to one side versus figure on the other) latency is less than 25 msec (Zhou et al., 2000). The latency is shorter than that of the figure-related response enhancement in V1 (Lamme, 1995).

The importance of the discovery of border-ownership neurons in primate visual cortex is difficult to overstate. Border-ownership represents a vector code. It simultaneously signals two pieces of information: the orientation of a figure’s boundary contour and position of the figure, relative to the receptive field. This fundamentally differs from the response properties of neurons involved in figure-ground segregation early in visual cortex. Neurons tuned to color (Zeki, 1973), edge contrast polar-
ity (Hubel and Wiesel, 1962), curvature and shape (Pasupathy and Connor, 1999),
brightness (Roe and Ts’o, 1995), and disparity and depth (Poggio and Fischer, 1977)
signal scalar codes in the sense that neurons are tuned to and signal the presence of a
single visual attribute. Identification of neural responses to scalar properties of objects
in the visual scene has been the approach in electrophysiology of the visual system
since the time of Hubel and Wiesel (Hubel and Wiesel, 1962). Motion-sensitive neu-
rons in V1 and MT that simultaneously signal motion direction and speed represent
a vector code in the dorsal stream. Responses occur due to the local presentation
of the motion pattern within the receptive field and often not require contextual
integration across large parts of the visual scene (DeAngelis and Newsome, 1999).
Border-ownership neurons are fairly unique early in the visual system in general due
to their vector code, the nonlocal integration of information outside the classical
receptive field, and the direct representation of a figure-ground property that is so
directly present in human visual perception.

Vector coding in border-ownership neurons has significant implications for many
existing computational models of figure-ground segregation. Many existing models
for figure-ground segregation assume that ‘featural’ brightness or color signals propa-
gate omnidirectionally throughout a visual area to isomorphistically construct a figure
surface representation (Grossberg and Mingolla, 1985b; Arrington, 1994; Grossberg,
1997; Roelfsema et al., 2002). Regions are labeled or tagged according to their cor-
responding brightness percepts, and segregated by completed boundary signals. Fea-
tural properties spread in all directions (i.e. “fill-in”) because they implicate scalar
information: the brightness at a particular visuotopic location only signals its value
at that location. Boundary signals are therefore required to constrain the unoriented
featural propagation, else brightness would leave its associated surface region. The
vector code conferred by border-ownership if sufficient for figure-ground segregation
and surface regions do not explicitly need to be tagged by isomorphic brightness. Border-ownership also has the potential to greatly improve filling-in models, as the spread of featural properties can be guided by the border-ownership vector code at region boundaries. For example, border-ownership signals could bias featural signals in one direction, but not another, based on large-range contextual properties of the visual scene.

With their small receptive field sizes and rapid response latency, how do border-ownership neurons acquire information about side-of-figure location? Von der Heydt and colleagues demonstrated that boundary contour fragments of figures far outside the receptive field individually contributed to border-ownership signals. Interestingly, modulation due to distant fragments occurred earlier in time than nearby fragments (Zhang and von der Heydt, 2010). Some existing models have proposed that intra-areal (i.e. within a visual area) horizontal fibers propagate the contextual information (Pao et al., 1999; Nishimura and Sakai, 2004; Zhaoping, 2005; Baek and Sajda, 2005; Kogo et al., 2010). While these models do in fact yield border-ownership representations, intra-areal connections do not satisfy biological timing constraints. Horizontal fibers in early visual areas, such as V1 and V2, are not myelinated (gray matter) and conduct signals far slower than myelinated inter-areal (i.e. feedforward and feedback connections between visual areas) (Angelucci et al., 2002). Moreover, horizontal connections have limited lengths, yet computational models assume that all the connections fully span a visual area. Given the large receptive field sizes in visual areas further down the visual hierarchy and the sufficient speed of inter-areal connections, it is likely that feedback plays a crucial role in the development of border-ownership signals.
2.2 The RGB model of border-ownership

We present a computational model of figure-ground segregation that clarifies the mechanisms by which border-ownership cells obtain their sensitivity to the location of figure. The model contains stages that correspond to areas LGN, V1, V2, and V4 in primate visual cortex. Model neurons in V1 and V2 develop selectivity for border-ownership (B cells), model neurons in V2 group contrast edge signals in convex annular shapes (G cells), and model neurons in V4 perform on-surround competition (R cells). For this reason, the model is called the RGB model.

The model demonstrates a number of important properties. First, model B cells demonstrate side-of-figure selectivities toward regions in a visual image that are perceived by humans as figure, which is consistent with known neurophysiological border-ownership signals. Second, cortical networks that propagate global contextual information to small receptive field B cell units bias border-ownership signals toward the inside rather than the outside of figures. Third, the model demonstrates the border-ownership selectivity, despite the size of figures on the retina. Fourth, the model uses inter-areal connections with fast conduction latencies within the physiological range between B, G and R cells to explain how border-ownership signals emerge. G and R cells, with their larger receptive field sizes, propagate the information about grouped edge signals and likely figure location over large areas of the visual field to border-ownership cells.

B cells interact with spatial scale G cells with annular receptive fields, sensitive to Gestalt convexity and closure. While G cells receive feedforward input from B cells within the receptive field, G cells also feedback to inhibit B cells with side-of-figure preferences away from the center of the annulus. The selective feedback biases the border-ownership signals toward the center of the annulus by suppressing the activity of B cells with inconsistent side-of-figure preferences. R cells pool over G cells and
compete across scale. This resolves the likely spatial scale of figures in the visual scene, and demonstrates an emergent property whereby occlusion is detected without explicit junction detectors. Existing models of border-ownership require junction detectors or explicit rules to resolve the occlusion of one figure by another (Finkel, 1992; Zhaoping, 2005; Craft et al., 2007; Mihalas et al., 2011). R cell competitive dynamics “invert” G cell activity in the presence of T-junctions to signal occlusion. Following the competition, R cells feed back to B cells to inhibit B cells with side-of-figure preferences away from the peak R cell activity, which produces a bias in the border-ownership signals toward the figure.

2.3 Contributions

The following section introduces the RGB model in the form of a peer-reviewed publication (Layton et al., 2012b). As first author of the study, I developed the model architecture, implemented and tested it on the set of visual displays, analyzed the results, and wrote and revised the manuscript. The coauthors of the manuscript oversaw progress and contributed in the capacity of academic advisors.
Chapter 3

Dynamic coding of border-ownership in visual cortex
3.1 Introduction

Most physiological studies of the primate visual system in the past half-century have followed the path established by Hubel and Wiesel (1962) and largely focused on the function of individual areas or subpopulations of cells within a visual area. Investigations outside the visual system indicate that the cortex can solve complex problems with networks that span multiple areas and whose functionally equivalent circuits are widely distributed throughout the cortex (Nieder & Miller, 2004). Our computational analysis indicates that the visual system may also rapidly recruit an assembly of cortical areas to determine border-ownership in figure-ground segregation, a single emergent function. Neuroanatomical evidence indicates that early visual areas such as LGN, V1, V2, and V4 are massively interconnected with numerous feedforward and feedback connections (Sincich & Horton, 2005). Feedforward connections are believed to quickly propagate sensory visual information to cortical areas further up the visual hierarchy to subserve a rich perception of the visual scene. Feedback projections are often said to play a modulatory role with respect to bottom-up sensory visual signals by increasing the gain of neuronal responses in attended regions and performing contextual integration. To date, few studies have hypothesized that feedback projections subserve crucial as opposed to supplementary roles for the functions of early visual cortices. It is not clear whether the simultaneous activation of multiple areas early in the visual system only performs modular functions that are later combined or whether such activation can collectively solve problems that individual cortical areas cannot solve alone. We here introduce a computational model that provides a unified explanation for how several cortical areas act coherently to perform figure-ground segregation.
Figure 3-1. (a) Bistable image in which observers either perceive a saxophone player or a female face; from Shepard (1990). The central white-black borders alternate their direction of ownership with the scene interpretation. (b) The cell has the same light gray and dark gray in the left and right halves of its receptive field (ellipse), respectively, in both (b) and (c), but elicits a larger response (border-ownership preference) when the light gray patch is attached to a figure located to the left of the receptive field (b; d, light gray curve) compared to when the dark gray patch is attached to a figure on the right (c; d, dark gray curve). (e) The same B cell continues to prefer figures whose border enter the cell’s receptive field and are located to the left despite the presence of transparent overlays. Although the cell yielded a stronger response when a light square appeared to the left compared to a dark square appearing to the right (d), the presence of a dark transparent overlay to the right (e) diminishes the cell’s response (g; light gray curve). (f) When the local luminance configuration remains the same, but a light transparent overlay appears within and to the left of the cell’s receptive field, the cell’s response increased (g; dark gray curve). Panels (d) and (g) are adapted with permission from Figure 3-1 of Qiu et al. (2007). Panel (a) excerpted from MIND SIGHTS by Roger N. Shepard. Copyright © 1990 by Roger N. Shepard. Reprinted by arrangement with Henry Holt and Company, LLC. All rights reserved.

Distinguishing between an object (figure) and its background (ground) in a visual scene is required for performing important higher-order visual functions, such as object recognition. Although figure-ground segregation is fundamental to visual perception, how the visual system performs it is not well understood. A direct link between visual figure-ground perception and the responses of certain single neurons has, however, been established in the early visual system. These cell responses may require the simultaneous activation of parts of visual areas V1, V2, and V4 acting as a functional network.
Researchers have found that as many as 59% or 53% of sampled cells from primate visual areas V2 and V4, respectively, preferentially respond to borders when they form a certain side of a figure (Zhou, Friedman, & von der Heydt, 2000). This side-of-figure selectivity is known as border-ownership (see Figure 3-1a). For example, when the receptive field of a border-ownership cell (B cell; Craft, Schutze, Niebur, & von der Heydt, 2007) is centered on a vertical edge of a square, a stronger response (Figure 3-1d) may be elicited when the square is located to the left (Figure 3-1b) as compared to the right of the edge (Figure 3-1c), although the local contrast in the cell’s receptive field remains the same. Such side-of-figure selectivity could indicate a neurophysiological correlate of the percept that a border is owned by either the region to one side or another of that border, but not both. Figure 3-1a shows a well-known bistable display from Shepard (1990) in which the percept of figure alternates between a saxophone player and a female face. The central black-white borders is said to ‘belong to’ or ‘be owned by’ whichever portion of the display is perceived as the figure. B cells have been shown to exhibit modulation due to bistable visual scenes.

Although B cells in monkey V2 may demonstrate a border-ownership preference to a light square (Figure 3-1b) when the border intersects the cell’s receptive field and the light square is positioned to the left compared to a dark square (Figure 3-1c) positioned to the right of the cell’s receptive field (Figure 3-1d, light gray curve), the presence of a light square to the left and a dark transparent overlay to the right (Figure 3-1e) diminishes the border-ownership response (Figure 3-1g, light gray curve). Relative to this particular B cell whose receptive field is indicated by the ellipse, the figure is located to the right. When a border of a dark square appears within the receptive field of the same cell, the
dark square is located to the right of the receptive field, and a light transparent overlay to the left (Figure 3-1f), the cell exhibits an increased response (Figure 3-1g, dark gray curve). This particular B cell prefers figures whose borders locally intersect the cell’s receptive field and are located to the left of its receptive field center.

B cells demonstrate selectivity to figures in the global scene context far outside the cells’ classical receptive fields. By virtue of their prevalence in early primate visual areas, B cells have small receptive field sizes both compared to many other visual cortical areas and the figures to which they respond. For example, Zhou et al. (2000) report that median B cell receptive field sizes at foveal eccentricities are 0.5°, 0.7°, and 3.6° in monkey visual areas V1, V2, and V4, respectively. Despite their small receptive field sizes, B cells respond to a consistent side-of-figure irrespective of the figure size as long as it still perceptually appears as a figure, a property known as size invariance. Evidence also indicates that an inter-cortical network that spans V1, V2, and V4 allows B cells to access global information about the figures in the visual scene irrespective of potentially conflicting local information, such as motion (von der Heydt, Qiu, & He, 2003), luminance (Zhou et al., 2000; Qiu & von der Heydt, 2007; Zhang & von der Heydt, 2010), or disparity (von der Heydt, Zhou, & Friedman, 2000). That is, B cells can register the appropriate side-of-figure response, despite their small receptive field size and potentially ambiguous local information. How B cells do this — which must surely involve interactions with cells in other higher cortical areas and with larger receptive fields — is the point of our model.

The time it takes for border-ownership signals to emerge constrains the type of inter-cortical network that can plausibly perform figure-ground segregation. Researchers
have proposed that B cells access global information either *intra-areally*, i.e., by lateral connections within a single visual cortical area, such as V2 (Zhaoping, 2005), or *inter-areally*, i.e., where cells with larger receptive fields communicate contextual information about the scene via feedback projections to visual areas with small receptive field cells fewer synapses away from the retina (Angelucci et al., 2002). Intra-areal and inter-areal axonal conduction velocities have been estimated to be 0.3 m/s (Nowak, Munk, Girard, & Bullier, 1995; Nowak & Bullier, 1997; Angelucci et al., 2002) and 3.5 m/s (Girard, Hupe, & Bullier, 2001) in early visual areas, respectively (Bullier, 2001). Hence, inter-areal connections *can be an order of magnitude faster* than intra-areal connections for propagating information across the visual field. Sugihara et al. (2003) showed that B cell responses to 3° squares did not differ in latency compared to those to a 8° square, which is consistent with the use of inter-areal connections, but not intra-areal connections, to propagate contextual figure-ground information. Although a variable amount of time is required to propagate information about a figure within a single cortical area, transmitting the information to another area with large receptive field cells could afford a roughly fixed delay irrespective of the figure size in the visual field. Using published neuroanatomical data (Gattass, Gross, & Sandell, 1981; Gattass, Sousa, Mishkin, & Ungerleider, 1997), we estimate the cortical distance along the horizontal meridian and spanning 0°–5° eccentricity within V2 as 22.25 mm, which agrees with prior estimates (Craft et al., 2007). Traversing such a distance at 0.3 m/s would take approximately 75 msec, which cannot account for border-ownership latency of 10–25 msec reported in neurophysiology (Zhou et al., 2000). Hence, *it appears that connections within a single cortical area alone could not plausibly account for the fast global scene integration that*
is observed in B cell border-ownership responses (but see Zhaoping, 2005, who argues otherwise as we describe in the Discussion section).

The visual system connects meaningful properties in the world to mechanisms in cortex. In nature discrete objects that primates interact with tend to be convex. Humans are more likely to interpret convex regions as figure compared to those that are concave, irrespective of texture, color, and other low-level characteristics (Kanizsa & Gerbino, 1976; Peterson & Salvagio, 2008). One might expect that the natural convexity bias for figures has been mapped directly onto visual cortices. Cells throughout the visual system demonstrate on-center/off-surround selectivity and variants thereof, including off-center/on-surround. V4 neurons exhibit sensitivity to parametrically defined convex and concave curves of various orientations, acutenesses, and partial occlusion of figures (Bushnell, Harding, Kosai, & Pasupathy, 2011). Aside from possessing receptive fields larger than those at equivalent eccentricities in V1, V2, and V3, large numbers of V4 cells appear to have curved, radially symmetric, ‘on-surround’ receptive fields (Hegde & Van Essen, 2006). With an appropriate receptive field size, neurons with on-surround receptive fields are capable of detecting conjunctions of curved contours. Depending on the number and alignment of the detected contours, the co-occurrence of multiple curved contours within the on-surround region of a receptive field may be important for forming partial shape representations. The interactions between neurons that associate multiple curved contours may be instrumental for the generation of more complete object representations. Neurons in V4 with large, ‘donut-shaped’, on-surround receptive field organization can respond to the co-occurrence of curved contours that fall within the donut, which suggests the neurons are sensitive to Gestalt properties such as convexity.
and closure (Pizer, Burbeck, Coggins, Fritsch, & Morse, 1992; Pizer, Eberly, Morse, & Fritsch, 1998). V4 cells may communicate convexity information about figures in the visual scene with B cells in V2 via fast inter-areal connections (Craft et al., 2007).

The goal of the present paper is to demonstrate that a simple model of neural competition between convexity-sensitive units with different receptive field sizes in model V2 and V4 connected by fast inter-areal fibers can account for a number of the properties of B cells reported by neurophysiological studies to date (Zhou et al., 2000; Qiu & von der Heydt, 2007). A successful model of border-ownership and figure-ground segregation in the primate visual system should satisfy the following constraints. 1) Model B cells should demonstrate side-of-figure selectivities toward regions in a visual image that are perceived by humans as figure, which is consistent with known neurophysiological border-ownership signals. For example, in the case of a square occluding a rectangle (Figure 3-2e), model B cells with receptive fields centered around the square’s perimeter should elicit stronger border-ownership toward the square than the rectangle (Zhou et al., 2000). 2) Cortical networks that propagate global contextual information to small receptive field B cell units should bias border-ownership signals toward the inside rather than the outside of figures. Convex (on-surround) receptive fields are essential characteristics of cells extracting information about objects, figures, and surfaces in the environment. However, detecting convexity alone is not sufficient to determine figure-ground relations in a visual scene. The model should not indicate side-of-figure preferences toward concave regions (Figure 3-2f) that are not perceived by humans as figure (Zhou et al., 2000). 3) The model should demonstrate the B cell size invariance property and have a mechanism to determine border-ownership assignment
despite differences in the relative size between the figure and ground. In other words, how does the visual system identify figures that can appear at a wide range of spatial scales? As tested in the present model, we hypothesize that feedback from ‘on-surround’ model units with different receptive field sizes that undergo normalized competition propagates information about figure locations to B cells irrespective of the particular figure sizes. Our model, the R cell, G cell, B cell (RGB) model, predicts that a substantial portion of the large number of cells in primate V2 and V4 with ‘on-surround’ receptive fields (Pasupathy & Connor, 1999; Hegde & Van Essen, 2004, 2006; Bushnell et al., 2011) compete and group convexity information inter-areally to rapidly resolve figure-ground segregation. Akin to the bistable perceptual figure-ground reversals observed in displays such as Figure 3-1a, the RGB model predicts similar reversal phenomena occur at a local level within the early visual system due to competition between cells with ‘on-surround’ receptive fields. The dynamics of on-surround competition (Figure 3-4b) between units in different cortical areas have not been extensively studied, and we suggest on-surround competition is an integral part of figure-ground segregation in the primate visual system.

3.1.1 Visual Displays
Figure 3-2: Visual displays that are used in model simulations. Luminance junctions (a-d) represent important tests for the model because similar local junctions appear so frequently within natural and synthetic visual scenes. Visual displays that have been tested on B cells in the electrophysiological literature afford the comparison between model and cell responses (c, e, f). The model makes B cell response predictions in displays that have only been tested psychophysically for border-ownership (g, h). (a) A T-junction that is frequently associated with the occlusion of surfaces (left half; T-junction ‘stem’) by another surface (right half; T-junction ‘hat’) and the border typically is owned by the occlude. (b) L-junctions do not necessarily implicate occlusion, as the junction may appear at a corner of a figure. (c) The presence of X-junctions that reverse contrast polarity once may elicit the percept of a transparent surface (top left square) occluding another (bottom right square). The borders of the small centrally located square are owned by the transparent occluding surface. (d) When X-junctions reverse in contrast polarity twice, the percept of occlusion vanishes, and the borders of the centrally located small square may either be owned by the square or the surrounding L-shapes. (e) A square occludes a rectangle and contains two T-junctions. B cells recorded from in vivo signal border-ownership of the occluding square near the T-junctions (Zhou et al., 2000). (f) C-shape display that contains a concavity. B cells demonstrate side-of-figure preferences to the C-shape, and not to the concave region (Zhou et al., 2000). (g) Convexity display of Peterson & Salvagio (2008). Human subjects are more likely to indicate that the convex region is the figure than the ground compared to the concave region. (h) Kanizsa square. When the pacmen inducers are appropriately aligned, an illusory square is seen in the center that is a brighter white than that on the periphery.

Humans responses in psychophysical studies show that T-junctions, which mark the confluence of three luminance values (Figure 3-2a) in natural and synthetic scenes, indicate the presence of occlusion, but the confluence of two luminance values at L-
junctions (Figure 3-2b) do not (McDermott, 2004). Because local junctions appear so ubiquitously in visual scenes and their border-ownership cell response properties are known, these junctions provide important tests for any model of border-ownership. The model should yield border-ownership signals consistent with those obtained in neurophysiological and psychophysical studies. In Figure 3-2a, the left and right half regions are associated with the occluded and occluding figures, respectively. The central vertical border is owned by the figure on the right half side. For a convex figure, the L-junction in Figure 3-2b would be owned by the figure attached to the bottom-right region. Moreover, X-junctions represent local luminance constellations whereby four luminance values converge at a point and may coincide with the presence of a transparent occluding surface (Figure 3-2c). In Figure 3-2c, the borders of the central small square region are owned by the top left, not the bottom right, L-shaped region and together the small central square and the top left L-shape form the percept of a transparent filter. Studies have identified heuristic rules about the contrast polarity relations around the junction that generally give rise to the percept of a transparent surface (see Adelson & Anandan, 1990 and Anderson, 1997 for a review). When the contrast polarity at the X-junction reverses twice compared to once (Figure 3-2c), the percept of transparency is abolished (Figure 3-2d). In this case, the borders of the small central square may either by owned by the surrounding L-shapes or the square.

Displays that have been extensively tested in neurophysiological studies also represent important tests for any model of border-ownership. Figure 3-2e shows the occlusion of a rectangle by a square. Even at the two T-junctions, B cell responses indicate border-ownership by the square and not the rectangle (Zhou et al., 2000). The
gray C-shape in Figure 3-2f provides another test for border-ownership models because B cells provide border-ownership signals toward the C-shape and not toward the concave region to the right of the C.

Although the convexity displays of Peterson & Salvagio (2008) and the Kanizsa square have not been tested neurophysiologically on B cells, border-ownership properties have been identified from human psychophysical studies. Testing the model on these displays will check its psychophysical consistency and predict border-ownership cell responses to inform future experiments. Consider a display (Figure 3-2g) that contains convex and concave segments (Peterson & Salvagio, 2008). Humans exhibit a bias to report convex regions as the figure more often than concave ones, even when the convex region area equals that of the concave region. Non-rectangular displays that have curvature and do not have T-junctions may pose a challenge to figure-ground segregation models. Although we are not aware of any direct evidence on neural border-ownership data of curved displays or a border-ownership bias for convex or concave shapes, indirect evidence suggests this may be the case. First, B cells have been shown to produce border-ownership signals in the C-shape displays toward the C-shape and not the concave region (Zhou et al., 2000). Second, when a square region is defined by random dot motion and luminance contrast with its surrounding, the square region may be interpreted as an ‘object’ or a ‘window’. Over 80% of B cells in the sample that had a side-of-figure selectivity to the convex square when it appeared as figure (‘object’) compared to less than 20% when it appeared as the surrounding concave region (‘window’) appeared as figure (von der Heydt et al., 2003). Therefore, we predict that B cells are more likely to elicit side-of-figure preferences to the convex rather than the concave regions of the
convexity displays of Peterson & Salvagio (2008) irrespective of region area differences, which is consistent with human psychophysical judgments.

In the Kanizsa square display, which is formed by four ‘pacmen’ inducers, human subjects see an illusory square in the center that appears brighter than the white luminance to the periphery (Figure 3-2h). The Kanizsa square is an important test for border-ownership because based on B cell responses to individual convex shapes, such as squares, it is expected that B cells would indicate ownership toward the individual pacmen inducers. We hypothesize that B cells with receptive fields centered along the concave borders of the pacmen demonstrate border-ownership toward the illusory square when the inducers are aligned as in Figure 3-2h.

3.2 Methods

3.2.1 Model Overview
**Figure 3-3:** Schematic depiction of the model response to a square visual display. Border-ownership cells (designated by ellipses) with different side-of-figure preferences become active due to the presence of a bottom-up edge signal. Due to the lack of global information about the visual scene, all B cells at each spatial location initially are equally active and do not show a direction of border-ownership bias. B cells interact with small (G1, red) and large (G2, blue dashed lines) spatial scale grouping cells (G cells) with annular receptive fields, sensitive to convexity and closure. Note, connections for the larger scale (blue) are only shown. While G cells receive feedforward input from B cells within the receptive field, G cells also feedback to inhibit B cells with side-of-figure preferences away from the center of the annulus. The selective feedback biases the border-ownership signals toward the center of the annulus by suppressing the activity of B cells with inconsistent side-of-figure preferences. R cells pool over small (R1, red) and large (R2, blue dashed lines) spatial scale G cells and compete (black solid line) across scale to resolve salient scales in the visual scene. Following the competition, R cells feedback (blue solid lines) to B cells to inhibit B cells with side-of-figure preferences away from the peak R cell activity, which produces a bias in the border-ownership signals toward the figure. Note, for visual clarity not all connections in the model are shown.

Figure 3 shows the architecture of the RGB model. B and grouping (G) cells, as reported by Craft et al., 2007, cannot alone account for the properties of border-ownership cells reported in neurophysiological studies, because convexity sensitive G cells in V2 respond to regions that may be correlated with, but often do not represent figures or surfaces in the visual scene. Because grouping cells respond strongly to convex regions, a C-shape, as shown in Figure 3-2f, elicits high G cell activity outside of the ‘C’ due to the exterior convexity, but the ‘C’ is nonetheless interpreted as the figure. How does the cortex differentiate between regions of convexity that may or may not be associated with a figure? In addition to using well-known LGN and V1 complex cell units, border-ownership (B) cells in V1/V2, and grouping (G) cells in V2, as described by Craft et al. (2007), another unit type is required, called the R cell in our model, that reflects known properties of ‘shape’-sensitive cells in primate V4 (Hegde & Van Essen, 2006; Bushnell et al., 2011). The RGB model predicts that competition between R cells with different receptive field sizes plays a fundamental role in figure-ground segregation in concert with B cells by identifying candidate locations for figures in the visual scene.
Inter-areal connectivity between B and R cells is required because of their differing functional properties: R cells, with large receptive fields, can detect which side of an edge the figure is on, but cannot determine the precise location of the boundary; B cells, with small receptive fields, can locate a boundary, but cannot alone determine which side of an edge the figure is on.

**Figure 3-4.** (a) Border-ownership (B) cells in visual cortex preferentially respond to borders when they represent a certain side of a figure. Grouping (G) cells have an ‘on-surround’ or ring-like receptive field structure (yellow-green) and respond to convexity and bias the competition between two overlapping and similarly oriented but opposite-pointing B cell units (see legend). For example, G cells with appropriately sized receptive fields respond preferentially to a square (panel a; outline of filled square shown) and B cells with border-ownership preferences toward the center of the square are enhanced relative to those with preferences away from the square center. Brighter shades of green indicate stronger connection weights between border-ownership cells (denoted by ‘B’) and the G cell. (b) G cells bias border-ownership direction by inhibiting B cells with side-of-figure selectivities that point away from the radial center of their receptive fields.

### 3.2.2 B and G cells

Figure 3-4b summarizes the B-G cell microcircuit. This component of the model is similar to the model of Craft et al. (2007), except the RGB model does not have
junction detectors, and we analyze the temporal dynamics of B and G cells. Due to the prevalence of right angles and rectangular objects in the visual displays examined by neurophysiological studies of border-ownership (Zhou et al., 2000; Qiu & von der Heydt, 2007), we simulate four different border-ownership cells that sample each visuotopic location: left, right, up, and down. Graphically, we depict B cells by arrows, whose length indicates the magnitude of response and whose direction indicates the net direction of border-ownership at that visuotopic location (Figure 3-4a). Directions are determined by a vector, denoted vectorial modulation index ($V_{mod}$; Craft et al., 2007; Mihalas, Dong, von der Heydt, & Niebur, 2011) at each visuotopic location: the $x$ and $y$ components are a difference divided by a sum of the activities of left/right and up/down B cell activations. (See Model Equations for details.)

Blue and red connections in Figure 3-4b signify excitatory feedforward input to and inhibitory feedback from G cells, respectively. By virtue of the radial symmetry and ring-like receptive field shape, each G cell receives input from B cells whose receptive field centers are located at a distance relative to the G cell receptive field center. G cells thus perform ‘on-surround’ integration of their inputs over certain locations in the visual field. Larger receptive field G cells integrate a larger number of B cells over a wider area in the visual input. With an appropriate receptive field size, G cells can elicit strong responses when figure borders enter their on-surround receptive fields and communicate that information via feedback to border-ownership cells that also have the figure border in their receptive fields. Figure 3-4b schematically shows how a G cell with the square borders in its receptive field can feed global convexity information back to B cells with a border of the square in their receptive fields to reinforce the presence of the square figure.
Note, Figure 3-4b shows that the feedback connections from a G cell to its associated B cells are inhibitory, otherwise an unstable positive feedback loop could occur. We follow the convention of Craft et al. (2007) that G cells monosynaptically integrate spatially offset B cells within their on-surround receptive fields with a preference of ownership toward the center of the ring, and feed back to inhibit active B cells with the opposite ownership preference. The RGB model reinforces the presence of a convex figure by having G cells *inhibit a subset of B cells that supply it input*, but only those with the opposite ownership preference. B cells with active bottom-up input provide a local estimate of the relative figure position.

**Figure 3-5.** Model R cells produce activity peaks at the location of an occluding surface. Panels (a) and (b) focus on the dynamics of R cells with receptive fields located on the ‘stem’ and ‘hat’ sides of T-junction, respectively. G cells in panel (a) with receptive fields near the T-junction on the ‘stem’ side are highly active (bright green disks). The one-dimensional cross-section of G cell activity positioned above the model input shows the expected G cell responses at different locations along the ‘stem’ side. R cells perform on-surround integration of G cell units. Before R cell competition occurs in the model, R cells that receive inter-areal projections from the highly (moderately) active G cells will also be highly (moderately) active. The expected R cell response before competition
occurs is shown at the top. R cells with receptive fields centered on the same visuotopic location compete in shunting manner: the higher the unit activation, the more it inhibits other units. Due to the concentration of highly and moderately active R cells on the ‘stem’ side, competition will be fierce, and cell responses will drop precipitously following the competition. Conversely, G cells in panel (b) with receptive fields on the ‘hat’ side of the T-junction indicated by the oval overlaid on the input, are only weakly active, as indicated in the one-dimensional cross-section of G cell activity shown above the input. Before R cell competition occurs in the model, R cells that receive projections from the weakly active G cells will also be weakly active. The expected R cell response before competition occurs is shown at the top. Because R cell activity is weak on the ‘hat’ side, the competition is less fierce and the activity following competition is higher on this side compared to the ‘stem’ side. R cells indicate via feedback projections to B cells that the ‘hat’ side of the T-junction is the occluding figure. Note that at the top of the diagram only small-scale activity is shown, but the large-scale will begin with similar peaks.

3.2.3 R cells

R cell units compete in a shunting fashion—*R cells receive inhibition from other R cells with different receptive field sizes that are centered at the same visuotopic location proportional to their current activation* (Grossberg, 1973). Shunting interactions are well known to occur between cells in visual cortex and have *canonical, universal properties, such as divisive normalization present in many places throughout the cortex* (Carandini & Heeger, 2011). R cells propagate global information about the location of salient figures in the visual scene inter-areally to B cells. As dynamics unfold, reversals in the polarity of border-ownership signals occur toward R cell receptive field centers (left vs. right, up vs. down).

Figure 3-5 shows how R cells use shunting competitive dynamics between units with different receptive field sizes to produce peaks in the location of the occluding surface in a T-junction and suppress activity in the location of the occluded surface. T-junctions mark the confluence of three luminance values, and the regions on the “hat” and “stem” sides of the junction are associated with the *occluding and occluded* surfaces,
respectively. Figure 3-5a focuses on the dynamics of G and R cells with receptive fields on the side of the occluded surface (stem). Because G cells respond to convexity, they elicit high activation on the stem side (bright green circles superimposed on input). As shown above the input in the one-dimensional cross-section of the G cell activity, the strong response to the corners is concentrated in certain visuotopic locations: G cells with receptive fields far away from the T-junction on the stem side will elicit a weak response (blue). R cells with larger receptive fields integrate G cell activity in an on-surround manner. The R cell that monosynaptically receives inter-areal projections from both highly active G cells (bright green) will also be highly active before the competition occurs. Conversely, the R cell with a larger receptive field that misses the highly active G cells receives projections from moderately active (green) G cells and will therefore be moderately active. The R cells shown will inhibit one another proportional to their present level of activity due to their shunting interactions. Because both units are at least moderately active, the activity of both units will be strongly suppressed relative to their initial levels after the competition occurs (top of Figure 3-5a).

Figure 3-5b illustrates on the dynamics of G and R cells whose receptive fields are on the occluding surface side (hat). Because G cells prefer convexity and do not receive information about corners as they did on the occluded surface side, the G cell response will uniformly be much weaker on this side (green ellipse superimposed on the input). As shown above the input in the one-dimensional cross-section of the G cell activity, G cells still receive contrast information from the line, so the G cells whose receptive fields intersect with the line from the occluding surface side will be moderately active. The R cell that integrates the moderate G activity will also be moderately active
before competition (green). Conversely, the R cell with a larger receptive field size does not integrate the moderate G cell activity and produces a weak response (blue). As in Figure 3-5a, R cells inhibit each other proportional to their current level of activity. Because the R cells shown are only weakly or moderately active initially, the inhibition they receive from one another is weak. After the competition occurs, their responses will decrease, but not as much as the R cells shown in Figure 3-5a. Hence, *R cells reverse border-ownership polarity (left vs. right, up vs. down) at a T-junction and respond more strongly to the occluding surface:* two strong G cell activity peaks are annihilated by R cells on the ‘stem’ side and one shallow peak on the ‘hat’ side survives. The surviving peak on the occluding surface is conveyed to B cells via inter-areal projections to suppress the response of B cells that have preferred directions away from the occluding surface.

### 3.2.4 Model Equations

Differential equations in the model specify the activity of cells or cell populations with receptive fields centered on each pixel of the input displays. Since the operations within each equation apply to all cells, we use matrix notation. For example, \( \mathbf{x} \) stands for the set of cells at every spatial location \((p, q)\) in the input display. Convolution between a matrix \( \mathbf{x} \) and kernel \( \mathbf{F} \), specified by the \( * \) operator, is always centered at each cell position \((p, q)\). In all convolutions, we tile the boundary values beyond the image border as far as necessary. All ordinary differential equations were numerically integrated using a Runge-Kutta routine. Our simulations were performed on a 2.66 Ghz 8-core Apple Mac Pro with 64 GB RAM in Wolfram Mathematica 8.
Model neurons are represented by a single-compartment voltage $V(t)$ that obeys the following shunting equation:

$$C_m \frac{dV(t)}{dt} = -(V(t) - E_{\text{leak}}) \delta_{\text{leak}} - (V(t) - E_{\text{excite}}) \delta_{\text{excite}}(t) - (V(t) - E_{\text{inhib}}) \delta_{\text{inhib}}(t)$$

(1)

In Eq. 1, $C_m$ specifies the membrane capacitance, $\delta_{\text{leak}}$ denotes the constant leakage rate, $\delta_{\text{excite}}(t)$ signifies the lumped excitatory inputs to the cell over time, and $\delta_{\text{inhib}}(t)$ specifies lumped inhibitory inputs to the cell over time. The terms $E_{\text{leak}}$, $E_{\text{excite}}$, and $E_{\text{inhib}}$ refer to the cell’s leak, excitatory, and inhibitory reversal potentials.

3.2.5 Model LGN Cells

The first stage of the model consists of isotropic on-center/off-surround processing. Eq. 1 can be rewritten by setting $x = V$, $\alpha_{\text{LGN}} = \delta_{\text{leak}}$, $E_{\text{leak}} = 0$, $\beta_{\text{LGN}} = E_{\text{excite}}$, and $\gamma_{\text{LGN}} = -E_{\text{inhib}}$:

$$\frac{dx}{dt} = -\alpha_{\text{LGN}} x + (\beta_{\text{LGN}} - x) F_{\text{excit}}^{\text{LGN}} * I - (x + \gamma_{\text{LGN}}) F_{\text{inhib}}^{\text{LGN}} * I$$

(2)

In Eq. 2, $x$ signifies the activation of model LGN neurons with receptive fields centered at each pixel in the input display at time $t$, $\alpha_{\text{LGN}}$ represents the passive decay of the model cell, $\beta_{\text{LGN}}$ specifies the saturation upper bound, $\gamma_{\text{LGN}}$ is the inhibitory lower bound, and $I$ refers to the input image. Both $F_{\text{excit}}^{\text{LGN}}$ and $F_{\text{inhib}}^{\text{LGN}}$ are Gaussian kernels with $\sigma_{\text{excit}}^{\text{LGN}} = 0.25$ and $\sigma_{\text{inhib}}^{\text{LGN}} = 0.5$. In our simulations, we set $\alpha_{\text{LGN}} = 5$, $\beta_{\text{LGN}} = 1$ and $\gamma_{\text{LGN}} = 0.5$. To interpret Eq. 2 as cell or population firing activity, we half-wave rectify the output (i.e. consider the non-negative component of the response):
3.2.6 Model Complex Cells

We construct complex cell units by summing the responses garnered by pairs of simple cell units that differ in orientation preference by π radians. We fix \( \theta \in \{0, \frac{\pi}{2}\} \). Simple cell units also possess orientation preferences, but are sensitive to the polarity of contrast. According to the Steering Theorem (Freeman & Adelson, 1991) we can express a simple cell kernel of arbitrary orientation preference \( S_\theta \) via a weighted sum of the partial derivatives of a 2D Gaussian kernel \( F^{V1} \) with respect to the \( x \) and \( y \) directions.

\[
S_\theta = \cos(\theta) S_0 + \sin(\theta) S_\pi
\]

In Eq. 4, the notation \( S_0 \) means the first derivative of kernel \( F^{V1} \) in the direction of \( \theta \), rotated such that the kernel has a preferred orientation \( \theta \) radians. Since \( S_0 \) and \( S_\pi \) are orthogonal, they form a basis. Hence, we may interpret Eq. 4 as basis elements rotated such to yield a simple cell kernel, which has a preferred orientation \( \theta \) radians. In our simulations, we chose \( \theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\} \). Complex cell units \( C_\theta \) can be obtained by summing anti-phase kernels:

\[
C_\theta = S_\theta + S_{\theta+\pi}
\]

Because we are simply concerned with horizontal and vertical directions of contrast, we fix \( \theta \in \{0, \frac{\pi}{2}\} \). We set \( \sigma^{V1} = 0.5 \). To compute the complex cell \( Y \) unit activity in model V1, we convolve the LGN unit responses with the horizontal and
vertical complex cell kernels $C_\theta$, threshold by $\Gamma$, and half-wave rectify in the following additive equation:

$$\frac{dY_\theta}{dt} = -Y_\theta + \max(C_\theta * X - \Gamma, 0)$$  \hfill (6)

3.2.7 Model Border-Ownership Cells (B cells)

We model B cells, with border-ownership direction $\theta$, using the following equation that integrates bottom-up input from complex cells and top-down signals from G and R units (described below):

$$\tau^B \frac{dB_\theta}{dt} = -B_\theta + (\beta^B - B_\theta)Y_\theta$$

$$- \gamma^B B_\theta \left( \sum_s \zeta^s (G^s * K^s_{\theta+\pi}) + \sum_l \eta^l (R^l_2 * L^l_{\theta+\pi}) \right)$$  \hfill (7)

In Eq. 7, $G^s$ and $R^l_2$ signify the grouping cell (G cell) and second stage R cell activity at scale $s$ and $l$, respectively. $K^s_{\theta+\pi}$ and $L^l_{\theta+\pi}$ may be viewed as subunits of the G and R cell units’ receptive fields. We set the scale indices $s$ and $l$ of $K^s_{\theta+\pi}$ and $L^l_{\theta+\pi}$ such that the feedback from G and R cells projects to nearby B cells that receive bottom-up input from complex cells, respectively. $\zeta^s$ and $\eta^l$ serve to differentially weight feedback contributions to B cells due to scale-dependent receptive field differences. We perform a scale-dependent weighting by $\zeta^s = \eta^l = \sqrt{4.5r}$ similar to that of Craft et al. (2007), where $r$ is the radius of the G or R cell receptive field kernel. The parameter $\gamma^B$ regulates the gain of inhibitory G and R feedback signals. In Eq. 7, we set $\tau^B = 10$ ms, $\beta^B = 1$, $\gamma^B = 100$. At every spatial position, we model B cells with four directions of border-
3.2.8 Model grouping cells (G cells)

G cells possess radially symmetric annular or ‘ring-like’ receptive fields that integrate B cell activity. We achieve feedback between B and G units in Eq. 7 by convolving $G^s$ with either the left, top, right, or bottom piece of the G cell receptive field annulus $K_{\theta + \pi}^s$, which selects B cells in regions of the G cell receptive field and suppresses B cells with direction of border-ownership preference $\pi$ radians away from that pointing inward toward the annulus center.

We construct the G cell kernels by taking a difference-of-Gaussians (i.e. Mexican hat) between kernels $F_1^G$ and $F_2^G$ of radius $r^G$ and standard deviations $\sigma_1^G$ and $\sigma_2^G$, respectively. For the Gaussians $F_1^G$ and $F_2^G$, we constrain the ratios $\frac{r^G}{\sigma_1^G} = 2$ and $\frac{r^G}{\sigma_2^G} = 2.22$. For computational tractability and simplicity, we select two G cell receptive field sizes within the range reported in the neurophysiological literature. We set $r^{G_1}$ and $r^{G_2}$ to 2 and 3, respectively. In order to obtain the kernel fragments $K_{\theta}^s$ we simply extract the necessary half of the annulus. For example, to obtain $K_{\pi}^s$, we take the left half component of the G cell kernel at scale $s$.

In order to model the G cell dynamics, we employ an equation similar to that of Craft et al. (2007):

$$\tau^G \frac{dG^s}{dt} = -G^s + \xi^s(1 - G^s) \sum_{m,n} (B_{\theta_m} * K_{\theta_m}^s)(B_{\theta_n} * K_{\theta_n}^s)$$

(8)
In Eq. 8, we multiplicatively combine all pairs of directions of ownership permutations $m$ and $n$. For example, when $m = 1$ and $n = 2$, $\theta_m = 0$ and $\theta_n = \frac{\pi}{2}$, since $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. We fix $\tau^G = 10$ msec. Each combination may be interpreted as functional subunits of the G cell receptive field. The multiplicative subunit structure affords G cells a nonlinear response to regions with convexity and closure. Hence, G cells detect co-occurrences between pairs of edge signals with different orientations.

### 3.2.9 Model R cells

The R cell layer consists of stages of competition across scale and distance-dependent spatial competition. R cells pool over local populations of G cells that possess a common scale. We introduce a temporal competitive neural network to model R cells in model area V4 (Grossberg, 1973). In the following R cell network, we select a faster-than-linear signal function $f(w) = w^2$ to form a winner-take-all network.

\[
\tau^{R_1} \frac{dR_1^l}{dt} = -R_1^l + \left( \beta^{R_1} - R_1^l \right) \left( f(R_1^l) + \delta^l(G^s \ast F_{excit}^{l,R_1}) \right) - R_1^l \sum_{k \neq l} F_{inhb}^{R_1} \ast f(R_1^k) \]

In our simulations, we let $\tau^{R_1} = 10$ msec and $\beta^{R_1} = 2$. By having R units that perform inter-scale competition, the network selects the spatial locations of salient figures at each scale. As with G cell spatial scales, we employ two R cell receptive field sizes. For the purposes of our simulations, we assume G and R cells each possess two scales. We assume $R_1^l$, which represents the set of R cells of scale $l$, may only receive projections from $G^s$, which represents the set of G cells of scale $s$, when $s = l$. In all
simulations, we set $r^l \in \{2, 3\}$. $\delta^l$ specifies a scale-dependent proportionality constant. We set in all simulations $\delta^1$ and $\delta^2$ to 1200 and 350, respectively. We produced the ring-shaped kernel $F^{l,R_1}$ by subtracting a 2D disk kernel of radius $r$ with another of radius $r/2$. The radius 3 disc kernel $F^{R_1}$ specifies which proximal R cells with different receptive field sizes enter in the inter-scale competition.

Finally, we have a second R cell stage that performs local spatial competition to enhance the contrast in the network responses, which is useful for high spatial frequency displays:

$$
\tau^{R_2} \frac{dR^l_2}{dt} = -R^l_2 + A(\beta^{R_2} - R^l_2)(f(R^l_2) + R^l_1) - (R^l_2)(f(R^l_1 * F^{l,R_2}_{inhb}))
$$

In Eq. 10, we set $\tau^{R_2} = 10$ msec and $\beta^{R_2} = 1$. Eq. 10 represents a choice network as well with $f(w) = w^2$. $A$ specifies the scale-specific attentional signal, which we model with a broad 2D Gaussian, spatially centered at the locus of attention. The attentional signal multiplicatively enhances the R cell signal at a particular spatial scale. Hence, it only enhances existing R cell activity. $A$ was set to 1 in all simulations except for in the double-reversal transparency display (Figure 3-7e) to differentially weight small and large R cell peak activities. Figure 3-7f (left) shows the model output when $A = 1$ and Figure 3-7f (right) shows the model output when $A$ is changed to amplify the activity peaks shown in Figure 3-7e (right). $F^{l,R_2}$ represents an annular kernel with diameter $3^\circ$. R units feedback to B cells (Eq. 7) to suppress those with anti-preferred directions of border-ownership.

3.1.10 Vectorial Modulation Index
We use the following vector modulation index $\mathbf{v} = (m_x, m_y)$ to relate our B cell responses to those reported in other studies (Craft et al., 2007; Mihalas et al., 2011):

$$m_x = \frac{B_0 - B_\pi}{B_0 + B_\pi}$$

$$m_y = \frac{B_\frac{\pi}{2} - B_\frac{3\pi}{2}}{B_\frac{\pi}{2} + B_\frac{3\pi}{2}}$$

(11)

Due to the normalizing difference over a sum of B cell activity with anti-preferred directions of border-ownership, each component lies between -1 and 1. Negative values correspond to computed border-ownership in the leftward and downward directions for $(x, y)$ components at each spatial location $(p, q)$, respectively. By contrast, positive values correspond to computed border-ownership in the rightward and upward directions for $(x, y)$ components at spatial location $(p, q)$, respectively. A zero valued component indicates an indifference in border-ownership in that particular axis. Since Zhou et al. (2000) discovered border-ownership modulation emerging 10-25 msec from the onset of the visual presentation, we ran the model for 25 msec in model time.

**Figure 3-6.** T Junction simulation. The V1 complex cell response to the T-junction
display is shown in (a). The top (b-c) and bottom (d-g) rows show simulation results with and without V4 R cells, respectively. Without feedback from R cells, the vectorial modulation index (Vmod) of B cells points toward the stem side of the T-junction, rather than to the occluding surface (b). This is due to the high G cell response on either side of the T-junction stem, indicated by the dashed white contour (c and e). The solid white lines are drawn to indicate the contrast boundaries present in the T-junction display. R cells perform on-surround competition across scale (Stage 1) and respond with high activity on the side of the T-junction hat (the occluding surface) due to shunting inhibition in the network competition, indicated by the dashed white contour (f). A second stage of R cells performs local spatial competition to identify peaks from Stage 1 that may represent figure locations (g). With R cell feedback, the Vmod of B cells indicates the presence of a figure on the side of the T-junction hat (d), consistent with Zhou et al. (2000). The lengths of each vector component in (b) and (d) are proportional to the difference in activity between cells signaling border-ownership along that particular axis. 31 B cells (19 are shown in (b) and (d) for clarity) and 22 x 22 G and R cells were each simulated.

3.3 Results

3.3.1 T- and L-junctions

Figure 3-6 shows simulation results to a T-junction, which is an important test for the model, because model B cells should indicate border-ownership toward the occluding surface, as reported by Zhou et al. (2000). Figure 3-6a shows the vertical and horizontal complex cell activity that projects to B cells (Figure 3-6b and d). Figure 3-6c depicts the activity of G cells with different receptive field sizes. Consider the case wherein the model consists only on B and G cells (Figure 3-6a – c); that is, B cells do not receive feedback from R cells. Because the G cells prefer convexity and the geometry of a T-junction includes corners of contrast on the ‘stem’ (Figure 3-6a), the response of G cell is greatest inside the corners of the T-junction ‘stem’ (Figure 3-6c; region enclosed by the dashed white contours). Conversely, G cells respond poorly to T-junction ‘hat’ (indicated by weak green color), because G cells may at best make one point of contact with the complex cell signal. The G cells that respond well on the left side project to B cells that
respond to bottom-up contrast on the bottom left half of the vertical edge (Figure 3-6a) to enhance the activity of B cells with side-of-figure preferences toward the G cells’ receptive field locations. As shown in Figure 3-6b, the B-G interaction results in a preference toward the corners on the side of the ‘stem’. However, this region is not always consistent with the usual interpretation of occlusion. The B-G microcircuit alone cannot account for the figure-ground properties associated with the T-junction or the border-ownership response observed by Zhou et al. (2000) (see Figures 23 and 24). This means that with only B and G cells, G cells can respond outside of figures. Their responses are also biased by high contrast regions. Since G cells respond to any region of convexity at all spatial scales, G cells alone cannot convey the appropriate scale range within which an object may be perceived a figure.

Figures 6f – g show simulation responses of the activation of R cells. R cells with receptive fields centered at the same spatial location but with different receptive field sizes compete with each other, which results in the pattern of activity shown in Figure 3-6f. R cells are inhibited proportional to their present level of activation (shunting inhibition). A highly active R cell at a particular spatial scale receives strong inhibition when spatially proximal R cells with different receptive field sizes are also active. The result is a shift in the activity distribution of R cell units compared to that of G cells (compare activity peaks in Figure 3-6e and g): R cells that have the same receptive field location as a highly active G cell will have low activity and vice versa. R cells that have receptive fields on the side of the T ‘stem’ integrate two G cell peaks across scales and as a result encounter strong suppression. When an R cell feeds back to B cells to enhance units with side-of-figure preferences toward the location of the R cell, Figure 3-6d shows
B cells now show a net modulation in activity toward the side of the T-junction typically associated with the occluding surface. This reversal is consistent with the B cell responses reported neurophysiologically (see Zhou et al., 2000, Figures 23 and 24).

Although R cells exhibit an inversion with respect to G cell peak activity in the case of T-junctions, this does not occur for L-junctions (Figure 3-7a). For the sake of the present discussion, a T-junction can be thought as the combination of an L-junction and its reflection along an axis defined by the T-junction ‘stem’. As we would expect from the T-junction simulation, G cells respond most strongly inside the L-junction. Although R cells are engaged outside the L-junction, the activity is lower than that within. Without an adjacent corner, which would be present with a T-junction, the R cell on-surround shunting competition does not receive as much suppression as it would in the T-junction case. Hence, a shift in peak R cell activity does not occur, which is consistent with the tendency for the L-junction to contain a figure on the inside of its corner. The junction geometry is important because when four L-junctions are joined to form a square, the R cell peak inversion does not occur and R cells maintain the G cell peaks inside the square’s region. Hence, the mere adjacency of two L-junctions does not induce a R cell peak shift — their spatial relations matter. Junctions with different spatial configurations give rise to different sets of G and R cell dynamics, which induces border-ownership signals consistent with neurophysiological findings and human figure-ground perception.
Figure 3-7. The top and bottom rows show the model R and B cell responses, respectively, to a number of visual displays. Unlike the T-junction simulation in which the G and R cells respond maximally to different regions of the visual input, G and R cells both respond inside the L-junction contrast-defined corner (a). The presence of two strong G cell activity peaks in the T-junction simulation on the ‘stem’ side and a weak activity peak on the ‘hat’ side induced the peak R cell activity to shift to the hat side. The L-junction display only results in one distinct G cell peak, which does not produce enough inhibition to move the peak R cell activity to a different location (b). The response of R and G cells to similar locations extends to simple rectangular shapes. Panel (c) shows a configuration of X-junctions (defined in the text) that produces a percept of a transparent overlay on top of a square. R cells respond highest in the center of the transparent overlay and signal to B cells that it is the occluding surface (d). When X-junctions are arranged such that they do not support the percept of transparency (e), R cells yield three activity peaks of different magnitudes: the peak due to the center of the visual display (e, left) is the strongest, and the other two peaks near the corners of the two ‘L-shapes’ are of equal magnitudes (e, right). The left subpanel of (f) shows the B cell response, which favors the center region. If the activity peaks produced by the large receptive field R cells (e, left) are weighted higher than those yielded by the small receptive field R cells (e, right), the B cell response favors the ‘L-shapes’. Panel (g) shows the model R cell response to a square occluding a rectangle. Due to the presence of T-junctions, R cells respond on the interior of the occluding surface, and bias B cell responses (h) toward to square.

3.3.2 Transparency

Although the present model cannot fully explain perceptual transparency, the model correctly assigns border-ownership to the perceived figure in the case shown in Figure 3-7c. Figures 5c – f shows the model output to a transparency display with two X-junctions. When the contrast polarity reverses once about the X-junctions, the
configuration favors the percep of a lighter square serving as a transparent filter over an equally-sized darker square. Figure 3-7c show the peak R cell activity after spatial competition, which is concentrated close to where the top left corner of the centrally located medium gray square intersects with the lighter luminance ‘L-shape’. Feedback from the R cells to B cells generates modulation that encloses the square that perceptually appears as a transparent overlay (Figure 3-7d). The average magnitude of the B cells that enclose this region is higher than that of any other closed region within the visual display, such as the center medium gray square (Figure 3-7d). The nontrivial response obtained around the borders of the center square shows the model can simultaneously maintain more than one figure-ground segmentation of a visual scene.

Figure 3-7e contains two X-junctions that exhibit a double reversal in contrast polarity, which does not perceptually support transparency. The display is perceptually segmented into two dark gray ‘L-shapes’ and a centrally located light gray square. R cells yield three activity peaks of different magnitudes: the peak due to the center of the visual display (e, left, large spatial scale) is the strongest, and the other two peaks near the corners of the two ‘L-shapes’ are of equal magnitudes (e, right, small spatial scale). Figure 3-7f (right panel) shows that the chosen parameter configuration favors the segmentation of the dark ‘L-shapes’ as figure. If we give higher weight to the peak produced by the larger spatial scale R cells (e, left), which is centered on the square, compared to the peak produced by R cells with smaller receptive fields (e, right), the B cell responses reverse in their modulation along the borders of the ‘L-shapes’ and square to support the square as being the figure (Figure 3-7f, left panel).

3.3.3 Occlusion
We show in Figure 3-7g – h an occlusion display containing a square in front of a rectangle. As in Zhou et al. (2000), we set the visible rectangle area to match that of the square. In the absence of R cells, the B cells exhibit response modulation preferring the higher contrast areas. As predicted by our T-junction simulation (Figure 3-6g), the activity of R cells peak on the ‘hat’ side of the T-junction, and the B cells reverse their overall direction of modulation to prefer the light gray square.

**Figure 3-8.** C-shaped display simulation. By virtue of their annular receptive field structure, G cells respond maximally to the top and bottom portions of the convex C-shape and the concave region (b, indicated by dashed white ellipses). In the absence of R cells, some B cell vectorial modulation indices show preferred direction of ownership in the direction of the concavity (circled in red), which is not consistent with the cell data, indicating the C-shape as the figure (d). The distribution of G cell activity is similar to that produced in the T-junction case (see Figure 3-6): there are proximal regions in which G cells elicit high activation (indicated by dashed white ellipses) and a nearby region in which there is uniformly weak G cell activity inside the vertical part of the convex ‘C-shape’ (b). Competition between units with different receptive field sizes produces peak R cell activity within the C-shape (c), which biases the B cells to demonstrate vectorial modulation indices toward the C-shape interior (e).

### 3.3.4 Displays with concavities
Figure 3-8 shows the simulation results to a ‘C-shape’ display. The display is an important test for models because the shape is not convex and cells whose receptive fields are centered around the corners of the concave region demonstrate a border-ownership preference toward the ‘C-shape’ instead of the background (Zhou et al., 2000). Without R cells, the B cells mostly develop preferences consistent with the convex ‘C-shape’ appearing as figure, except for some B cells whose receptive fields are centered along the concavity corner (Figure 3-8d, see red circles). This is because G cells respond maximally to the interior of both the ‘C-shape’ (Figure 3-8b, see top and bottom dashed ellipses) and concave region (Figure 3-8b, see middle dashed ellipse). G cells with receptive fields centered along the interior of the ‘C-shape’ generally elicit weak activity, except for the cases indicated by the top and bottom dashed ellipses (Figure 3-8b). Note the similarity between the G cell activity distribution in the ‘C-shape’ and T-junction cases: there are proximal regions of high G cell activity (dashed ellipses) and an extent of weak activity on the interior of the ‘C-shape’. As in the T-junction simulation (Figure 3-6), R cells yield a peak shift relative to the G cell peak activity distribution, in this case from the peaks indicated by ellipses to where there was weak G cell activity on interior of the of the ‘C-shape’ (Figure 3-8c). Hence, B cells that respond inconsistently with the representation of the ‘C shape’ as figure (Figure 3-8d, indicated by red circles) in the absence of R cells, now exhibit consistent modulations (Figure 3-8e).
Figure 3-9. (a) Convexity display used in Peterson & Salvagio (2008). Human subjects were more likely to indicate the ‘convex’ region (right) as figure, as compared to the ‘concave’ region (left). (b) Stage 2 R cell responses. R cells respond to the convex, concave, and far-left regions, but exhibit peak activity to the convex region. We argue this reflects the human bias to indicate the convex region as figure compared to the concave region. Some R cells responded on the far left side too, because their receptive fields are on the convex side of the farthest left contour. (c) B cells indicate directions of border ownership toward the convex regions, consistent with human psychophysical data (Peterson & Salvagio (2008)). We downsampled (10x) the convexity display shown in (a) within our simulations and the luminance was inverted in (c) for visualization.

The convexity display of Peterson & Salvagio (2008) may be challenging for the model as well as others, because G units with appropriate receptive field sizes would respond to the interior and exterior cavities of the segments just as well. A solution has been proposed in the form of a Bayesian belief propagation network to bias border-ownership cells toward a “skeleton” along the medial axis of a shape (Feldman & Singh, 2006; Froyen, Feldman, & Singh 2010). It has also been shown that radial kernels can locate the medial axis of a shape (Pizer et al., 1998). The combination of radially symmetric R cell kernels and network dynamics in our model can also bias border-ownership toward the medial axis of a shape, as shown in Figure 3-9. Because G cells respond to convexity, units with appropriate receptive field sizes yield higher activities in the locally convex portions (*1) in the convex region than in the locally concave portions (*2) of the concave region. In addition, the sharp protrusions (*3) from the curved vertical contour toward the medial axis of the convex region elicit weaker G cell
responses than the convex protrusions into concave region (*4). When R cells with different receptive field sizes compete and undergo normalization across scale, similar to the T-junction case, R cell activity in the locally convex portions in the convex region (*1) is low because G cell activity is high. R cells in the convex region with receptive fields near the medial axis that integrate the sharp protrusions (*3) elicit higher activation than the concave region because G cell activity is low. The R cell dynamics bias B cells toward the convex region (Figure 3-9b), which is consistent with the human judgments found by Peterson and Salvagio (2008). The B cell bias toward the convex region is independent of the distance between the curved vertical contours, and rather is related to the relative locations of concave and convex turns around each vertical contour. The normalized inter-scale competition in the RGB model assists in identifying the location of a proximal figure candidate and a ‘best fit’ approximation to its size.

**Figure 3-10.** Irrespective of the shape of the figure present in the visual display and its contrast polarity, the model B cell demonstrates border-ownership of the figure when it is to the left of the receptive field. The model B cells (top right bar graph) exhibit the same relative mean firing rates as reported in neurophysiology (bottom right bar graph), when reflecting the figure position about the B cell’s receptive field, while maintaining the same local contrast. Consistent with the cell responses, for example, model B cells elicit greater activation when the square is positioned to the left (a) compared to the right (b) of the unit’s receptive field. Red ellipses designate the location of the B cell classical receptive field both in the simulation and in the experiments of Zhou et al. (2000).
Bottom bar graph adapted from Zhou et al. (2000).

In Figure 3-10, we compare the ordinal relationship between B cell unit responses to square, ‘C-shape’, and square occluder displays in our model (top right) and those shown in Figure 3-23 of Zhou et al. (2000) (bottom right). Irrespective of the shape of the figure present in the visual display and its contrast polarity, the model B cell, whose classical receptive field is indicated by the red ellipse, demonstrates border-ownership of the figure when it is to the left of the receptive field. At the spatial location indicated by the red ellipse, we take the mean activity of the most responsive B cell in time bins of 1 msec over 25 msec following the onset of the visual display. The model cells demonstrate the same relative mean firing rates as reported in neurophysiology when reflecting the figure position about the vertical axis defined by the center of the B cell’s receptive field, while preserving the same contrast that is present in the B cell’s receptive field (Craft et al., 2007). For example, when the ‘C-shape’ appears darker than the background (Figure 3-10c) and the B cell has a receptive field centered on the border, indicated by the red ellipse, the cell yields less mean normalized activity compared to when the ‘C-shape’ is lighter than the background and the ‘C-shape’ configuration is reflected about the vertical axis (Figure 3-10d).

**Figure 3-11.** Kanizsa square simulation. In the absence of R cells, B cells along the pacman inducer borders have vectorial modulation indices consistent with the percept that the inducers are figures (a). Through competition and feedback from R cells, B cells...
along the interior borders of pacmen reverse their vectorial modulation index directions to point toward the illusory square (b). Since progressively removing inducers diminishes the percept of the illusory square, we studied the B cell vectorial modulation indices as a function of number of inducers remaining. After removing an inducer (c), all B cells still exhibit vector directional components in the direction of the illusory square. With two inducers, 40% of B cell vectorial modulation indices have vector components toward the illusory square (d). Finally, the presence of only one inducer clearly favors it as the figure.

3.3.5 Kanizsa square

Finally, we examined the model behavior to the Kanizsa square display. In Figure 3-11a-b, we show the differences in B cell modulation in the presence and absence of R cells in the circuit, respectively, and inspect the modulation of border-ownership signals as inducers are progressively removed (Figure 3-11c-e). Without R cells, border-ownership signals favored the representation of the four individual inducers as figures (Figure 3-11a). The addition of R cells reversed the direction of B cell modulation for those located along the concave borders of the inducers. The vectorial modulation is toward the center of the illusory square (Figure 3-11b). When one inducer is removed (Figure 3-11c), most B cells have vectorial modulation indices that still are directed toward the illusory square. With only two inducers (Figure 3-11d), few B cells now exhibit vectorial modulation indices that are directed toward the illusory square, while most are directed toward the inducers. Only one inducer yields vectorial modulation indices that indicate that the inducer is the figure. The behavior of the model as inducers are removed is similar to that of the ‘C-shape’ (Figure 3-8), since it detects the salient figure despite the presence of local concavity.

3.4 Discussion
We have shown that the RGB model network of a subset of cells identified in primate V1, V2, and V4 connected by fast inter-areal connections gives rise to border-ownership selectivity in a variety of visual displays in a manner that is consistent with neurophysiological and psychophysical data. The model predicts that competition between convexity-sensitive cells with on-surround receptive fields acting in a network across primate visual areas V2 and V4 can explain the response of border-ownership cells in V2. B cells quickly provide vector information (i.e. a direction of ownership and a confidence measure in the strength of the response) that may prove useful in domains other than figure-ground segregation, such as the coordination of motor actions. For example, in situations whereby a rapid motor response is required to act on rapidly approaching objects, B cells could quickly via inter-areal connections provide enough information to motor cortex without fully analyzing the visual scene. Rapid development of vector information about figures could also contribute to the stability of visual perception in the presence of eye movements and moving objects (Ballard & Hayhoe, 2009). Our analysis focused on the resolution of border-ownership signals in visual scenes ‘at a glance’. Evidence exists that when the interpretation of the visual scene changes, border-ownership cell activity may take some time (~100 msec) to adjust to the new information about the figures (O’Herron & von der Heydt, 2009).

3.4.1 Are local visual junctions special?

The occlusion of one by another is a frequent occurrence in nature. B cells in neurophysiology demonstrate sensitivity to changes in perceived occluding figure when the properties of small local junctions are altered. For example, B cells show a diminished response when a dark gray rectangle occludes a light gray rectangle (Figure
3-10f), compared to when the light gray rectangle was in front (Figure 3-10e). We tested the model’s competency in resolving border-ownership in the presence of T-junctions. Without feedback from R cells in model V4 to B cells in model V2, border-ownership signal were always directed toward the T-junction corners on the occluded surface (stem side), because of the increased convexity compared to the occluding surface (hat side; Figure 3-6). We showed that R cells changed the border-ownership signals to point toward the occluding surface (hat side). This reversal occurs due to the model’s on-surround competition. Qiu and von der Heydt (2007) showed that in a transparency display, border-ownership signals reversed direction within 50–100 msec of the presentation of the display. This reversal is consistent with an initial bias by G cell activity and eventual R cell modulation. It may take 50–100 msec for the competition within the inter-cortical network across V1, V2, and V4 to stabilize border-ownership signals in such situations. Unlike some other types of models, such as Bayesian belief propagation networks, wherein connecting model results to timing is an additional degree of freedom (free parameter), the dynamic structure of our model naturally affords a link to actual temporal units (msec). We believe that the latency that is required for border-ownership signals to develop is evidence for the dynamic coding of figure-ground segregation in the cortex.

Many have proposed that the visual system detects local luminance junctions and that they are important for determining whether objects occlude others (Finkel, 1992; Bayerl & Neumann, 2006; Craft et al., 2007; Weidenbacher & Neumann, 2009). The on-surround competitive dynamics between B, G, and R cells in model V1, V2, and V4 allow the model to determine the occluding surface or figure without explicitly detecting
local luminance junctions. If the primate visual system detects local luminance signatures such as L- and T-junctions and uses that information to determine whether an object is in front of or behind another, humans should be good at determining if a local luminance pattern marks a point of occlusion in synthetic and natural images. McDermott (2004) studied human performance on this task when subjects viewed through a small aperture synthetic junctions or junctions sampled from larger images. His analysis showed that local information alone cannot fully explain the psychophysical results. Humans erred ~11% and ~11-27% in determining points of occlusion in the synthetic and natural junction conditions, even for the largest aperture sizes of degrees of visual angle. After considering scaling artifacts, ~25% of points that indicated the presence of occlusion could not be judged as such based on local junction information alone (McDermott, 2004). T-junctions are also not necessary to obtain a robust perception of occlusion (Shimojo, 1990; Zaidi, Spehar, & Shy, 1997; Peterson & Salvagio, 2008). There is limited neurophysiological evidence that supports the existence of cells that selectively respond to luminance junctions in “lower” visual areas, presumably where such an operation would need to take place (Lazareva et al., 2002). If local junctions were critical to the detection of occlusion, the scale at which they are viewed matters (Koenderink, 1984). Since V1 cells possess small receptive field sizes, any potential junction detector system would need to only consider junctions in a select number of spatial scales. Such a system would require more context outside the small V1 receptive field and very specific connectivity anatomies.

Occlusion and figure-ground interpretation in visual scenes does not always accompany the presence of T-junctions. Figure 3-9 shows that the RGB model also yields
border-ownership results in non-rectangular displays without T-junctions, consistent with the psychophysical data that humans are more likely to indicate that convex regions are figures compared to those that are concave. The simple ‘ring-like’ receptive fields of G and R cells resemble the on-surround receptive fields of cells in primate V2 and V4 and intrinsically respond to regions of convexity. The RGB model predicts that border-ownership cells would prefer the convex regions over the concave one in display shown in Figure 3-9.

3.4.2 Border-ownership recruits an inter-areal cortical network within the early visual system

Intra-areal fiber properties and conduction speeds impose constraints on how contextual information may propagate without feedback. Considering that the inter-areal distance between V1 and V2 is only several millimeters and axonal diameters are approximately 1μm, conduction times may be as small as 1msec (Rockland & Virga, 1990; Nowak & Bullier, 1997). The rapid speed of inter-areal connections is consistent with data showing minimal V2 response latencies of roughly 10-20 msec after V1 becomes active (Bullier, 2001), and certain “higher” visual areas having lower response latencies than “lower” ones in the visual hierarchy (Paradiso, 2002; Hochstein & Ahissar, 2002). The data of Sugihara et al. (2003), do not show changes in the slope of B cell responses as a function of figure size. Although the model of Zhaoping (2005) that exclusively used intra-areal connections showed a roughly equivalent latency in the onset of border-ownership signal development in the case of these data, the slope of the border-ownership signal strengths decreases as a function of the square display side length (see Zhaoping, 2005 Figure 3-6E-G). Unlike the measured cell responses, border-ownership
units in Zhaoping’s model take longer to develop the same strength responses for larger figure sizes. Consistent with the data of Sugihara et al. (2003), inter-areal conduction speeds would not predict a difference in the slopes of border-ownership signals as a function of the figure size (Craft et al., 2007). It is unclear whether models with extensive intra-areal connections can handle the description of border-ownership responses at arbitrary scales within which humans can perform figure-ground segregation. Our model relies on fast inter-areal connections rather than lateral connections among retinotopically proximal units in the same area.

Border-ownership sensitivity occurs in large populations of neurons in visual areas V2 and V4, yet approximately 20% of cells in V1 also possess this property. Without feedback, it remains unclear how 20% of V1 cells can be robustly sensitive to border-ownership via intra-areal connections alone. Intra-areal interactions between neurons of different orientation and border-ownership preferences, such as those used by Zhaoping (2005), may locally contribute to border-ownership, but we suspect inter-areal feedback is necessary to satisfy the neural timing constraints observed in the data.

3.4.3 Comparison with existing models

The use of feedback between units that detect edges and those with on-surround receptive fields has existed in the literature for some time. Instead of the B/G cell nomenclature, Pizer et al. (1992) describe the theory of Cores in which these units are called “boundariness” and “medialness” detectors, respectively (Pizer et al., 1992, 1998). Boundariness and medialness detectors feature a similar connectivity anatomy by which boundariness detectors do not directly communicate via horizontal-like connections, but
rather indirectly propagate information via the medialness detector. “Hot spot” clusters that obtain many votes from the edge detectors become known as cores, and may reside on a so-called medial axis of a figure. While the theory of cores provides an algorithmic description of a system capable of identifying figures, our model proposes a biological implementation that considers known neurophysiological information about B cells.

Existing models either employ junction detectors or extensive intra-areal connectivity. Kogo, Strecha, Van Gool, & Wagemans (2010) introduced the Differentiation-Integration for Surface Completion (DISC) model to reconstruct relative surface brightness and border-ownership. Although the DISC model performs many operations often attributed to neurons, such as reconciling border-ownership, it is algorithmic instead of neurally described. In a process similar to that performed by the Retinex algorithm (Land, 1986), the DISC model first constructs a map representing luminance ratios between different regions of the input display. In parallel, the algorithm assigns border-ownership throughout the image. DISC determines border-ownership based on a priori knowledge of the junction distribution, and labels T- and L-junctions using explicit rules. The DISC model simulates the illusory brightness percept in the Kanizsa square display as a function of different choices of inducers. Unlike the DISC model, the RGB model can simulate the predicted border-ownership responses yielded to the Kanizsa square display (Figure 3-11) in a neurally described rather than an algorithmic framework.

Zhaoping (2005) developed a neural model to describe how in the absence of feedback from higher visual areas cells in V2 could exhibit border-ownership sensitivity. The model uses intra-areal connectivity rules built on Gestalt grouping assumptions, such
as convexity. For example, the model confers larger synaptic weights between neurons separated by a right-turn compared to a left-turn, and adjacent neurons sampling like contrast or possess like border-ownership preferences signals facilitate or otherwise suppress each other. Units in the model are also assumed to possess end-stopped properties. Although the model does not incorporate an explicit T-junction detection stage, the Gestalt connectivity rules effectively afford border-ownership cells with receptive fields centered at T-junctions a bias consistent with occlusion. Because T-junctions feature bottom-up contrast signals of orthogonal contrast orientations, border-ownership units sampling the T-junction ‘hat’ and ‘stem’ would compete if the units had differing border-ownership preferences. Competitive ties are broken through the use of random noise or the inhibitory/facilitatory influence of adjacent units with (in)consistent border-ownership or contrast orientation preferences. Unlike the model of Zhaoping (2005), which cannot reconcile the presence of figures at multiple spatial scales with connectivity limited by known physiological fiber lengths, our model is fundamentally based on physiological data of inter-areal connections that do not face the same temporal constraints as intra-areal fibers and cells with a variety of receptive fields sizes.

While the Form-And-Color-And-DEpth (FACADE) model does not address the dynamics of border-ownership cells in primate visual cortex, its goal is to explain how presentations of 2D displays can give rise to 3D figure-ground perception by reconstructing surface brightness at different depth plains, despite whether the surface is visible (modal) or not (amodal). FACADE theory argues that representations of both modal and amodal percepts are required to act on visual information and recognize partially occluded objects, respectively (Kelly & Grossberg, 2000). FACADE builds on
the boundary contour system / feature contour system (BCS/FCS) (Cohen & Grossberg, 1984; Grossberg & Mingolla, 1985) that describes how a complementary neural representation of borders and surfaces can give rise to form and brightness in visual perception. First, the system performs on-center/off-surrond processing to discount the illuminant, but maintain relative contrast strengths in the borders. The BCS performs filtering to determine the borders in the visual display, regardless of contrast polarity and orientation. The FCS performs boundary-gated, nearest-neighbor diffusive filling-in jointly based on the outputs of the output BCS and on-center/off-surrond processing. Through the use of recurrent competition between bipole “long-range” grouping units, which possess a ‘bow-tie’-like receptive field with two excitatory lobes, and hypercomplex units, which act as end-stopped complex cells, the FACADE model does not require explicit T-junction detectors. While obtaining a border-ownership readout based on the FACADE model is possible, the process is indirect. Our model specifically focuses on the dynamics of cells that produce border-ownership signals.

3.4.4 Free-space border-ownership

Our results to the Kanizsa square display demonstrate that the model can reconcile figure-ground segregation in the presence of illusory figures. Importantly, we show with the progressive removal of inducers, the border-ownership responses become less consistent with an illusory form and more consistent with the individual inducers serving as the figure. The DISC model of Kogo et al. (2010) uses an interesting phenomenon called free-space border-ownership whereby units develop border-ownership preferences in the absence of bottom-up contrast signals. Some models (Zhaoping, 2005; Craft et al., 2007), including our own, assume border-ownership units
require bottom-up stimulation to acquire side-of-figure preferences. Current neurophysiology has not reported the existence of cells with free-space border-ownership sensitivity and indicates that border-ownership cells are edge-gated (e.g. see Figure 3-11 of Zhou et al., 2000). Nevertheless, the possible existence of free-space border-ownership cells has theoretical interest and should be investigated further in displays with illusory contours, such as the Kanizsa square. Such cells would complement our results shown in Figure 3-11 obtained with only edge-gated B cells.

The RGB model performs figure-ground segregation and is sensitive to occlusion without the use of junction detectors. Perception of local junctions may be an outcome of operations such as captured in our model, as suggested by the data of McDermott (2004), which show that allegedly “local” junctions can best be discerned only with large context in natural scenes. Evidently the visual system does not rely on specialized junction circuits to perform figure-ground segregation. Occlusion information conferred by T-junctions may be a specialized case of a more general process of inter-scale competition within a multi-area inter-cortical network.
Chapter 4

Border-ownership is a general-purpose signal for figure-ground segregation

4.1 Functional segregation in the visual system

The parsellation of neural function into discrete areas and pathways has dominated in visual neuroscience for the past half decade. Since Hubel & Wiesel discovered simple cells in area 17 of cat visual cortex, neuroscientists have focused on identifying the function of single neurons or functional organization of populations of neurons within a visual area. Complex cells, end-stopped cells, blob cells, and single-, double-, and triple-component cells in MSTd are some examples of the tendency to classify neurons based on their responses to characteristic low-level properties of visual displays (Duffy and Wurtz, 1991b; Maunsell and Newsome, 1987). Based on the functional classification of neurons in different visual areas, Mishkin and Ungerleider devised the notion that visual cortex contains two distinct processing pathways (Mishkin et al., 1983). Processing of the visual scene begins in the retina, and there is a proposed division of labor between the magnocellular and parvocellular pathways, which process motion and form information, respectively. These pathways remain segregated through LGN, V1, and extrastriate cortex to become the dorsal and ventral streams, respectively. The dorsal stream includes V1, MT, and MST, and the ventral stream includes V1, V2, and V4. Elements of human visual perception have been assigned to one of the streams by virtue of correlations with physiological properties. For example,
figure-ground segregation and movement detection have been suggested to be processed by the magnocellular pathway, and shape discrimination by the parvocellular pathway (Livingstone and Hubel, 1988).

More recent work has challenged the notion of segregated processing pathways by demonstrating extensive connectivity between ventral and dorsal stream areas and intermingling between inputs to the two streams. Originally, thin and thick stripes in V2 were thought to independently process fine and coarse information in the visual scene. However, studies have shown that both receive input from magnocellular and parvocellular signals (Sincich and Horton, 2002; Sincich and Horton, 2005). Moreover, extra-striate areas, such as V4 and MT, are anatomically far more interconnected than once thought, which may facilitate form-motion interactions (Gattass et al., 1988; Gattass et al., 1997; Ungerleider et al., 2007). Even in V1, neurons once thought to subserve a particular function, such as occlular dominance, motion, contrast, or color, exhibit multi-dimensional tuning to many properties (Schiller, 1996). This calls into question the utility of prevalent conceptualizations of functional organizations of visual areas in terms of visuotopic maps of color, orientation, motion, etc. Evidence suggests that a strict hierarchical organization may not best characterize the complexity found in the visual system. Many interesting aspects of visual perception may emerge through the multi-dimensional tuning of neural populations distributed across multiple visual areas.

Evidence indicates that border-ownership coding may occur independent of low-level tuning to motion, color, contrast, disparity, and so on. As described in the previous section, border-ownership cells respond to oriented edges of figures defined by luminance contrast (Zhou et al., 2000). Von der Heydt and colleagues demonstrated that border-ownership neurons in V1, V2, and V4 also demonstrate selectivity to color. For example, a neuron may produce border-ownership signals when
the figure is blue, not not red (Friedman et al., 2003). Color selectivity was fairly uniformly distributed across neurons tuned to border-ownership in the sample. When presented stereoscopically defined squares in random dot displays, neurons in V1 and V2 yield border-ownership signals, despite the absence of luminance contrast (Qiu and von der Heydt, 2005). Neurons that show side-of-figure tuning to gray squares exhibit response suppression when square is embedded into a perceptual transparency crossbar display, indicating the shift in global scene context (Qiu et al., 2007). Border-ownership neurons demonstrate side-of-figure selectivities to figures defined by Cornsweet cusp luminance gradients (Zhang and von der Heydt, 2010). Evidence indicates that border-ownership of occluded figures is modulated by attentional signals (Dong et al., 2008) and border-ownership activity is remapped when primates make saccades (O’Herron and von der Heydt, 2013).

4.2 Model of kinetic border-ownership

The independence of border-ownership coding to luminance contrast, color, disparity, transparency, and so on indicates its importance to figure-ground segregation. Border-ownership neurons respond to perceived edges of figures, irrespective of how they are defined. von der Heydt and colleagues showed that border-ownership signals also develop for kinetically-defined figures in V1, V2, and V4 (von der Heydt et al., 2003). Displays in the study contained a square defined by grayscale contrast and kinetic dot motion surrounded by an annular region with a different amount of uniform luminance and dots moving in a different direction. More neurons yielded border-ownership when the central region appeared as a figural ‘object’ than a ground ‘window’. Border-ownership neurons also responded to figures defined exclusively by random-dot texture accretion/deletion (personal communication, 2012).

Why are border-ownership neurons concerned with motion? The survival of ani-
mals depends on the rapid detection of predators. When stalking their prey, predators often use adaptive camouflage to home into striking range to launch an attack. Therefore, prey must be capable of detecting the form of predators with extreme rapidity when they break camouflage. The low latency of that border-ownership cells exhibit differential side-of-figure selectivity (≈25 msec) affords a feasible mechanism by which prey can detect the breaking of camouflage of a predator. Combined with color, disparity, and luminance domains in which border-ownership cells exhibit tuning, border-ownership neurons may be an integral part of an important fight-or-flight cortical network.

In the following section, we introduce a model that extends the RGB model by showing how kinetic border-ownership signals can emerge in cortex in visual displays wherein figures and their background are defined only by the texture accretion/deletion or shearing motion of random dots. The model predicts a dynamic recruitment of V4 to group static textured regions flanked by accretion/deletion and MT to pool over large regions of uniform motion. Feedback from neurons with larger receptive field sizes in V4 and MT are shown to be instrumental in the kinetic border-ownership signals. The model proposes functional roles for known anatomical connectivity between V2, V4, and MT. Model border-ownership outputs are consistent with human judgments of depth ordering in a number of psychophysical studies using kinetic random-dot displays.
Chapter 5

The border-ownership of kinetic edges: a neural model

5.1 Introduction

The process by which the primate visual system differentiates visual objects from their background is known as figure-ground segregation and represents a fundamental aspect of visual perception. Rapidly perceiving figure-ground relationships in the world is an important ability that allows humans to effectively act on and navigate about their environment. Objects perceived to be in front of (occluding) others is a natural instance of figure-ground segregation. Occlusion occurs when a physical object (occluder) blocks the line of sight between the observer’s eye and another object (occluded). Objects in the world are often opaque, in which case any incident light reflects off the object surface. When an occluding object is opaque, only part of the occluded object is seen, while the other part is covered by the occluding object. It has been hypothesized that local junctions, such as T-junctions, where luminance values due to multiple objects at different depths come together at a point, facilitate the identification of occluding and occluded opaque objects (Metelli, 1974; Beck et al., 1984; Anderson, 1997; McDermott, 2004). However, when objects are camouflaged with their surroundings, figure-ground segregation is much more challenging. Figure 5.1a shows an example of a mossy frog (*theloderma corticale*), which is camouflaged due to its textual similarity with a surrounding rock.
Figure 5.1: (a) A mossy frog (*theloderma corticale*) is camouflaged with a surrounding rock. Although the mossy frog may blend in well with the rock when stationary, the frog appears salient when in motion. (b) Schematic of a kinetic random dot display containing two different regions, labeled ‘A’ (left half) and ‘B’ (right half), separated by a vertical kinetic edge (marked in green), which is not seen unless dots in at least one region are in motion. When the dots are stationary, humans see a dense uniform texture, but when the dots in ‘A’ move differently than in ‘B’ humans report seeing the presence of two surfaces separated in depth. For example, if the dots in ‘A’ move to the right, but are deleted upon reaching the kinetic edge between the regions (marked in green) and the dots in ‘B’ remain stationary, region ‘B’ is perceived as its own surface in front of a surface composed of the texture in ‘A’. (c) A schematic illustrating the percept of the visual display shown in (b). Humans report seeing a surface made up of the dots in ‘B’ occluding a surface in region ‘A’ that moves underneath. The surfaces appear to be separated by a paper-thin margin in depth.
Luminance junctions are generally not necessary to form percepts of occlusion. When a camouflaged object moves in front of a textured background, as would be the case for a tiger moving in tall grass, the space occupied by the object is replaced with a very similar texture. When an object moves in front of a textured background, it is said to produce kinetic occlusion. Although a moving object may possess a texture that is similar to its background, humans can perceive salient figure-ground relationships due to the kinetic occlusion. The local luminance junctions within and around the object and background are statistically similar, yet humans can easily segment the two. Despite the textural similarity between the figure and ground surfaces and their motion, the borders of the figure are perceived to belong to the figure and not the ground (border-ownership). How do humans perceive robust figure-ground relationships in camouflage displays? We introduce a model that performs figure-ground segregation on kinetic displays based on differences in movement patterns (spatio-temporal correlations) between the occluding and occluded surfaces. The model addresses the rich percept of moving borders appearing to belong to or be owned by the occluding, but not the occluded, object by assigning border-ownership to the kinetic edges that separate the moving surfaces.

Kinetic random dot displays contain textured surfaces, each of which is composed of uniformly random luminance values, and resemble extreme cases of camouflage wherein luminance contrast provides no reliable information to the observer for determining figure-ground relations. Kinetic random dot displays are important, challenging tests to models of figure-ground segregation. Figure 5·1b depicts a kinetic random dot display that contains a far (ground) and near (figure) surface, labeled ‘A’ and ‘B’, respectively. When the surfaces are stationary, humans see a single uniform texture. However, if the surface defined by the dots to the left of the kinetic edge (vertical green dashed line, not actually part of the display) in ‘A’ moves coherently, yet
differently than the dots on the right hand side in ‘B’, humans perceive two surfaces at different depths separated by a kinetic edge (Kaplan, 1969; Yonas et al., 1987). As the dots in ‘A’ move rightward, the dots disappear (texture deletion) upon reaching ‘B’ at the kinetic edge. The dots in ‘B’, however, remain stationary over time. The percept is schematically illustrated in Figure 5·1c — humans perceive ‘B’ as the figure surface (closer in depth) and ‘A’ as the ground surface (farther in depth).

Over the past century, Gestalt psychologists have developed the concept of border-ownership and its integral role in figure-ground segregation (Rubin, 1921; Koffka, 1935; Kanizsa, 1979). When humans see a border separating a figure from its ground, it is grouped with and perceived as belonging to only the figure and not the ground. It is said that the border is owned by the figure. Recently, neurophysiological studies have identified neurons in primate visual cortex that are selective to border-ownership (Zhou et al., 2000; Qiu and von der Heydt, 2007). A border-ownership cell (B cell) elicits a higher response when there is a border within its receptive field that is attached to a figure to one side than when the border is attached to a figure on the other side of the receptive field (left vs. right, up vs. down). Border-ownership cell responses in non-human primates have been extensively studied in the context of simple static displays, such as rectangles, occluding squares, and C-shapes (Zhou et al., 2000). Von der Heydt et al. (2003) demonstrated that cells in primate visual area V2 elicit border-ownership responses to kinetic random dot displays with squares defined by coherently moving random dots that possess a different spatio-temporal correlation than surrounding dots.

The present article addresses how cells in primate V2 may elicit border-ownership signals based on motion alone, without reliable contrast information. Our model results indicate that recurrent feedback projections from cells in visual areas V4 and MT explain how cells in V2 can produce border-ownership signals in kinetic random
dot displays. The remainder of the Introduction focuses on psychophysical studies that assess human perception of kinetic random dot displays. The Methods section introduces the model circuits, simulation techniques, and equations. The Results section presents simulation results that demonstrate the competency of the model in performing border-ownership assignment of kinetic edges, relates the model performance with human psychophysics data, and compares the temporal dynamics of model border-ownership neurons to those reported in neurophysiology. The Discussion section describes a novel neural strategy for representing the relative depth ordering of kinetically-defined surfaces.

5.1.1 Kinetic random dot displays

Stationary & Encroach Conditions

Psychophysical studies have identified the conditions under which humans perceive occlusion in kinetic random dot displays. Kaplan first studied whether differences in the movement patterns of two surfaces alone yield percepts of relative depth and kinetic occlusion (Kaplan, 1969). He tested the hypothesis that accretion/deletion of texture elements optically specify “the direction of depth” or figure-ground relations about a kinetic edge. Subjects viewed moving ink blots that occupied one of two regions separated by a vertical kinetic edge on an overhead projector and underwent horizontal motion. Kaplan manipulated the relative velocities of the ink blots in either region and the initial placement of the kinetic edge. The ink blots were sufficiently fine so that the display appeared as a continuous texture when the blots were stationary. When texture in either region moved to the left or right, but the kinetic edge remained fixed over the ~4 sec trial, subjects responded that the stationary surface occluded the moving surface (stationary condition, Figure 5.1.1a). Subjects reported that the stationary surface was “covering” the moving surface, whose texture elements were deleted at the stationary kinetic edge. When dots in the two regions moved at
**Figure 5.2:** Kinetic random dot displays tested by the model, each represents an important test for neural models of kinetic border-ownership. The dashed green lines show the location of the kinetic edges that separate different regions of the display. Only dense random dots actually are in the displays. 

(a) **Stationary condition.** A display similar to those used in Kaplan (1969) whereby dots in ‘A’ move to the right, dots in ‘B’ remain stationary, and the kinetic edge between them remains stationary and therefore is correlated with the dots in ‘B’. Humans perceive surface ‘B’ (figure) as occluding surface ‘A’ (ground).

(b) **Encroach condition.** A display similar to those used in Kaplan (1969) whereby dots in ‘B’ move to the left and encroach on those in ‘A’. When dots in ‘B’ are displaced leftward toward ‘A’, dots in ‘A’ are deleted. The kinetic edge is thereby correlated with dot motion in ‘B’. Humans perceive surface ‘B’ (figure) as occluding and sliding over surface ‘A’ (ground).

(c) **Gap condition.** Similar to (b), except a white gap moves leftward with the dots in ‘B’ such that no texture accretion/deletion occurs (Yonas et al. 1987). Human subjects report seeing ‘B’ as occluding ‘A’, but with decreasing probability as the gap width increases.

(d) **Shear condition.** Dots in ‘A’ move parallel to the kinetic edge such that no texture accretion/deletion occurs and dots in ‘B’ remain stationary (Royden et al. 1988). Humans report seeing surface ‘A’ as figure and ‘B’ as ground.

(e) **Strip condition.** Dots in the center region ‘B’ remain stationary and surrounding dots in ‘A’ move orthogonally to the vertical kinetic edges. Human responses indicate that subjects see ‘B’ in front of ‘A’ (van Doorn & Koenderink 1982).

(f) **Object and Window conditions.** A square region is defined by dot motion and luminance contrast differences from the surrounding annular region (von der Heydt et al. 2003). The square region either appears as an ‘object’ (bottom left panel) or ‘window’ (bottom right panel). Top panel taken with permission from von der Heydt et al. (2003).
different speeds and potentially different directions or when the texture to one side moved while the other remained stationary, but the kinetic edge moved with one of the surfaces, the surface that was correlated with the edge was seen as in front of the other (*encroach condition*, Figure 5.1.1b). Both conditions are examples of the Gestalt principle of common-fate — the surface that moves together with the kinetic edge is seen as the figure and the other surface is the ground. The kinetic edge is perceived as owned by the occluding surface. Border-ownership reports in humans are consistent with analyses performed by Gibson and his students, which indicate that the covering and revealing of a surface under certain conditions may be optically specified by texture accretion/deletion (Gibson et al., 1969). Subjects in Kaplan’s study did not always perceive figure-ground relationships, as was the case when the horizontal dot velocities in ‘A’ and ‘B’ were equal in magnitude, differed in direction, the kinetic edge remained fixed, and moving dots in ‘A’ and ‘B’ were deleted upon arriving at the kinetic edge (*degenerate condition*). Instead of perceiving figure and ground surfaces, subjects perceived a thin kinetic edge that did not belong to either region. Models of figure-ground segregation should differentiate between the cases wherein humans do and do not perceive strong depth orderings around a kinetic edge.

**Gap Condition**

In Kaplan’s study, texture elements in the figure surface always were correlated with the kinetic edge direction of motion, texture accretion/deletion always occurred at the kinetic edge, and texture elements in both surfaces moved relative to one another. From the experiments of Kaplan, it is not clear whether the visual system relies on accretion/deletion or relative motion information to determine the figure-ground relationships of the two surfaces. Yonas tested whether texture accretion/deletion between two kinetic random dot surfaces is necessary to obtain figure-ground percepts
by introducing a white, variably-sized “gap” between the independently controlled left and right regions of the display (gap condition, Figure 5.1.1c) (Yonas et al., 1987). The percentage of trials in which subjects responded that there was a depth ordering in the displays was highest when there was texture accretion/deletion (zero gap). Depth ordering percepts dropped off monotonically as the size of the gap between the two regions grew. The results are compatible with the idea that neurons in the visual system with limited receptive field sizes are involved in the determination of kinetic border-ownership and figure-ground segregation. By virtue of their prevalence in early primate visual areas, B cells have small receptive field sizes — both compared to many other visual cortical areas and the figures to which they respond. For example, von der Heydt and colleagues report that median B cell receptive field sizes at foveal eccentricities are 0.5°, 0.7°, and 3.6° in monkey visual areas V1, V2, and V4, respectively (Zhou et al., 2000). The rich percept of border-ownership at the kinetic edge and the dependence on narrow gap sizes for robust depth ordering percepts to occur suggest that B cells may be involved in the same cortical network that determines figure-ground relations in kinetic random dot displays.

**Shear Condition**

The displays used in the experiments of Kaplan and Yonas required horizontal dot motion, perpendicular to a vertical kinetic edge. Royden studied kinetic random dot displays that involved a motion-defined rectangle surrounded on all sides by another region of stationary dots (Royden et al., 1988). Dots moved either horizontally or vertically within the rectangle, and the aspect ratio of the rectangle was manipulated. Kinetic edges were defined either by texture accretion/deletion, as was the case in the displays of Kaplan and Yonas, or shearing motion whereby the dot motion was parallel to the kinetic edge (shear condition, Figure 5.1.1d). When the rectangle aspect ratio made the kinetic edges where accretion/deletion occurred much longer
than those where shearing motion was present, subjects responded that the rectangle appeared to be occluded by the surrounding stationary dots, similar to human reports in the stationary condition (Figure 5.1.1a). Conversely, when the rectangle aspect ratio made the kinetic edges where shearing motion occurred longer than those where accretion/deletion occurred, subject responses indicated that the rectangle appeared in front of the surrounding region. Moreover, the data indicate that the figure-ground percept is more powerful when the rectangle was defined predominately by accretion/deletion than shearing motion. Any model that makes mechanistic predictions about how the visual system performs kinetic figure-ground segregation must address the difference in percepts that occur when a kinetic edge is defined by texture accretion/deletion or shearing motion.

**Strip Condition**

Rather than dividing the kinetic random dot display into two adjacent regions, van Doorn studied the effects that tiling parallel kinetic edges across the display had on figure-ground perception (van Doorn and Koenderink, 1982). The displays of van Doorn consist of rectangular regions of dots whereby dot velocities alternate in adjacent regions (strip condition). When the strips were thick, subjects indicated that they perceived adjacent strips as separate surfaces. Figure 5.1.1e shows an example of a strip condition display with vertically-oriented kinetic edges. Because the kinetic edges remain stationary, deletion of the left peripheral texture occurs at the left kinetic edge, and accretion of the right peripheral texture occurs at the right kinetic edge, the central strip appears in front of the surrounding two strips. When the strips were very narrow, responses indicated that the subject saw dots occupying two superimposed transparent planes. Subjects saw alternating motion-defined surfaces when the thickness was as narrow as 2”, which demonstrates the high spatial resolution of the mechanisms in the visual system involved in determining
kinetic figure-ground relations. Thicknesses in between resulted in incoherent, white noise “snow” percepts. Models of figure-ground segregation in kinetic random dot displays should assign border-ownership of the kinetic edges inward toward the strip centers when humans perceive the strips as figures. The presence of multiple strips adds to the complexity of the displays because the visual system determines that certain pairs of adjacent kinetic edges belong to the same strip, and models must perform the same border-ownership assignment.

Object & Window Conditions

Von der Heydt and colleagues tested whether kinetic random dot displays elicit border-ownership responses in primate visual areas V2, where there are cells known to exist that produce border-ownership signals to static displays (von der Heydt et al., 2003). von der Heydt and colleagues presented monkeys visual displays that consisted of two regions that differed due to dot motion differences and luminance contrast (Figure 5.1.1f). The square region has a different mean grayscale luminance value than the surrounding region and either region contained sparse moving or stationary dots. The kinetic edges that border the square region were either correlated with the dots in the square (object condition) or with the surround region (window condition). In the object condition, the square appeared as the figure surface. The dots within the square region were constrained to move with the same velocity as the kinetic edges (the kinetic edges remained fixed if the dots in the square had zero velocity). In the window condition, the square region appeared as a hole or aperture and the surrounding region served as the figure surface. The texture in the square region was accreted/deleted and the surrounding texture moved with the kinetic edges. In both conditions, there were cases in which either the kinetic edges had a zero (“stationary edge” condition) or non-zero (“moving edge” condition) velocity. von der Heydt and colleagues also reversed the luminance contrast polarity
on separate trials. For example, there were window stationary edge conditions in which the mean square luminance was either lighter or darker than the surroundings. Border-ownership neurons whose receptive fields were centered on the kinetic edges of the square elicited *ceteris paribus* differential responses in the object and window conditions, which suggests B cells incorporate texture accretion/deletion information in border-ownership signals. Interestingly, more neurons in the sample preferentially signaled border-ownership when the square was consistent with the object interpretation than the window interpretation. The results indicate that there are B cells in V2 that produce border-ownership signals near kinetic edges.

Kinetic random dot displays represent important tests for any model that attempts to describe the underlying neural dynamics of kinetic border-ownership because reliable luminance contrast information is absent. In particular, the model should yield border-ownership signals consistent with human figure-ground perception in the displays of Kaplan (Figure 5.1.1). The model should show that texture accretion/deletion is sufficient, but not necessary, to elicit kinetic border-ownership signals (Yonas et al., 1987), and it should be able to differentiate between accretion/deletion (Figure 5.1.1a) and shearing motion (Figure 5.1.1d). When there are multiple kinetic edges, under appropriate conditions the model border-ownership signals should indicate the presence of different strip surfaces (van Doorn and Koen-derink, 1982). Finally, when random dots within a shaped region (e.g. square) move differently than in the surrounding region, model border-ownership signals should differ when the square region is interpreted as an object or a window (von der Heydt et al., 2003). We present a simple neural model that performs border-ownership assignment consistent with human perception in these kinetic random dot displays, using grouping and competition between units that respond to different spatio-temporal correlations. Our model is consistent with known neurophysiology and the results
explain how cells in primate V2 may produce border-ownership signals of kinetic edges (Zhou et al., 2000; von der Heydt et al., 2003). The model specifically predicts that known anatomical connectivity between areas V2, V4, and MT at least in part subserves motion-induced border-ownership signal generation.

5.2 Methods

![Diagram of visual input and cell types in the primate visual system](image)

In this section, we introduce the core model circuits, visual displays that are simulated in the model, and mathematical equations.

5.2.1 Magnocellular and parvocellular pathways

Figure 5.3 summarizes the model architecture. The model contains stages that correspond to populations of cells within the primate retina (not shown), lateral geniculate nucleus (LGN), primary visual cortex (V1), V2, V4, and medial temporal area (MT).
Figure 5-3: Model diagram. The model consists of stages corresponding to primate visual areas LGN, V1, V2, V4, and MT. The top (bottom) set of panels define model magnocellular (parvocellular) pathways. Magnocellular model units detect non-zero (moving) spatio-temporal correlations and determine border-ownership of regions with motion, while parvoceullar units detect stationary correlations and determine border-ownership of static regions. The model uses two sub-populations of LGN cells: transient and tonic. The former decays faster than the latter. V1 units detect spatio-temporal correlations by receiving feedforward input from several spatially displaced LGN cells with different conduction delays. V1 units with different spatio-temporal correlations compete in a contrast-enhancing network. Each type of V1 unit that is tuned to a different spatio-temporal correlation projects separately to B cells in V2. Magnocellular B cells project to MT and parvocellular B cells project to V4. MT cells send feedback to both inhibit magnocellular and parvocellular B cells in V2 in locations where MT receives excitatory feedforward input. V4 units also feedback to V2, but only target parvocellular units and units with spatio-temporal correlation sensitivities in directions that are orthogonal with respect to the V4 cell circular receptive field shape (see Figure 6 and Table 1 for more details).
Because the visual displays used to investigate the perception of figure-ground segregation and border-ownership in humans involve dots that either remain stationary or move, our model contains parallel subsystems that correspond to the magnocellular (Figure 5·3, top pathway) and parvocellular (Figure 5·3, bottom pathway) pathways of the primate visual system. Cells within the primate parvocellular pathway have higher spatial, but lower temporal resolution compared to those in the magnocellular pathway (Livingstone and Hubel, 1988). Therefore, model cells in the parvocellular and magnocellular pathways are more sensitive to static surfaces and moving dot patterns in moving surfaces, respectively.

Model V1 cells are tuned to different spatio-temporal correlations due to the convergence of inputs from LGN cells with spatially displaced receptive field centers and conduction delays. Each type of correlation cell in V1 projects to corresponding B cells in V2, which are sensitive therefore to the same range of spatio-temporal correlations. Possible mechanisms by which B cells acquire their side-of-figure selectivities in static displays with shapes defined by luminance contrast have been studied (Craft et al., 2007; Layton and Browning, 2012). The mechanisms underlying motion-induced border-ownership sensitivity have not been extensively investigated. Border-ownership units in our model develop side-of-figure selectivities via feedback from units with larger receptive fields within areas V4 and MT. After model area V2, the parvocellular and magnocellular pathways bifurcate — the former independently projects to model V4 and the later to model MT. When the pathways diverge, both projections to V4 and MT include both feedforward and feedback connections.

Although there is evidence for separate magnocellular and parvocellular pathways in the primate visual system, recent data suggest that there are substantial interactions between the pathways (Sincich and Horton, 2005). Our model includes considerable competitive/cooperative interactions between the parvocellular and magnocellul-
lar systems. Units in V1 compete across correlations to perform contrast-enhancement on spatio-temporal correlation signals. MT units feedback to inhibit B cells with different spatio-temporal correlation selectivities than the B cell that supplied the MT unit input. B cells tuned to moving correlations modulate V4 cells that are driven by B cells tuned to stationary correlations by enhancing the V4 unit’s activity when accretion/deletion is detected within the V4 unit’s receptive field.

Transient cells in model LGN respond to changes in luminance, irrespective of the dot motion direction. Magnocellular and parvocellular LGN cells both elicit a transient response, except parvocellular cells decay much faster (Figure 5·3, tonic and transient LGN cell panels). The response of model LGN cells is consistent with the neurophysiological data of Maunsell (Maunsell et al., 1999). In the study, macaque monkeys were presented spots of light and single-cell recording was performed on neurons in LGN. Magnocellular response latencies preceded the fastest parvocellular response latencies by approximately 10 msec. In one monkey, magnocellular minimum and median response latencies were 16 msec and 21 msec, respectively, and parvocellular minimum and median response latencies were 24 msec and 31 msec, respectively. M retinal ganglion cells and magnocellular LGN cells have a higher ratio of early to late average firing than P retinal ganglion cells and parvocellular LGN cells, which suggests that magnocellular neurons in the early visual system exhibit a higher degree of transience and decay faster than parvocellular neurons (Schiller and Malpeli, 1977; Schiller and Malpeli, 1978). Indeed, evidence suggests that magnocellular LGN neurons conduct impulses faster than parvocellular LGN neurons (Dreher et al., 1976). Although it has been estimated that cells in the magnocellular pathway transmit impulses to LGN and V1 ∼3 msec and ∼5 msec quicker, respectively, than cells in the parvocellular pathway, LGN neurons in both pathways have highly overlapping response latency distributions (Nowak and Bullier, 1997). In macaque V1 neurons,
lesioning parvocellular LGN neurons increased the transience of V1 cell responses, but lesioning magnocellular LGN cells had no effect (Maunsell and Gibson, 1992). This suggests that the responses of V1 cells obtain their transience at least in part by feedforward projections from magnocellular LGN neurons, but not parvocellular LGN neurons.

5.2.2 Spatio-temporal correlation detection

Model V1 cells are sensitive to spatio-temporal correlations that are present in the kinetic random dot displays and compete in a contrast-enhancing network to enhance the dominant correlation. Figure 5·4a shows a schematic depiction of the model motion detection circuit. We use fast ($\delta_f$) and slow ($\delta_s$) conduction delays between LGN and V1 to detect spatio-temporal correlations in the input video, consistent with neurophysiological evidence for considerable variability in V1 response latencies and conduction delays in LGN (Nowak et al., 1995; Maunsell et al., 1999). For example, a V1 cell that receives input from three pairs of LGN cells, which have horizontally-adjacent receptive fields and the conduction delay of the right cell is longer than the left, will be sensitive to leftward motion assuming the movement occurs within a limited speed range (Figure 5·4c). Our correlation detection mechanism is a modified Reichardt circuit that pools over the response of several small subpopulations of LGN.
**Figure 5-4**: Spatio-temporal correlation detection circuit and V1 sub-unit structure. (a) The model implements a modified Reichardt motion detection circuit that affords V1 units sensitivity to different spatio-temporal correlations. LGN units have slow ($\delta_s$) or fast ($\delta_f$) conduction delays between LGN and V1. Each ellipse indicates a single spatial location where a number of LGN cells with different conduction delays have overlapping receptive fields. When a dot moves to the left, the V1 unit sensitive to leftward spatio-temporal correlations is active because it receives input from a LGN cell in the right position with a slow conduction delay and a LGN cell in the left position with a fast conduction delay. As the dot moves to the left within the V1 cell’s receptive field, the two signals arrive at the same time. Each model V1 cell integrates multiple LGN cell pairs at different spatial locations to increase the motion detection confidence. We normalize by the luminance magnitude so that the units are sensitive to correlations and not luminance fluctuations. (b) The subunit structure of a sample V1 cell with a left preferred motion direction (Livingstone, 2003). The cell has facilitatory and suppression regions aligned orthogonally with respect to the preferred motion direction (left). (c) Model V1 units also pool over LGN subunits orthogonal to the preferred motion direction.
cells to increase confidence about the dominant spatio-temporal correlation that is detected (Adelson and Bergen, 1985; Reichardt and Egelhaaf, 1988). Reichardt-style motion detection models require the comparison of at least two spatially and temporally separated signals, and a motion output signal is only produced when the convergent input signals are appropriately correlated (Hassenstein and Reichardt, 1956). By contrast, Barlow-Levick models assume that for a motion detector is sensitive to any motion direction for which the detector does not receive nulling inhibition (Barlow and Levick, 1965; Barnes and Mingolla, 2012). Evidence for Barlow-Levick models comes from Meynert cells in layer six of V1 that have large asymmetrically shaped dendritic trees, which would allow nulling inhibition to act on the signal propagating along the dendrites (Livingstone, 1998). A large dendritic tree could also support a Reichardt detection mechanism. If the cell has a high firing threshold and two input signals from LGN cells converge on the dendritic tree at the appropriate time and location to summate and exceed a firing threshold, the cell could elicit a direction-selective response. We take this possibility as well as the extensive magnocellular and parvocellular LGN latency scatter as evidence for a Reichardt correlation-detection mechanism, but it is possible that the two motion detection methods co-exist. Gradient and Fourier models for motion detection have also been proposed, though it is less clear how the latter may be implemented in cortical circuits (van Doorn et al., 1984).

We obtain better sensitivity to kinetic edges that occur over the width of a single pixel by pooling over convergent LGN responses in the direction orthogonal to the cell’s preferred motion direction (Figure 5·4c). This mimics the subunit structure of V1 complex cells that have facilitatory zones aligned perpendicularly to the cell’s preferred motion direction (Livingstone and Conway, 2003) (Figure 5·4b). Both complex cells in neurophysiology and in the model do not require this subunit structure to
obtain motion sensitivity, but it assists model V1 cells in detecting the fine spatial resolution kinetic edges.

### 5.2.3 Border-ownership cells

Neurons in V2 have been shown to respond to kinetic edges (Chen et al., 2012), which are defined only by the separation of two adjacent regions of randomly moving dots that move with different spatio-temporal correlations (Figure 5·1a). Because neurons in V1 and V2 also tend to be sensitive to uniform local motion patterns, Orban and colleagues determined which neurons actually responded to oriented kinetic edges rather than to the local motion patterns (Marcar et al., 2000). A neuron is sensitive to an oriented kinetic edge if the neuron responds to the kinetic edge irrespective of the local dot motion directions to either side. A neuron sensitive to a horizontally-oriented kinetic edge responds not only when there is rightward motion in the top region and leftward motion in the bottom region, but also when there is upward motion in the top region and downward motion in the bottom region. If the neuron only is sensitive to local motion patterns on either side of the kinetic edge, it responds in one case but not the other. Orban and colleagues showed that 11.5% of neurons in their V2 sample met their requirements for kinetic edge sensitivity compared to only 3.7% in V1, which indicates that kinetic edge sensitivity is more common in V2 than V1 (Marcar et al., 2000). Similar to area MT, far more cells in V1 responded better to local motion patterns and 90° rotations deteriorated the responses (Marcar et al., 1995). More neurons in V2 responded to both the rotated and non-rotated displays (see Figure 6a of (Marcar et al., 2000)). Neurons sensitive to kinetic edges have not been tested for border-ownership selectivity. Von der Heydt et al. (2003) tested border-ownership neurons on moving random dot displays with kinetic edges and reported 18% to be sensitive to the kinetic edge polarity in V2 and none in V1, which are similar numbers reported by Orban (Marcar et al., 2000). We hypothesize
that a subpopulation of kinetic edge sensitive neurons are sensitive to kinetic edge border-ownership and is crucial for kinetic figure-ground segregation in primates.

Model B cells in V2 develop side-of-figure selectivities near kinetic edges. Due to the rectilinear structure of the kinetic random dot psychophysics displays, we simulate B cells with four possible border-ownership directions specified by \( \theta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\} \): right, up, left, and down. Each B cell in V2 receives feedforward projections from V1 correlation cells of a common type over a small spatial extent defined by a 2D Gaussian kernel \( G_{V_2} \). Receiving input from V1 correlation cells of a single correlation type bestows the B cells sensitivity to the same spatio-temporal correlation. Therefore, V1 cells in the model parvocellular pathway that respond to stationary patterns project in parallel to different B cells in V2 than V1 cells that respond to non-stationary patterns in the magnocellular pathway.

5.2.4 Magnocellular and parvocellular interactions: feedback from V4 and MT to V2

The model predicts that B cells in the parvocellular and magnocellular pathways project differently to V4, which has implications for how feedback acts on B cells within each respective pathway. As depicted in Figure 5.5a, V4 cells receive feedforward projections from parvocellular B cells (see Model Equations for more details). The model assumes that feedforward projections from V2 to V4 take on an annular structure. That is, V4 does an on-surround integration of B cell inputs in V2 (Pasupathy and Connor, 1999; Craft et al., 2007; Layton and Browning, 2012). This connectivity affords V4 with round, ring-like receptive field shapes, larger in size compared to those of B cells. A V4 cell with a receptive field center located at position \((i, j)\) does not receive input from B cells located at \((i, j)\), but does so from cells at a distance (on-surround projections; see Figure 5.5a). A model V4 cell sends inhibitory feedback to B cells at the same locations that supplied it feedforward input,
Figure 19: Connections and correlations between V2, V4, and MT areas.

**Legend**
- **Connections**
  - Excitatory
  - Inhibitory
- **Border-ownership Strength**
  - Strong
  - Weak
- **Border-ownership**
  - Left Motion Selective
  - Right Motion Selective
  - Static Selective
- **Correlation Directions**
  - Stationary
  - Left Border-ownership Selective
Figure 5-5: (a,b) A schematic of the feedforward and feedback connectivity between V4 and V2. (a) Model V4 cells receive driving, excitatory feedforward input from parvocellular V2 B cells (black ellipses) that preferentially respond to stationary correlations. When magnocellular B cells are active within the V4 cell’s receptive field and are sensitive to spatio-temporal correlations that are orthogonal in direction to the V4’s receptive field (red and blue B cells in (a)), the magnocellular B cells send modulatory excitatory input to the V4 cell. Parvocellular B cells provide driving input and gate any response elicited by the V4 cell. Subpopulations of magnocellular B cells sensitive to orthogonal correlations are capable of boosting the V4 cell response gain. (b) V4 cells send inhibitory feedback to parvocellular B cells in locations in which the V4 cell received excitatory input, but the inhibition targets B cells with outward border-ownership preferences relative to the V4 receptive field center. V4 cells also inhibit magnocellular B cells with inward border-ownership preferences within their receptive fields. This connectivity affords V4 cells sensitivity to stationary regions, particularly those surrounded by texture accretion/deletion. (c,d) The feedforward and feedback connectivity between model MT and V2. Model MT additive cells a perform long-range inter-areal grouping of magnocellular B cells with the same moving correlation sensitivity and inhibit B cells that respond to different moving correlations with inward border-ownership selectivities. (e) The shear condition. The top region (figure surface) has rightward moving dots and the bottom region has stationary dots (ground surface). Model V4 cells do not inhibit B cells whose correlation direction is tangent to the V4 cell receptive field, and therefore rightward correlation B cells (red) with upward border-ownership will be the most active along the kinetic edge.
but the inhibition targets B cells with outward side-of-figure selectivities (differing by 180°) within the parvocellular pathway (Figure 5·5b, feedback connections to black B cells). Inhibiting the border-ownership direction that points away from the center of the V4 annular receptive field implicitly enhances border-ownership in the inward direction toward the V4 receptive field center. Note that excitatory feedback to B cells with side-of-figure selectivities directed inward toward the center of the annulus may result in an unstable positive feedback loop. Inhibitory feedback from V4 to V2 also results an annular spatial arrangement.

V4 cells send inhibitory feedback to B cells in the magnoceullar pathway in locations where they received feedforward input from parvocellular V2 cells, targeting cells with a side-of-figure selectivity directed inward toward the annular center (see Figure 5·5b, feedback connections to colored border-ownership units). This connectivity pattern signals that a kinetically-defined surface should be composed of elements moving with a single spatio-temporal correlation. In other words, if a V4 cell responds within a stationary textured surface, the cell signals that the surface has a stationary spatio-temporal correlation and none other. However, the model predicts that V4 cells only inhibit magnocellular cells sensitive to a correlation that involves movement in the radial direction, which is orthogonal to the annular receptive field. That is, model V4 cells only inhibit magnocellular B cells that would respond when texture accretion/deletion is present at the borders of the V4 receptive field. When there is a stationary surface with fixed kinetic edges and texture accretion/deletion occurring at the kinetic edges, humans perceive the stationary surface as figure. In this case, the V4 cells that respond to the stationary surface inhibit magnocellular B cells at the kinetic edge due to the presence of texture accretion/deletion, which implicitly strengthens the parvocellular B cells with inward side-of-figure selectivities toward the stationary surface. Therefore, border-ownership signals emerge at the
kinetic edges directed toward the stationary surface. Shearing motion occurs under the same circumstances, but without texture accretion/deletion at the kinetic edges. In this case, V4 cells do not inhibit the magnocellular B cells with receptive fields centered along the kinetic edges because the surrounding motion pattern is tangent to the annulus. The model implicitly enhances border-ownership signals toward the surrounding surface (see Figure 5.5e), which is consistent with the human percept that the moving surface is seen as figure (Royden et al., 1988). See Table 1 for a summary of V2 to V4 connectivity in the model.

MT cells are tuned to specific spatio-temporal correlations and respond to large fields of coherent motion (Snowden et al., 1991). Model MT units have annular receptive fields and group border-ownership unit activity in the magnocellular pathway. Border-ownership units that are tuned to a specific spatio-temporal correlation project to corresponding MT cells, thereby affording MT additive cells sensitivity to the same spatio-temporal correlation over a larger area (Figure 5.5c). MT has been shown to send feedback projections to V2 (Born and Bradley, 2005; Rockland and Knutson, 2000). Model MT cells inhibit magnocellular B cells in V2 that are tuned to the same spatio-temporal correlation within the receptive field that have outward side-of-figure selectivities (differ by 180°) relative to the center (Figure 5.5d, feedback connections to red B cells). A MT cell in the model nonspecifically inhibits all other B cells in positions where the MT cell received input with side-of-figure selectivities that are directed inward toward the center, except for V2 cells that have the same correlation sensitivity (Figure 5.5d, feedback connections to border-ownership units that are not red). This feedback connectivity affords a similar effect to that achieved by the V4-V2 feedback connectivity — implicit enhancement of border-ownership in the direction of moving surfaces that have the same spatio-temporal correlation preference as the MT cell, thereby signaling that the surface should move with a uniform
### Table 5.1: Summary of feedforward and feedback connections between V2 and V4 and MT. “Same type” means projecting to units that are sensitive to the same spatio-temporal correlation. “Radial” means projecting to border-ownership cells with spatio-temporal correlation sensitivities that are orthogonal relative to the projecting unit’s preference. “Inward” (“outward”) means projecting to border-ownership cells with side-of-figure selectivity that is directed toward (away from) the projecting unit’s receptive field center. “All” means projecting to both orthogonal and same type border-ownership cells.

<table>
<thead>
<tr>
<th>Feedforward connections (excitatory)</th>
<th>Feedback connections (inhibitory)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>V4</td>
<td>MT</td>
</tr>
<tr>
<td>Magno V2</td>
<td>N/A</td>
</tr>
<tr>
<td>Parvo V2</td>
<td>Same type</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Same type</td>
</tr>
</tbody>
</table>

*Spatio-temporal correlation.* This type of surround inhibition in physiology, possibly through feedback, has also been reported to occur in other visual areas (Tanaka et al., 1986; Zipser et al., 1996). A summary of the V4 and MT feedback connectivity in the model is shown in Table 5.1.

#### 5.2.5 Visual Displays

Seven types of kinetic texture displays served as the inputs to the model that are schematically depicted in Figure 5.1.1. All displays subtended 100 x 100 px, were presented at 20 frames/sec, and appeared for a total of 2 sec. The displays initially consist of randomly generated grayscale values, and as time evolves, dots within each of the regions marked ‘A’ and ‘B’ in Figure 5.1.1 are correlated with one another to yield uniform displacements $\mathbf{v}_A$ and $\mathbf{v}_B$ between successive video frames, respectively. We constrain $\mathbf{v}_A$ and $\mathbf{v}_B$ to unit speed for non-zero velocities. Unless otherwise specified, we configure the visual displays such that the regions ‘A’ and ‘B’ constitute the left and right sides of the display and correspond to the surfaces perceived as ground and figure by human observers, respectively. The kinetic edge that is formed between the regions labeled ‘A’ and ‘B’ is always vertically-oriented and initially positioned at the center of the display, unless noted otherwise.
In the **Stationary condition** (Figure 5.1.1a), a kinetic edge appears between a stationary dense random texture (‘B’) and another surface (‘A’) that moves perpendicularly to the kinetic edge. Similar to the Encroach condition, the kinetic edge is correlated with the movement of the ‘B’ surface. Therefore, because the ‘B’ surface remains stationary over time, the kinetic edge is not displaced throughout the video.

The **Encroach condition** (Figure 5.1.1b) consists of a dense random texture ‘B’ that moves with a perpendicular kinetic boundary separating the ‘A’ surface from another surface ‘B’. That is, the motion of ‘B’ surface is correlated with the kinetic edge displacement over time and covers over part of ‘A’ over time, indicated by the purple shaded region. For example, dots in surface ‘B’ could be uniformly displaced by one pixel to the left thereby encroaching on and inducing texture deletion of dots in a stationary ‘A’ surface. The surface ‘B’ appears to grow larger as it encroaches on ‘A’. Dots in the ‘A’ surface may move in any one direction not equal to that of ‘B’ or remain stationary throughout the video. The motion of the ‘B’ surface is perpendicular to the orientation of the kinetic edge and texture elements are accreted/deleted depending on whether the ‘B’ surface appears to grow/shrink.

In the **Gap condition** shown in Figure 5.1.1c, dots in region ‘B’ that move orthogonally with respect to the vertical kinetic edge are correlated with the kinetic edge as in the Encroach condition, except a white gap moves with the region ‘B’ and separates the two surfaces (Yonas et al., 1987). The gap width in simulations is 0-5 px wide, in 1 px increments.

We also simulate the **Shear condition** (Figure 5.1.1d) of Royden whereby dots in a random texture surface ‘A’ move parallel to the stationary surface’s (‘B’) kinetic edge. For example, dots in surface ‘A’ that move horizontally vis-à-vis those that are stationary in surface ‘B’ yield shearing motion. The kinetic edge is not displaced throughout the video. ‘A’ and ‘B’ are considered the figure and ground surfaces,
respectively.

The Strip condition (Figure 5.1.1e) is composed of three (one center, two peripheral) textured regions and two kinetic edges. The two peripheral surfaces possess the same spatio-temporal correlation, which differs from that of the central surface (strip). Dots move perpendicularly to the vertically-oriented kinetic edges, thereby resulting in texture accretion/deletion. The texture surfaces evenly partition the visual display and kinetic edges are not displaced throughout the video. Therefore, the one or more regions that are stationary appear as figure and own the kinetic boundary, and the one or more surfaces with moving dots appear as ground. For example, the peripheral surfaces, when stationary, appear in front of the central surface whose dots move laterally and are accreted at one kinetic edge and deleted at another. The Strip condition is qualitatively similar to the “broad strip” condition of van Doorn.

Finally, we simulate the Object and Window conditions shown in Figure 5.1.1f (von der Heydt et al., 2003). Because model units are not sensitive to luminance contrast and only spatio-temporal correlations, we construct two regions ‘A’ and ‘B’ that are composed of randomly moving dots as follows: region ‘A’ is a square region surrounded by an annular region ‘B’. For simplicity, we simulate cases wherein the square’s kinetic edges remain stationary. Either the dots in the square move and get deleted/accreted at the borders while the surrounding dots remain stationary (Window) or the surrounding dots move and are accreted/deleted at the square boundary while the dots in the square remain stationary (Object). The Object and Window conditions are 2D versions of the Strip condition and the displays possess four kinetic edges.

### 5.2.6 Model Equations

Model equations assume the form of a system of delay differential equations (DDEs) that reflect the dynamics of individual neurons within a brain area. The DDEs we consider are first-order differential equations where the temporal derivative at the
present time depends on the solution at present and past times.

\[
\begin{aligned}
\dot{x}(t) &= F(t, x(t), x(t - \tau_1), \ldots, x(t - \tau_n)) & t > t_0 \\
x(t) &= \phi(t) & t \leq t_0
\end{aligned}
\]  

(5.1)

In Eq 5.1, \(x(t)\) signifies the response of model unit \(x\) at time \(t\), the initial history function \(\phi(t)\) specifies the solution at times earlier than some initial time \(t_0\), and \(\tau_1, \ldots, \tau_n\) specify \(n\) possible conduction delays. For simplicity, our model uses a single delay \(\tau\) that describes the monosynaptic conduction delay between units in different early visual areas. Since the inter-areal distance between V2 and V4 is no more than 20 mm (Gattass et al., 1988) and the median inter-areal conduction velocity is approximately 3.5 m/s in early primate visual areas (Girard et al., 2001), a signal propagating between two early visual area takes approximately 5 msec. Therefore, we set \(\tau = 5\) msec in all simulations. In all our equations, we employ the initial history function \(\phi(t) = 0\) for all \(t \leq t_0\) to indicate that the cell response before the stimulus appears is zero.

Since the operations within each equation apply to all cells, we use matrix notation. For example, \(\mathbf{x}\) stands for the set of cells at every spatial location \((i, j)\) in the input display. Convolution between a matrix \(\mathbf{x}\) and kernel \(\mathbf{F}\), specified by the * operator, is always centered at each cell position \((i, j)\). The notation \([.]^+\) denotes the half-wave rectification \(\max(., 0)\). All DDEs were numerically integrated using explicit Euler's method with a fixed time step of 1 msec. Our simulations were performed on a 2.66 Ghz 8-core Apple Mac Pro with 64 GB RAM in Wolfram Mathematica 8.

Model neurons obey equations that feature shunting competitive dynamics (Grossberg, 1968). These equations perform a leaky integration of their inputs and simulate many known properties of neurons, such as divisive normalization (Heeger, 1992; Carandini and Heeger, 2011) and automatic gain control (Grossberg, 1983).
equations generally resemble the following membrane equation, termed a \textit{recurrent competitive field} (Grossberg, 1973):

\[
\dot{x}_i = -\alpha x_i + (\beta - x_i)(f(x_i) + I_i^+) - x_i \sum_{k \neq i} (f(x_k) + I^-). \tag{5.2}
\]

Eq. 11.2 is a shunting equation that describes the activity, \(x\), of the \(i^{th}\) cell in a neural network layer. The parameters \(\alpha\) and \(\beta\) define the passive decay rate and saturation upper bound, respectively. The terms \((\beta - x_i)(f(x_i) + I_i)\) and \(-x_i \sum_{k \neq i} f(x_k)\) of Eq. 11.2 specify the shunting excitation by input \(I_i^+\) and surround inhibition, respectively. The variable \(I^-\) represents inhibitory input the cell may receive. In Eq. 11.2, \(f(x)\) is a \textit{signal function} that specifies the nature of the feedback from cells in the same network layer. We often configure \(f(x)\) as a faster-than-linear function, which results in a contrast-enhancing network. Eq. 11.2 differs from typical model equations in that time delay \(\tau\) is omitted.

\textbf{Model Photoreceptors and Retinal Ganglion Cells}

When photoreceptors are stimulated by pulses or steps of increasing light intensities, their graded potential is proportional to the change in light intensity (Baylor et al., 1974). While some retinal ganglion cells in primates have also been observed to respond to changes in luminance intensity, others respond tonically to constant luminance intensities (Schiller and Malpeli, 1977; Kayama et al., 1979; DeYoe and Bartlett, 1980). We model both populations of retinal ganglion cells, which affords sensitivity to a range of spatio-temporal correlations, including stationary luminance patterns. We assume both subpopulations in the model receive input from change-sensitive photoreceptors, with the \textit{tonic retinal ganglion cell} \(a_p\) subpopulation directly integrating the input signal \(I\) and the \textit{transient retinal ganglion cell} subpopulation \(a_m\) responding to the change detected by the photoreceptors. The tonic and transient
subpopulations are indexed by \( p \) and \( m \) because these cells form the first segregation between parvocellular and magnocellular pathways in the model, respectively. Model tonic retinal ganglion cells code stationary input patterns with high fidelity, while model transient cells are simple change-sensitive units that respond well to movement.

\[
a_p(t) = \begin{cases} 
I(t) & \text{for } |r| \leq \Omega \\
0 & \text{otherwise}
\end{cases} \tag{5.3}
\]

\[
a_m(t) = \begin{cases} 
[I_r(t) - I_r(t - \eta)]^+ & \text{for } |r| \leq \Omega \\
0 & \text{otherwise}
\end{cases} \tag{5.4}
\]

Eqs. 5.3 & 5.4 compute the model tonic and transient retinal ganglion cell responses \( a_p \) and \( a_m \), respectively, where \( I \) corresponds to the video input at time \( t \), \( \eta \) defines how large a temporal interval over which the units \( a_m \) should integrate in the recent past, \( r = (x, y) \) specifies the 2D coordinate that marks the receptive field center of the retinal ganglion cell, and \( \Omega = 1 \) denotes a radius 1 spatial region (eight-neighborhood) about each position \( (I_{i,j}) \). For simplicity, we set \( \eta = 5 \) msec, which is the duration of each video frame. The transient retinal ganglion cell at position \((i,j)\) elicits a response equal to the change in luminance of the input if change is detected within a one pixel radius at time \( t \). Model transient retinal ganglion cells \( a_m \) are half-wave rectified to yield a response only when the luminance intensity increases, which is consistent with the behavior of ON retinal ganglion cells (Schiller, 1996). Our model only uses the ON class of units, but OFF units may be also be included (Baloch et al., 1999).

Because the retinal ganglion cell equations thus far give square wave responses during each video frame, we apply the following exponential moving average on the
retinal ganglion cell activity:

\[ n(t) = n(t - \eta) + \nu [a(t) - n(t)] \quad (5.5) \]

The exponential moving average in Eq. 5.5 temporally smooths the retinal ganglion cell responses yielding \( n(t) \), which gives higher weight to more recent responses and less weight to present responses. The parameter \( \nu \), which may assume values between 0 and 1, differentially weights the degree to which new information from the signal at the present time contributes to the temporally accumulated signal. The weights of retinal ganglion cell responses decay exponentially, with higher values for \( \nu \) implicating faster decay rates. For example, \( \nu = 1 \) gives the original retinal ganglion cell time series and \( \nu = 0 \) allows no new information to contribute to the temporal average, resulting in a constant signal. We apply the same smoothing to both tonic and transient retinal ganglion cells with \( \nu = 0.3 \).

**Model LGN: Transient Cells**

Model tonic and transient retinal ganglion cells directly project to corresponding parvocellular \( x_p \) and magnocellular \( x_m \) LGN units, which follow the same transient behavior observed in neurophysiological studies (Hammond, 1974; Hochstein and Shapley, 1976). Model LGN cells \( x \) perform a leaky integration of their retinal ganglion cell inputs and possess habituative gates \( z \) (Grossberg, 1980), which show activity-dependent changes in synaptic efficacy.

\[ \dot{x} = \epsilon_{LGN} (-\alpha_{LGN} x + (\beta_{LGN} - x) \odot n) \quad (5.6) \]

Eq. 5.6 specifies the dynamics of transient LGN cells \( x \), where \( \epsilon_{LGN} \) is inversely proportional to the cell time constant and scales the temporal sensitivity of the LGN response and \( \odot \) signifies element-wise multiplication. In all simulations we set \( \epsilon_{LGN} \)
= 5, resulting in a 20 msec time constant, which is consistent with the temporal responses of primate LGN cells (Maunsell et al., 1999). The parameter \( \beta \) is fixed at 1. To account for the qualitative differences in the decay rates of parvocellular and magnocellular cells in primate LGN, model parvocellular \( x_p \) and magnocellular \( x_m \) LGN cells have different decay rates of \( \alpha_{LGN} = 2 \) and \( \alpha_{LGN} = 40 \), respectively. The following equation describes the dynamics of the slow-adaptation habituative gates that act postsynaptically on LGN cell responses.

\[
\dot{z} = \epsilon_{LGN} (1 - z - \kappa_{LGN}(x \odot z))
\] (5.7)

Eq. 5.7 defines the dynamics of habituative gates \( z \), which models synaptic depression due to neurotransmitter depletion at the synapse. The parameter \( \kappa_{LGN} \) determines the neurotransmitter depletion rate, which is proportional to the LGN activity \( x \). When \( x \) does not respond (is equal to zero), the synaptic efficacy recovers to 1 at rate \( \kappa_{LGN} \). Therefore, the postsynaptic LGN signal \( L = [x \odot z]^+ \) that in the model projects to V1 elicits a sharp transient response to visual inputs, with different decay behavior within parvocellular and magnocellular subpopulations, as shown in Figure 5·3. We set \( \kappa_{LGN} = 40 \) in all simulations.

**Model V1: Spatio-temporal Correlation Detection**

Each V1 cell receives feedforward input from a set of LGN cells denoted \( L^{\{\tilde{\delta}_1, \ldots, \tilde{\delta}_n\}} \). This set of LGN cells have receptive fields near the V1 cell receptive field center at location \((i, j)\), but the LGN receptive field centers are displaced by small amounts \( \{\tilde{\Delta}_1, \ldots, \tilde{\Delta}_n\} \). These LGN cell form the V1 subunits in the model V1 cell. Our model also considers a set of conduction delays between LGN and V1, denoted \( \{\delta_1, \ldots, \delta_n\} \). In simulations, we assume each V1 cell receives convergent input on its dendrites from two sets of LGN cells with relative receptive field displacements of \( \{\tilde{\Delta}_1, \tilde{\Delta}_2\} \). Each V1
cell receives input from pairs of LGN cells with different conduction delays \( \{\delta_1, \delta_2\} \) that have receptive fields spatially displaced by \( \Delta_i \). Hence, each V1 cell receives projections from a total of four LGN cells, with slightly different receptive field centers and conduction delays. For simplicity, we fix \( \delta_1 = 0 \) for instantaneous conduction, and \( \delta_2 = 5 \) msec, which is equal to the duration of each frame of video input. The values for \( \{\Delta_1, \Delta_2\} \) thereby determine the spatio-temporal correlation tuning of the V1 cell. For example, setting \( \Delta_1 = \{(0,0), (1,0)\} \) and \( \Delta_2 = \{(0,1), (1,1)\} \) in Cartesian coordinates will yield a V1 cell sensitive to unit speed leftward correlations. As this example shows, each \( \Delta_i \) contributes some redundancy in the correlation detection over different sets of LGN cells with proximal receptive fields locations, which augments the confidence in the motion detection. We had the best success selecting the relative orientation of \( \Delta_1 \) and \( \Delta_2 \) as orthogonal to the direction of motion. This choice enhances the detection of motion near spatio-temporal discontinuities or kinetic edges.

The correlation detection for input from the LGN units that arrives at time \( t \) on the dendrites of each V1 cell is computed according to the following multiplicative equation:

\[
C(t) = \frac{L^{\{\Delta_1, \ldots, \Delta_n\}}(t - \delta m) \cdot L^{\{\Delta_i, \ldots, \Delta_n\}}(t - \delta n)}{\|L^{\{\Delta_1, \ldots, \Delta_n\}}(t - \delta m)\| \odot \|L^{\{\Delta_i, \ldots, \Delta_n\}}(t - \delta n)\|}
\tag{5.8}
\]

In Eq. 5.8, \( C(t) \) is the correlation that is computed from the set of LGN cells with different receptive field locations scattered about the V1 cell receptive field center and conduction delays. The \( \cdot \) operator represents a dot product and \( \|\cdot\| \) indicates the L2 (Euclidean) norm. We divide by the magnitude of set of spatially displaced LGN cells in Eq. 5.8 to normalize by the magnitude of the transient response to contrast so that the V1 response is not confounded by luminance magnitude.

We simulate V1 units tuned to five different spatio-temporal correlations: stationary, up, down, left, and right. For the purposes of the present visual displays, we
assume dots within the random texture surfaces move at unit speed. The V1 units sensitive to stationary correlations receive input from parvocellular LGN, while V1 units sensitive to motion patterns receive input from magnocellular LGN. In both cases, the correlation within the magnocellular and parvocellular pathways is computed using Eq. 5.8.

The following shunting equation specifies the dynamics of model cells $v_k$ in V1 that are sensitive to correlation $k$.

$$\dot{v}_k = \epsilon_{V1}(-\alpha_{V1}v_k + (\beta_{V1} - v_k) \odot (v_k^2 + C_k(t)) - v_k \odot \sum_{m \neq k} v_m^2) \quad (5.9)$$

Eq. 11.4 describes the dynamics of V1 cells sampling the same visuotopic location that compete across correlation via the term $-v_k \sum_{m \neq k} v_m^2$ in a recurrent fashion. Each V1 cell receives a scalar value, $C_k$, as input from LGN units whose axons converge onto the dendrite of the V1 cell. In all simulations, we fix $\epsilon_{V1} = 5$, $\alpha_{V1} = 1$, and $\beta_{V1} = 2$.

**Model V2: Correlation-Sensitive Border-Ownership Cells**

Model parvocellular and magnocellular border-ownership cells are sensitive to correlation $k$ and have side-of-figure selectivity $\theta$ defined by Eqs. 5.10 & 5.11. We implemented four border-ownership directions ($\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$): right, up, left, and down. Note that the feedforward signal from V1 does not directly determine the border-ownership signal of each border-ownership cell — feedback from cells with larger receptive fields in MT and V4 are required by the model.

$$\dot{b}_{k,\theta}^p = \epsilon_{V2}(-\alpha_{V2}b_{k,\theta}^p + (\beta_{V2} - b_{k,\theta}^p) \odot (\chi_{V2}G_{V2} * v_k) - \gamma_{V2}(b_{k,\theta}^p)$$

$$\odot \left( \sum_s \sum_{r \neq k} K_{\theta+\pi}^s * m_{r}^s + \sum_{s} K_{\theta}^s * g_{s} \right) \quad (5.10)$$
Parvocellular \( b_{k,\theta}^p \) and magnocellular \( b_{k,\theta}^m \) B cells are defined in Eq. 5.10 & 5.11, respectively. B cells that are sensitive to correlation \( k \) in both pathways integrate V1 correlation cells also sensitive to correlation \( k \) over a local spatial area, determined in our simulations by the 2D Gaussian kernel \( G_{V2} \), that has radius 2, \( \sigma_{V2} = 1 \), and \( \chi_{V2} = 5 \). We determined \( \chi_{V2} \) by approximately setting the center weight to unity. The parameter \( \gamma_{V2} \) balances the ratio between excitatory feedforward to inhibitory feedback inputs. In all our simulations, we set \( \epsilon_{V2} = 5 \), \( \alpha_{V2} = 1 \), \( \beta_{V2} = 1 \), \( \gamma_{V2} = 5 \). The matrices \( K_{\theta}^s \) are the right, top, left, or bottom halves of the V4 or MT annular receptive field (depending on \( \theta \)) with receptive field size \( s \). More details about the annular kernels are defined below. Parvocellular B cells \( b_{k,\theta}^p \) with inward side-of-figure selectivities toward the MT cell receptive field center receive inhibition from MT cells \( m_r \) that are sensitive to different correlations \( (r \neq k) \) via the term \( \sum_s \sum_{r \neq k} K_{\theta+\pi}^s \cdot m_r^s \) in Eq. 5.10. In addition, the term \( \sum_s K_{\theta}^s \cdot g^s \) indicates that V4 cells send inhibitory feedback to parvocellular B cells whose border-ownership selectivity is directed \( 180^\circ \) away from the center of their annular receptive field.

The term \( \sum_s \sum_{r \neq k} K_{\theta+\pi}^s \cdot m_r^s \) in Eq. 5.11 indicates that MT cells send nonspecific inhibition to V2 B cells with inward border-ownership selectivities toward the MT receptive field center that have different correlation preferences than the MT cell. The term \( \sum_s K_{\theta}^s \cdot m_k^s \) specifies that MT cells inhibit magnocellular B cells with side-of-figure selectivities that point \( 180^\circ \) away from the center of the MT receptive field and the same correlation preference. Finally, the term \( \sum_s \sum_{n=(\theta|k,\theta)} K_{n+\pi}^s \cdot g^s \) specifies that only magnocellular B cells are inhibited by V4 cells if the correlation
preference implicates motion orthogonal to the annulus.

Model MT and V4: Large-Field Correlation Cells

Figure 5·6: Schematic depiction of the annular, “donut-shaped”, radially-symmetric, receptive field kernels of MT and V4 units (right hand side). Each MT and V4 unit performs a nonlinear integration of magnocellular and parvocellular B cells, respectively, in V2 that appear within the different “fragments” of the annulus ($K_{\pi}, K_{\frac{3\pi}{2}}, K_0, K_{\frac{\pi}{2}}$). All pairs of fragments are convolved with B cell populations and are multiplicatively combined. The input to V4 and MT is proportional to the sum of products of these convolutions. See Eq. 12 for details.

Following V2, the parvocellular and magnocellular pathways diverge: parvocellular B cells project to model V4 and magnocellular B cells project to MT. Since V2 cells that are sensitive to stationary correlations project to V4, V4 also prefers stationary correlations. Similarly, each set of magnocellular B cells in V2 independently projects to a different set of MT cells and MT cells develop the same correlation selectivities. As shown in Figure 5·5, both V4 and MT cells have receptive fields that are substantially larger than those in V2. We simulate two different receptive field sizes to integrate spatio-temporal correlations over a large area: one with a diameter of 3 px and another with a diameter of 5 px. We selected these receptive field sizes to pool over a large enough number of V2 units of a particular correlation, yet be small enough to be sensitive to correlation discontinuities (kinetic edges).

Both model V4 and MT cells perform a nonlinear grouping of their inputs within
their annular receptive fields. Subunits take the form of multiplicative combinations of pairs of half-annuli $K^{s}_{\theta}$ (see Figure 5·6). The parameter $s$ indicates the size of the receptive field and $\theta$ specifies which half of the annulus is extracted. For example, $\theta = 0$ indicates the right half of the annulus. Multiplicative combinations enhance the detection of correlations that exist on multiple sides of the annulus. Therefore, this nonlinear combination facilitates the detection of shapes defined by differential correlations in kinetic random-dot displays, such as the rectangular shapes used by Royden and the square shapes used by von der Heydt and colleagues. We construct annular kernels by taking a difference-of-Gaussians (i.e. Mexican hat) between kernels $F^{G}_{1}$ and $F^{G}_{2}$ with radii $r^{G}$ and standard deviations $\sigma^{G}_{1}$ and $\sigma^{G}_{2}$, respectively. We then half-wave rectify to preserve non-negative components and normalize the resulting annular kernel. For the Gaussians $F^{G}_{1}$ and $F^{G}_{2}$, we constrain the ratios between the kernel radii and standard deviations by setting $\frac{r^{G}_{1}}{\sigma^{G}_{1}} = 2$ and $\frac{r^{G}_{2}}{\sigma^{G}_{2}} = 2.22$.

Model MT cells $m^{s}_{k}$ receive excitatory input from magnocellular B cells that are sensitive to the same correlation within their annular receptive fields, and send feedback to the same correlation B cells that have side-of-figure preferences that point $180^\circ$ away from the annular center. MT cells also send inhibitory feedback nonspecifically to B cells sensitive to all other correlations that have side-of-figure selectivities toward the annular center. This connectivity affords enhancement of inward border-ownership responses toward surfaces that have the same uniform correlation as the MT cell. Model V4 cells receive excitatory input from parvocellular B cells and send feedback to parvocellular B cells that have side-of-figure preferences that point $180^\circ$ away from the annular center. Model V4 cells have similar connectivities, except V4 cells only inhibit magnocellular B cells within the annular receptive field that have correlations sensitivities orthogonal in direction with respect to the annulus. Because model V4 cells receive substantial input from the parvocellu-
lar pathway, they respond well to stationary patterns. *Our model predicts that V4 cells are sensitive to the ecological tendency that a stationary surface is the figure when texture accretion/deletion occurs at kinetic edges on its boundary, which do not move over time.* Therefore, while our model V4 cells are primarily driven by parvocellular B cells, each V4 cell is selectively enhanced by B cells that respond to dots that move orthogonally with respect to the annulus, which signals the presence of texture accretion/deletion. The term $q_{P_{m,n}}(b_{r,\theta_m} \star K_{\theta_m}^s) \circ (b_{r,\theta_n} \star K_{\theta_n}^s)$ in both equations takes the nonlinear multiplicative sum of all permutations of pairs $(m, n)$ of annuli halves $K_s^s$ and applies a square root compressive nonlinearity. In Eq. 5.13, the last term performs an integration of magnocellular B cells, which have correlation sensitivities that afford motion orthogonal to the annulus, but it is gated by the parvocellular input, where $k = r$ is the stationary correlation.

$$\dot{m}_{k}^{s} = \epsilon_{MT} \left( -\alpha_{MT}m_{k}^{s} + (\beta_{MT} - m_{k}^{s}) \circ \sum_{m,n}(b_{k,\theta_m} \star K_{\theta_m}^s) \circ (b_{k,\theta_n} \star K_{\theta_n}^s) \right) \quad (5.12)$$

$$\dot{g}^{s} = \epsilon_{V4}(-\alpha_{V4}g^{s} + (\beta_{V4} - g^{s}) \circ \sum_{m,n}(b_{r,\theta_m} \star K_{\theta_m}^s) \circ (b_{r,\theta_n} \star K_{\theta_n}^s) \circ \left( 1 + \sum_{p\neq r} \sum_{\{(m,n)\mid (\theta_m,\theta_n) \perp p\}}(b_{p,\theta_m} \star K_{\theta_m}^s) \circ (b_{p,\theta_n} \star K_{\theta_n}^s) \right) \right) \quad (5.13)$$

In Eqs. 9.5 & 5.13, we set $\epsilon_{MT} = \epsilon_{V4} = 5$, $\alpha_{MT} = \alpha_{V4} = 2$, and $\beta_{MT} = \beta_{V4} = 1$. The term $\sum_{m,n}(b_{r,\theta_m} \star K_{\theta_m}^s) \circ (b_{r,\theta_n} \star K_{\theta_n}^s)$ in both equations takes the nonlinear multiplicative sum of all permutations of pairs $(m, n)$ of annuli halves $K_s^s$ and applies a square root compressive nonlinearity. In Eq. 5.13, the last term performs an integration of magnocellular B cells, which have correlation sensitivities that afford motion orthogonal to the annulus, but it is gated by the parvocellular input, where $k = r$ is the stationary correlation.
Vectorial Modulation Index

We use the vector modulation index $\mathbf{v}_k = (m_{k,x}, m_{k,y})$ for correlation $k$, to relate our B cell responses to those reported in other studies (Craft et al., 2007; Mihalas et al., 2011):

$$m_{k,x} = \frac{b_{k,0} - b_{k,\pi}}{b_{k,0} + b_{k,\pi}}$$  \hspace{1cm} (5.14)

$$m_{k,y} = \frac{b_{k,\pi/2} - b_{k,3\pi/2}}{b_{k,\pi/2} + b_{k,3\pi/2}}$$  \hspace{1cm} (5.15)

Due to the normalizing difference over a sum of border-ownership cell activity with anti-preferred directions of border-ownership, each component lies between -1 and 1. Negative values correspond to computed border-ownership in the leftward and downward directions for $(x, y)$ components at each spatial location $(i, j)$, respectively. By contrast, positive values correspond to computed border-ownership in the rightward and upward directions for the $(x, y)$ components. A zero valued component indicates an indifference in border-ownership in that particular axis.

Algorithmic Correlation Detection Frontend

We were able to obtain our results both using the aforementioned biological model for motion detection and a computationally quicker algorithmic implementation using the raw video pixels as inputs. Using the input video $J(t)$ at frame $t$, we can compute the necessary correlations to provide input to model V1 (Eq. 11.4) via the following equation, which is similar to Eq. 5.8.

$$c(t) = \frac{J^{\Delta_1,\ldots,\Delta_n}(t) \cdot J^{\Delta_1,\ldots,\Delta_n}(t-1)}{\|J^{\Delta_1,\ldots,\Delta_n}(t)\| \odot \|J^{\Delta_1,\ldots,\Delta_n}(t-1)\|}$$  \hspace{1cm} (5.16)

In Eq. 5.16, each spatial displacement $\Delta_i$ in $J^{\Delta_1,\ldots,\Delta_n}$ specifies how many vertical and horizontal pixels are shifted between successive frames $t$ and $t - 1$. As in Eq. 5.8,
\( J \{ \Delta_i \ldots \Delta_n \} \) indicates that \( n \) local pixel neighborhoods should be grouped (in our simulations \( n = 2 \)) and then correlated via the dot product across successive frames. Although false matches occur, as in the biological implementation, the correlation yielded for surfaces with uniformly moving dots will equal 1 when the match is correct. In the algorithmic implementation, \( c_k(t) \) replaces \( C_k(t) \) in Eq. 11.4 and the rest of the model remains the same.

### 5.3 Results

#### 5.3.1 Stationary condition

Figure 5.7 shows simulation results for several types of kinetic random dot displays, each of which contains a figure surface (left in \( a, b, e \) and top in \( c \)) that is seen by human observers as closer in depth than a ground surface (right in \( a, b, e \) and bottom in \( c \)). Figure 5.7a summarizes the model performance in the stationary condition. The top panel schematically depicts the displacement of the random dots in the display over two consecutive frames of video. As indicated by the arrows, dots on the ground
**Figure 5-7:** Simulations demonstrate border-ownership assignment of the kinetic edge (green dashed line) to the figure surface in different kinetic random dot display conditions. Top panels schematically depict how dots within the figure and ground surfaces are displaced between successive frames in the kinetic random dot visual displays. The bottom panels depict the border-ownership vectorial modulation index produced by model border-ownership units, a normalized measure of border-ownership activity generated by border-ownership units between ±1. Vectorial modulation indices of ±1 correspond to strong border-ownership signals to the left and right of the border-ownership unit’s receptive field that elicited the signal, respectively, and a zero value indicates no border-ownership of the kinetic edge near the unit’s receptive field. The black, red, and blue curves correspond to the vectorial modulation indices produced by border-ownership units sensitive to stationary, rightward, and leftward spatio-temporal correlations, respectively. **(a) Stationary condition.** A subpopulation of borderownership units that prefers stationary correlations (black) produce a larger peak than another subpopulation of units sensitive to leftward (blue) correlations (see inset for magnitude comparison). Because the black peak is negative and larger than the positive blue peak, borderownership of the kinetic edge is assigned to the figure surface (left). **(b) Encroach condition.** A subpopulation of border-ownership units that prefer rightward correlations (red) produces a larger peak than another subpopulation of units sensitive to stationary (black) correlations. Because the red peak is negative and larger than the positive black peak, border-ownership of the kinetic edge is assigned to the figure surface (left). **(c) Shear condition.** A subpopulation of border-ownership units that prefer rightward correlations (red) produces a larger peak than another subpopulation of units sensitive to stationary (black) correlations. In the bottom panel, the green dashed line marks the position of the horizontal kinetic edge in the display, and vectorial modulation indices of ±1 corresponds to up and down border-ownership, respectively. Because the red peak is positive and larger than the negative black peak, border-ownership of the kinetic edge is assigned to the figure surface (top). **(d) Degenerate condition.** A kinetic edge divides the left and right regions but neither region is seen as figure or ground (Kaplan, 1969). Border-ownership unit subpopulations sensitive to correlations present in the display (red and blue) yield on average equal but opposite sign signals. Hence, the model does not assign ownership of the kinetic edge to either side. **(e) Encroach condition (motion in ‘A’ and ‘B’).** Border-ownership of the kinetic edge to the figure surface in the encroach condition (first panel) is assigned throughout the beginning (second panel), middle (third panel), and end (fourth panel) of the display presentation, and the border-ownership signals dynamically move with the kinetic edge.
surface advance by one pixel toward the figure surface and dots that collide with the
figure surface disappear. Dots in the figure surface remain stationary over time. The
dashed green line specifies the location of the kinetic edge, which also remains sta-
tionary over time in Figure 5·7a and only is seen by human observers when the dots
are in motion. The bottom panel shows the border-ownership vectorial modulation
index (see Methods) of border-ownership units in model V2, averaged across hori-
zontal cross-sections during the final video frame. The x-axis marks individual pixel
locations along the averaged cross-section, the dashed green line signifies the kinetic
edge position, and the y-axis shows the border-ownership vectorial modulation index.
Negative values correspond to a border-ownership signal toward the left, and posi-
tive values correspond to border-ownership signals to the right — all relative to the
receptive field location of the border-ownership unit eliciting the response. An index
closer to ±1 indicates that the unit is eliciting a stronger border-ownership signal in
that respective direction (e.g. left or right). A zero valued vectorial modulation index
indicates the lack of a border-ownership signal in either direction. The red and blue
curves plot the averaged cross-sectional border-ownership vectorial modulation index
of magnocellular border-ownership units that prefer left and right spatio-temporal
correlations, respectively. The black curve shows the averaged cross-sectional border-
ownership vectorial modulation index of parvocellular border-ownership units that
prefer static spatio-temporal correlations. In the stationery condition (Figure 5·7a),
consistent with human judgments that the left surface is figure, the peak of the black
curve, which occurs near the kinetic edge, is larger than the nearby peak produced by
the blue curve — indicating that the stationary border-ownership subpopulation in
model V2 more strongly signals ownership to the left than the leftward-correlation-
sensitive border-ownership unit subpopulation signals ownership to the right. The
inset compares the absolute value of the averaged border-ownership modulation in-
dices plotted in the bottom panel. In the model, the subpopulation of B cells that has the largest absolute magnitude assigns border-ownership of the kinetic edge toward the region with the texture that moves according to its preferred spatio-temporal correlation. Because the black peak is much higher than the blue peak, the model assigns border-ownership of the kinetic edge to the figure (left) surface.

5.3.2 Encroach condition

Figure 5·7b (bottom panel) shows the cross-sectional border-ownership vectorial modulation indices averaged during the final frame of the encroach condition, wherein dots in the figure surface (left) move toward and encroach on the ground surface (right) between consecutive video frames. When dots on the figure surface collide with others in the ground surface, the dots belonging to the ground surface are deleted and replaced with those belonging to the figure surface (top panel). As shown in the inset, the magnitude of the peak signal produced by the border-ownership unit subpopulation that prefers rightward correlations (red) is greater than that of the border-ownership unit subpopulation that prefers stationary correlations (black). Because the red peak is negative, the model demonstrates border-ownership of the kinetic edge toward the figure surface (left).

5.3.3 Shear condition

Figure 5·7c depicts the model results for the shear condition, wherein dots in the figure surface (top) move parallel to the kinetic edge that separates the figure and ground surfaces and the dots in the ground surface (bottom) remain stationary across successive video frames. The bottom panel shows the border-ownership modulation indices of border-ownership units in model V2 averaged across vertical cross-sections of the visual display. The dashed green line depicts the position of the horizontally-oriented kinetic edge, and locations to the left and right along the x-axis correspond
to locations in the figure (top) and ground (bottom) surfaces, respectively. The peak in the mean border-ownership vectorial modulation index of the border-ownership unit subpopulation sensitive to rightward correlations (red) is more positive than the peak produced by the competing subpopulation sensitive to stationary correlations (black) is negative. The inset shows a comparison between the peak magnitudes. Because the red peak is higher than the black peak, and the red peak is negative, the model results indicate that the kinetic edge is owned by figure surface (top).

Figure 5·7d shows a simulation of a case tested by Kaplan in which humans did not report the presence of depth ordering and saw a pencil-thin kinetic edge owned by neither surface. Model B cells that are sensitive to rightward (red) and leftward (blue) correlations present in the display elicit on average equal magnitude but opposite sign peak border-ownership vectorial modulation indices, as shown in the inset. Therefore, neither subpopulation assigns border-ownership to the kinetic edge.

Figure 5·7e shows the averaged border-ownership model output at the beginning (second panel), middle (third panel), and end (fourth panel) of an encroach condition video sequence wherein the figure surface (left) encroaches upon the ground surface (right) and dots in both surfaces move toward one another (first panel). Throughout the simulation, the peak vectorial modulation index of the border-ownership subpopulation that prefers rightward correlations (red) is more negative than the peak vectorial modulation index of the competing border-ownership subpopulation that prefers leftward correlations (blue) is positive. The insets in each panel depict comparisons of the absolute magnitudes of the peaks of the competing V2 border-ownership unit subpopulations. Because the red peak magnitude is always larger than the blue peak, and the red peak is negative, the model indicates that the kinetic edge is owned by the figure surface (left). Note that the border-ownership peaks dynamically move with the kinetic edge over time.
Figure 5-8: In the strip condition, the model border-ownership units assign border-ownership of the two kinetic edges to the figure surface, regardless of whether it is in the center (a) or periphery (b). (a) and (b) differ only based on whether there is rightward dot motion in the two peripheral regions (a) or in the center (b). Texture accretion/deletion occurs at the kinetic edges in both cases. (a) The border-ownership subpopulation that prefers stationary correlations (black) produces two peaks: one positive and one negative, indicating inward ownership of the two kinetic edges. The border-ownership subpopulation that prefers rightward correlation (red) produces two peaks: one negative and one positive, indicating outward ownership of the two kinetic edges. Because the black inward ownership peaks are larger in magnitude than the red ones, the model assigns ownership of the kinetic edges to the figure surface (center). (b) Similar to (a) except the larger black peaks indicate outward ownership of the kinetic edges to the peripheral surfaces.
5.3.4 Strip condition

Figure 5-8a-b show model border-ownership results in the *strip condition*. In Figure 5-8a, the stationary dots in the figure surface (center) are seen by human observers as in front of rightward moving dots in the ground surface (peripheral regions). The position of the two kinetic edges remain stationary throughout the video. The bottom panel of Figure 5-8a depicts the vectorial modulation index of model border-ownership units averaged over horizontal cross-sections. The border-ownership unit subpopulation that prefers stationary correlations (black) produces two peaks within the center region defined by the left and right kinetic edges: the positive and negative peaks to the left and right correspond to border-ownership of the left and right kinetic edges, respectively, by the figure surface (center). The competing border-ownership unit subpopulation also elicits two peaks: negative and positive ones to the left and right border-ownership assignment of the left and right kinetic edges, respectively, to the ground surface (periphery). The peaks produced by border-ownership units sensitive to stationary correlations (black) are larger than those produced by units sensitive to rightward correlations (red). The presence of a positive black peak near the left kinetic edge and a negative black peak near the right kinetic edge indicates *inward* ownership of the kinetic edges toward the figure surface (center). The presence of a negative red peak near the left kinetic edge and a positive red peak near the right kinetic edge indicates *outward* ownership of the kinetic edges toward the ground surface (periphery). As shown in the insets, the magnitudes of the black inward border-ownership peaks are higher than those of the red outward border-ownership peaks. This demonstrates that the border-ownership of the kinetic edges is assigned to the figure surface.

The strip condition in Figure 5-8b is the same as Figure 5-8a, except dots in the center region move rightward and dots in the peripheral regions remain stationary. Dots are accreted and deleted upon reaching the kinetic edges at the boundary.
between the center and peripheral regions. The bottom panel shows that the border-ownership unit subpopulation that prefers rightward correlations (red) produces two peaks inside the region defined by the two kinetic edges: a positive peak near the left kinetic edge and a negative peak near the right kinetic edge. The competing subpopulation of border-ownership units that prefer stationary correlations (black) produces two peaks as well: a negative peak near the left kinetic edge and a positive peak near the right kinetic edge. Because both black peaks (outward border-ownership) are greater in magnitude than those that are red (inward border-ownership), the model assigns border-ownership of the kinetic edges to the figure surface (periphery).

5.3.5 Object & Window conditions

Figure 5·9a-b show model border-ownership results for the object and window conditions, respectively. In Figure 5·9a, dots within the centrally located square region (figure) appear in front of a surrounding annular region (ground) composed of rightward moving dots (top panel). The kinetic edges at the four central square region boundaries, are labeled by the green dashed lines. The bottom panel shows border-ownership vectorial modulation indices averaged across horizontal cross-sections where there are vertically-oriented kinetic edges. Similar results were obtained when averaging over vertical cross-sections. The border-ownership unit subpopulation that prefers stationary correlations generated two peaks (black): positive and negative ones near the left and right kinetic edges, respectively. Because these peaks are greater in magnitude than other peaks nearby each respective kinetic edge, the two black peaks indicate inward ownership of the kinetic edges by the center square region.

Figure 5·9b shows the model simulation results for the window condition, wherein rightward moving dots within the centrally located square region (ground) appear as moving behind a surrounding annular region (figure) composed of stationary dots (top panel). As in the object condition, the four kinetic edges remain stationary
Figure 5.9: Model border-ownership assignment in the object (a) and window (b) conditions along horizontal cross-sections of the visual display. (a) In the object condition, the center square and surrounding regions represent the figure and ground surfaces, respectively. The border-ownership subpopulation (black) that prefers stationary correlations elicit two peaks that are larger than any others produced by competing subpopulations: a positive and negative peak near the left and right kinetic edges, respectively. Because the black outward peaks are the largest in magnitude nearby each of the kinetic edges, border-ownership of the kinetic edges is assigned to the figure surface (square). (b) The square and surrounding regions serve as the ground and figure surfaces in the window condition, respectively. As in (a) the stationary subpopulation of border-ownership units elicit the largest magnitude peaks, but unlike (a), these peaks indicate outward border-ownership. Hence, the model assigns border-ownership of the kinetic edges to the surrounding surface (figure).
throughout the display presentation. In contrast to the object condition, the sub-population of border-ownership units that prefer stationary correlations generate a negative peak near the left kinetic edge and a positive peak near the right kinetic edge (black). These outward ownership peaks are greater in magnitude than any of the peaks produced by border-ownership units sensitive to other correlations, and therefore the model assigns border-ownership of the kinetic edges to the surrounding surface rather than the central square surface.

Figure 5-10: (a) A comparison between the proportion of the time human subjects responded that there was a depth ordering between two surfaces in the gap condition for different gap sizes (replotted from Yonas et al., 1987) and (b) the magnitude of border-ownership vectorial modulation indices yielded in model simulations. (a) As the size of the gap between the left and right surfaces increases, subjects are less likely to indicate that there is a depth ordering in the visual display. (b) In model simulations of the gap condition, the border-ownership vectorial modulation index magnitudes declined as the gap size increased until the receptive field size of a MT/V4 unit was reached (5 px), at which point the magnitude dropped to zero, which indicates no border-ownership assignment. The zero vectorial modulation index value is analogous to the 50% chance criterion level in Yonas et al., (1987) where human subjects were equally likely to indicate the presence or absence of a depth ordering in the visual display.
5.3.6 Gap condition

Figure 5·10 compares human figure-ground judgments (Figure 5·10a) with border-ownership signals produced by the model in the gap condition (Figure 5·10b). The human data are replotted (Yonas et al., 1987). The study introduced small, medium, and large-sized gaps to displays that are similar to the encroach condition (Figure 5·7b, Figure 5.11b), except texture accretion/deletion between the left and right surfaces did not occur. Subjects were asked whether they perceived a depth ordering in the displays. Figure 5·10a shows the percent of the time subjects indicated that a depth ordering was present. 100% indicates a depth ordering between the left and right surfaces always was perceived and 50% indicates that subjects were equally likely to see or not see a depth ordering. When there was accretion/deletion (no gap), human subjects saw a depth ordering between the left and right surfaces 98.8% of the time. When a small, medium, and large sized gap was introduced, humans saw a depth ordering 87.2%, 78.8%, and 62.5% of the time, respectively. As the gap size increased, human observers were less likely to perceive a depth ordering between the surfaces, and approached chance performance for large gap sizes. Figure 5·10b shows the peak magnitude of the vectorial modulation index averaged over horizontal cross-sections (as in Figure 5·7 and Figure 5·8) during the entire trial of the maximally active border-ownership unit subpopulation in model V2. As the gap size between the moving surfaces increased, the model yielded lower amplitude border-ownership signals until the gap size exceeded the MT and V4 receptive field diameters (5 px), at which point the average border-ownership signal dropped to zero. Analogous to the 50% chance criterion level in the experiment of Yonas, a zero border-ownership magnitude model response corresponds to the border-ownership unit signaling ambivalence with respect to the direction of border-ownership. Both the human data and model demonstrate reductions of confidence in border-ownership as the gap size increases.
Although the human psychophysical and model result curves decrease with larger gap sizes, they do not agree quantitatively. It is not clear exactly how psychophysically derived figure-ground judgments map onto border-ownership signals and vice versa, as other factors are likely involved. However, the decrease seen in both curves supports the hypothesis that figure-ground judgments depend on neurons with limited receptive field sizes.

5.3.7 Temporal Dynamics

Figure 5·11 compares exemplar neuron spike rasters during the presentation of the object (Figure 5·11a) and window (Figure 5·11b) displays (von der Heydt et al., 2003) with the temporal response of single border-ownership units in model. The two neurons whose data are shown in (Figure 5·11a) and (Figure 5·11b) both demonstrated higher selectivity in the object condition than in the window condition. The time series were generated by neurons whose receptive fields were centered along the middle of the vertical square edge. In Figure 5·11a, the object-selective neuron elicited a much greater response when the square object appeared within and on the preferred side of the neuron’s receptive field (thick black line) compared to the nonpreferred side (thin black line). Von der Heydt et al. (2003) show that object-selective neurons demonstrated a higher separation between preferred and nonpreferred responses in the object than in the window condition (Figure 5·11b). Border-ownership units were tested in the model with the object (Figure 5·11c) and window (Figure 5·11d) conditions and units whose time series responses are depicted in Figure 5·11c-d had receptive fields centered on the same spatial locations as the neurons shown in Figure 5·11a-b. As observed in the single cell data, model border-ownership units in the object condition (Figure 5·11c) exhibited higher separations between their preferred (orange thick line) and nonpreferred (orange thin line) side activations in the object condition than in the window condition (Figure 5·11d). Moreover, model units with
Figure 5-11: (a-b) Mean time series data from single exemplar neurons that have receptive fields centered on the left vertical kinetic edge bordering the square and surrounding regions in the object (a) and window (b) conditions (data from von der Heydt et al. 2003). Both V2 border-ownership neurons prefer the object interpretation. When the square was on the preferred side (thick black line) compared to the non-preferred side (thin black line) and the square was interpreted as the object (a), the difference in firing rate was more pronounced than when the square was interpreted as the window (b). The neuron firing rates were lower in the window case than the object case. (c-d) Normalized cell activity of border-ownership units in the model with receptive fields centered on the same locations as the neurons shown in (a-b). As was the case with the neurons in the study of von der Heydt et al. (2003), units exhibited greater preferred (dark orange line) responses in the object condition than in the window condition. (e) Difference between the preferred and nonpreferred responses in (c) and (d) (separation between dark and light orange curves). The response separation is greater in the object condition (c, green) than in the window condition (d, pink).
receptive fields in the same location as the cell data showed greater preferred border-ownership activations in the object condition (Figure 5.11c) over time compared to the window condition Figure 5.11d. Figure 5.11e plots the difference between the preferred and nonpreferred border-ownership unit responses in Figure 5.11c (Object condition, green) and Figure 5.11d (Window condition, pink). The graph shows that the model unit exhibits greater separation in the preferred and nonpreferred responses in the Object condition than in the Window condition, which is also observed in the exemplar neurons (Figure 5.11a-b).

5.4 Discussion

In the present article, we presented a computational neural model of the early primate visual system that produces border-ownership signals that are consistent with human figure-ground percepts in kinetic random dot displays. Specifically, we tested kinetic random dot displays that contain texture accretion/deletion with stationary (Figure 5.1.1a) and moving (Figure 5.1.1b) figure surfaces. We tested more general versions of these displays with two or four kinetic edges wherein humans report seeing strip (Figure 5.1.1e) and square (Figure 5.1.1f) figure surfaces, respectively. Model performance on displays with shearing motion (Figure 5.1.1d) rather than texture accretion/deletion at kinetic edges was also tested. Kinetic random dot displays are challenging for models because at any time, the distribution of luminance around kinetic edges is random and figure-ground segregation cannot be performed based on luminance contrast. The model provides a mechanistic explanation for the tendency that the region that is correlated with the kinetic edge where texture accretion/deletion occurs is seen as the figure surface and owns the kinetic edge (Kaplan, 1969). Dynamically interacting neural pathways in the model that correspond to the dorsal and ventral streams group and compete across multiple cortical areas to
assign border-ownership to kinetic edges in the visual displays. The model predicts that the functional properties of known anatomical feedback connections between MT and V4 (Rockland and Knutson, 2000) at least in part serve to communicate contextual information from large receptive field cells to border-ownership cells in V2. The feedback loops between small receptive field border-ownership units in V2 and larger receptive field units that preferentially respond to stationary (V4) and moving (MT) textures are used to show how cells in primate visual cortex develop border-ownership signals solely due to dot motion patterns (von der Heydt et al., 2003).

5.4.1 The role of higher visual areas in figure-ground segregation

Border-ownership selectivity in the model emerges through the feedback from cells in V4 and MT, which have large but limited receptive field sizes. We demonstrated that when the gap separations between the figure and ground surfaces increased beyond the receptive field size of the larger model units in V4 and MT, the magnitude of the border-ownership vectorial modulation index drops to zero (Figure 5·10). This trend is consistent with psychophysical data that show that human reports of the presence of a depth ordering drop to near chance levels when the gap between the figure and ground surfaces becomes sufficiently large (Yonas et al., 1987). We predict that depth ordering is more likely to be perceived by human subjects for small gap sizes because the determination of depth ordering in the model depends on border-ownership of the kinetic edge. The slope of the mean border-ownership magnitude as a function of gap size obtained in model simulations (Figure 5·10b) does not fit the slope of the psychophysical data. While border-ownership magnitude may represent a neural correlate of the percept of depth ordering, the measure is indirect and many other decision-making and detection factors likely influence the data of Yonas. However, the qualitative decline in depth ordering report probability can be explained in the model via the recruitment of limited size receptive field units in V4 and MT.
5.4.2 Border-ownership cells and kinetic edges

In our model, we predict the existence of two distinct subpopulations of B cells in V2: those in the parvocellular stream that preferentially respond to stationary patterns, and those in the magnocellular stream that preferentially respond to texture that moves in different directions. We predict that B cells that are well-known to generate border-ownership signals to edges defined by luminance contrast (Zhou et al., 2000) compose a subpopulation of the border-ownership neurons in V2, and other subpopulations that project to more dorsal visual areas yield border-ownership signals to kinetic edges (von der Heydt et al., 2003). Conversely, cells in V2 that have been reported to preferentially respond to kinetic edges (Marcar et al., 2000; Chen et al., 2012) may possess border-ownership selectivities, akin to magnocellular B cells in our model.

In the object and window conditions, model border-ownership units exhibited similar temporal properties to neurons in primate V2 (Figure 5·11) with receptive fields centered at the same position along the square region boundary (von der Heydt et al., 2003). In particular, model border-ownership units exhibited higher sustained activation in the object condition (Figure 5·11c) than in the window condition (Figure 5·11d), and the difference between preferred and nonpreferred side-of-figure responses is larger in the object condition than in the window condition (Figure 5·11e). Despite the similarities between the model and primate neuron responses, there are notable differences in the object and window conditions used in the present study and by von der Heydt and colleagues. The square and surrounding regions used by von der Heydt and colleagues were defined not only by dot motion correlation differences (as in the present study), but also by grayscale luminance contrast. Border-ownership neurons in V2 are well-known to generate border ownership signals for edges defined by luminance contrast, so it is likely that border-ownership neurons integrate both lu-
minance and motion information. However, the data of by von der Heydt suggest that
the border-ownership responses produced by their sample of V2 neurons could only
be explained by a response to motion-induced figure-ground information. Our model
demonstrates that border-ownership neurons may elicit robust border-ownership sig-
nals based on the information present in the spatio-temporal dot motion patterns
alone. The presence of contrast information was not necessary for the model to pro-
duce border-ownership signals consistent with human percepts. Luminance contrast
differences between the figure and ground surfaces likely aids in the figure-ground
segmentation, and perhaps enhances the border-ownership signals. We know of no
study that directly tests for border-ownership in displays with surfaces defined solely
based on the spatio-temporal correlations of random dots.

Upper and lower halves of the visual field representations in V2 are well known
to project to V4 at foveal eccentricities up to 40° (Gattass et al., 1997). This sug-
gests that V4 may be important for processing of spatially precise objects in the
fovea (Piñon et al., 1998). Extensive feedforward and feedback projections between
V2 and V4 have been shown to exist (Ungerleider et al., 2007). Although V4 has
been implicated as an area that processes curved contours and shape, neurons in V4
have been shown to be sensitive to kinetic edges (Mysore et al., 2006; Orban, 2008).
Interestingly, between 10-20% of neurons sampled in V4 showed sensitivity to kinet-
ically defined shapes, and kinetic-edge-selective neurons also responded to the shape
when it was defined by luminance contrast (Mysore et al., 2006). The joint sensitivity
to luminance contrast and motion was not tested. We predict that the subpopula-
tions of V4 neurons that are driven by edges, either defined by kinetic dot motion
or luminance contrast, may exhibit an increased response when both properties are
present simultaneously and the amount of response gain may be proportional to the
area the two properties occupy within the receptive field. The finding of Orban and
colleagues that V2 cells elicit different sensitivities and tuning curves depending on whether texture movement was parallel or orthogonal to the kinetic edge is consistent with our model assumption that V4 cells feedback differently to V2 border-ownership units (Table 5.1) depending on whether there is texture accretion/deletion (orthogonal motion with respect to the kinetic edge) or shearing motion (parallel motion with respect to the kinetic edge).

5.4.3 Motion mechanisms

We showed that grouping and competition among units sensitive to different spatio-temporal correlations is sufficient for performing figure-ground segregation in kinetic random dot displays. This required discounting luminance contrast in the signal transmitted to model V1 units. Several neurophysiological studies of V1 have presented evidence that suggests that there may be separate mechanisms underlying motion and contrast sensitivity in V1 complex cells (Livingstone and Conway, 2003; Priebe et al., 2006). If the underlying mechanisms for V1 complex cell motion direction and contrast sensitivity are largely independent, an extension of our model may be able to address visual phenomena that require form-motion interactions. We assumed in the model that V1 cells receive projections from LGN units aligned orthogonally relative to the preferred motion direction (Livingstone and Conway, 2003). Neurophysiological evidence suggests that this orthogonal V1 subunit structure is not necessary to obtain motion selectivity in any particular direction. However, the LGN to V1 connectivity structure in the model assisted in obtaining greater sensitivity to the fine spatial resolution kinetic edges. The in vivo subpopulation that has facilitatory zones aligned orthogonally with respect to the preferred motion direction may be involved in detecting high resolution motion discontinuities.

Our model does not require the subtraction of adjacent region velocities (differential motion) to detect and perform border-ownership assignment of kinetic edges.
Several studies have identified motion sensitive cells in primate MT—with adjacent facilitatory and suppression zones within their receptive fields (Xiao et al., 1997; Born and Bradley, 2005), which some have interpreted as evidence that these neurons could be used to compare the velocities of adjacent visuotopic regions thereby supporting a differential motion processing strategy (Hildreth, 1992; Royden, 2002; Beck et al., 2008). A differential motion strategy, although not necessary in our model, may operate in parallel or in tandem with the competitive grouping one used in the present article. The sharp border-ownership peaks produced by the model near kinetic edges may also be involved in the crisp percept of kinetic edges. The model demonstrates that the sensitivity of cells in V2 (Marcar et al., 2000; Chen et al., 2012) and V4 (Mysore et al., 2006) to kinetic edges does not necessarily require a differencing operation on feedforward velocity signals.

5.4.4 The representation of relative depth

Some models have employed units that respond to discrete depth planes to represent the presence of multiple depths in a visual scene (Kelly and Grossberg, 2000; Berzhanskaya et al., 2007). Our model introduces a novel coding strategy whereby cortical circuits employ the ordinal magnitudes of border-ownership signals nearby a kinetic edge to represent the relative ordering of surfaces in depth. For example, in the stationary condition simulation (Figure 5.7a, bottom panel), the stationary border-ownership unit subpopulation yielded the highest magnitude vectorial modulation index near the kinetic edge, followed by the subpopulation sensitive to leftward correlations, and finally the subpopulation sensitive to rightward correlations. The border-ownership signal produced by the subpopulation sensitive to rightward correlation is much smaller in magnitude compared to the other subpopulations because that spatio-temporal correlation was not present in the visual display. The relative ordering of the border-ownership subpopulation peaks corresponds to the perceived
depth ordering in the visual display. Across our tests, when a visual display did not contain a particular spatio-temporal correlation to which a B cell subpopulation was selective, that subpopulation elicited a small border-ownership signal close to zero. Thresholding low amplitude signals, such as those yielded by the rightward-correlation-sensitive B cell subpopulation in the stationary condition, would leave an ordered magnitude gradient corresponding to correlations of surfaces present in the visual display. The relative depth ordering code by peak border-ownership signal magnitude also holds in more complex displays, such as the object and window conditions (Figure 5·9).

Neural representations that use the ordinal magnitudes of the firing rates of several subpopulations have been documented elsewhere in cortex. Averbeck and colleagues recorded from single neurons in macaque prefrontal cortex while monkeys performed a serial movement task that involved drawing simple geometric shape segments in a prescribed order. Averbeck found that the rank of the segments that would be drawn during each trial was associated with the relative activity strength of different neural subpopulations in prefrontal cortex. At the trial outset, when the monkey was trained to draw a square, for example, the neural subpopulation that would elicit peak activation during the drawing of the first segment was most active, followed by the subpopulation that elicited peak activation during the drawing of the second segment, and so forth. Similar to how the strength of the representation in different neural subpopulations in the Averbeck study was associated with the rank of the shape segments that would be drawn in the sequence, B cell subpopulation activity gradients may encode the relative depth ordering of proximal kinetically-defined surfaces. By representing depth ordering using a dynamic border-ownership magnitude code rather than by fixed depth planes, the cortex does not require additional resources to process ordinal depth. Studies have indicated there are a limited number of depth planes (e.g.
up to six) human subjects can segment at once within simple visual scenes (Tsirlin et al., 2008; Satgunam et al., 2009). If sufficiently many subpopulations produce border-ownership signals comparable in magnitude (Figure 5·7d), the signals may be too similar to decode at a population level and the border-ownership of a kinetic edge may become ambiguous. Studies that investigate the number of depths that can be simultaneously perceived in humans tend to use transparent motion displays, whereby dots move with different spatio-temporal correlations depending on which plane they occupy. Due to the lack of kinetic edges, however, it is not clear whether border-ownership circuits are recruited in the perception of depth in transparent motion displays. Perhaps different cortical mechanisms underlie the perception of depth globally (e.g. in transparent motion displays) and locally (e.g. at kinetic edges).

The relative separation between the border-ownership signal magnitudes of the competing V2 subpopulations (red v.s. black curve) in the shear condition (Figure 5·7c) is smaller relative to the separation obtained in the stationary and encroach conditions (Figure 5·7a – b). Due to random fluctuations in cortex, the model predicts that reversals in the border-ownership signal magnitude ordering produced by the competing V2 subpopulations is more likely in the shear condition than in stationary and encroach conditions. This is consistent with human psychophysical data that suggest that the percept of figures predominately defined by shearing motion is less powerful than those defined by texture accretion/deletion (Royden et al., 1988).

5.5 Conclusion

Our model uses dynamic grouping of spatio-temporal correlations and competition between border-ownership units in V2 to explain how humans perform figure-ground segregation in kinetic random dot displays. The model performs border-ownership of kinetic edges and determines the relative depth of proximal surfaces that have
different spatio-temporal correlations from the border-ownership signal magnitude gradient.
Chapter 6

Shape detection

Shape detection and processing occurs in the ventral pathway, beginning in the retina, and includes projections to the LGN, V1, V2, V4, and finally IT. LGN cells are sensitive to small spots of light, V1 cells are sensitive to oriented bars, V2 is the first extra-striate area with cells that respond to which side a figure appears relative to the receptive field, V4 neurons encode partial shape representations and are highly nonlinear in their responses. IT neurons respond to whole objects. How does processing within and between visual areas in the ventral stream afford the detection of whole shapes?

Pasupathy and Connor provide strong evidence that V4 neurons represent the curvature of shapes and contours, as an intermediate bridge between line segment representations in V1 and V2, and representations of complex shapes in IT (Pasupathy and Connor, 1999). Macaque monkeys were presented different curved contours, which systematically varied based on the local sign and magnitude of curvature, acuteness, and orientation. A considerable number of V4 neurons demonstrated selectivity to a particular orientation of the curved contour, and the response could not be explained by selectivity to constituent parts of the stimulus—e.g. by the orientation of local segments of the global shape or the angle of the contour separation. Within a typical V4 cells receptive field, the response to the preferred curved contour did not significantly change, which indicates the cells show position invariance that is important for downstream processes of object recognition. When a number of the curved
contours were arranged such to form closed, simply connected shapes and the V4 neurons receptive field was centered on the shape, V4 neurons demonstrated selectivity to certain neighboring curvatures when they appeared at particular angular positions with respect to the shape center (Pasupathy and Connor, 2001). Neurons generally were selective to convexities rather than concavities (Figure 4b). Interestingly, when the Pasupathy and Connor weighted each neurons curvature tuning peak response by the response garnered to each shape, the population response peaks matched the local curvature around the shape boundary. From the population response, Pasupathy and Connor were able to reconstruct the original shape, which suggests shape curvature is represented at a population level in V4 (Pasupathy and Connor, 2002). As many as 53% of sampled cells in V4 demonstrate border-ownership selectivity (Zhou et al., 2000), which is another indication that V4 neurons appear to hold partial shape and figure representations. Responses to partial shapes may be integrated in later visual areas to represent whole objects in a view-invariant fashion.

When humans recognize objects, such as familiar faces, they do so effortlessly despite changes in illumination, position, size, and view-point in the images projected onto the retina between multiple presentations of the same objects. Neurons in IT are thought to play an integral role in object recognition because they respond to their preferred properties despite position and size transformations within the receptive field (Tsao et al., 2006). The problem of object recognition is extremely challenging because of the tradeoff between specificity and invariance. In order to signal the recognition of a known object, a neuron must elicit a selective response to the recognized object that is different than that to other objects. However, the if the criteria employed by the neuron to elicit the response is too specific, the brain faces a combinatoric explosion because every imaginable object would require a unique neuron to represent it. On the other hand, the criteria to elicit a response needs to be general
enough to recognize the same object under a wide range of conditions. However, if the
criteria are too general, the response may no longer yield a selective response (Tsao
and Livingstone, 2008). Neurons in IT demonstrate both specificity and invariance
for complex objects, such as faces. For example, a neuron may respond to the picture
of a human hand despite its orientation, but not respond to five rectangles joined
together to approximate the shape of the hand (Desimone et al., 1984). Models of
IT neurons that perform object recognition (e.g. HMAX) are typically composed of
a series of simple and complex cell filtering processes, the responses from one stage
are combined in a weighted sum to generate the inputs to the next, and the final
stage performs a nonlinear operation such as the Max operation on its inputs to
elicit an object-selective response (Riesenhuber and Poggio, 1999; Riesenhuber and
Poggio, 2002; Serre et al., 2007). Units in intermediate layers acquire translation
and size invariance due to the model hierarchical structure, and intermediate units
may be tuned to objects at different views (Riesenhuber and Poggio, 2000). Due to
the complexity of the IT neuron responses and their intricate selectivities, models
have not had success simulating the spatio-temporal response properties of large IT
populations.

The temporal progression of neuron responses in V4 also provides important in-
sight that shape detection is a dynamical process. Individual neurons in V4 have been
shown to demonstrate differentiated responses depending on whether an visual object
appears in the foreground or background within 25 msec of its presentation (Zhou
et al., 2000) and complex objects, such as familiar faces, are recognized in only 100–
200 msec (Tsao et al., 2006). Strong cell selectivity in the early responses (\~200
msec or before) is thought to involve largely feedforward mechanisms, while later
modulation is thought to primarily involve feedback or both feedforward signals and
feedback (Angelucci et al., 2002; Schwabe et al., 2010; Craft et al., 2007; Layton et al.,
Hegdé and van Essen investigated the temporal dynamics of V4 neurons by presenting macaque monkeys with a spectrum of different visual displays sinusoidally-defined gratings, displays that have radial symmetry, spiral patterns, bars, crosses, two or more bars joined at a particular angle, and parts of circles. The visual displays were systematically manipulated across several dimensions, such as spatial frequency, number of blades in the spiral patterns, angular separations, and orientation (Hegdé and van Essen, 2005).

Hegdé and van Essen examined the temporal dynamics of V4 neuron responses for 300 msec following the presentation of each of the large set of parametrically-defined visual displays. Neurons generally exhibited a period of transient increased firing rate in which the responses of neurons were highly correlated with one another (low selectivity) followed by a period of firing rate decay in which the neurons were decorrelated from one another (high selectivity). While it was not clear what low-level properties of the visual displays contributed to the neuron responses, the neurons demonstrated a considerable amount of variability and selectivity to different types of visual patterns, which suggests the neurons perform a nonlinear integration of their inputs (Hegdé and van Essen, 2006). A multidimensional scaling (MDS) analysis revealed that the V4 population can discriminate well between stimuli that would drive V1 and V2 neurons, such as bars, and complex spirals or displays with radial symmetry.

Because V4 supplies a considerable amount of feedforward input to IT, an improved understanding of the representation of visual stimuli in V4 is crucial to construct better models of IT. Present models that, like models of IT neurons, involve a cascade of simple and complex cell processing stages and optimize the fit between the model subunits and in vivo neuron receptive field shapes at each stage perform no better in V4 than in other extrastriate areas, with at best 32% of the response
variance explained (Cadieu et al., 2007). A fundamental problem that limits present models of V4 is that the mathematical properties and geometric configuration of V4 neuron receptive fields is unknown. The HMAX model cannot detect convexities and concavities in objects that are presented because it lacks side-of-figure and surface representations (i.e. border-ownership). Sufficiently many hierarchical stages of simple and complex cell convolutions on the input image may produce a fit to a number of the qualitative receptive field properties observed in V4 neurons, however, due to the number of network layers, mathematical analysis of the receptive field structures is difficult to obtain. Although present models fall short in describing the characteristic properties of V4 neurons, neurophysiological studies provide insight into the receptive field organizations of V4 cells and their selectivities.

We present a dynamical neural model of primate V4 with convex receptive field units that elicit activity to the interior of figures at the appropriate size, but not to their exterior. This is, we propose a solution to the in versus out problem in shape processing: how are neurons sensitive when an interior portion of a shape appears inside the receptive field, and not the outside? We hypothesize that the spectrum of neurons with different receptive field sizes, ranging from small in V1 and V2 to large in V4, that project to IT such that multi-resolution represents can be multiplexed for shape detection purposes. Model V4 cells perform a nonlinear multi-scale integration of subunits that vary in retinotopic location and receptive field size, and feedback to units that are selective to figures of a single size. The known receptive field size and location scatter at a given visual eccentricity is exploited in the model to sense the extent and continuity of shape boundary contours. The model is tested on large sets of parametrically-varying visual displays, and model performance is assessed as a function of how V4 subunits are combined in the multi-scale shape integration.
Chapter 7

Dynamics of multi-scale shape integration and detection in the primate ventral stream

7.1 Introduction

Shape processing is thought to represent a primary function subserved by the ventral stream in the primate visual system. V4 is hypothesized to be essential for detecting shapes (Roe et al., 2012). Lesions to V4 induce behavioral deficits in discriminating images of objects (Merigan and Pham, 1998) and selecting objects from an array (De Weerd et al., 1999; De Weerd et al., 2003). Pasupathy & Connor provide strong evidence that V4 neurons represent the curvature of shapes and contours (Pasupathy and Connor, 1999). V4 neurons demonstrate selectivity to particular orientations of curved contours, which may possess different local signs and magnitudes of curvature and acutenesses (Pasupathy and Connor, 2001). V4 neurons demonstrate selectivity when contours are arranged to form closed, simply connected shapes that have specific signs and magnitudes of curvature at particular angular positions with respect to the shape center. From the population response, Pasupathy & Connor reconstructed the original shapes, which suggests shape curvature is represented at a population level in V4 (Pasupathy and Connor, 2002).

V4 projects to IT cortex, where partial shape representations may be collated into those of whole objects (Orban, 2008). Neurons in the anterior regions of IT
AIT exhibit high degrees of selectivity and invariance in their responses to complex shapes (Kobatake and Tanaka, 1994), such as faces (Tsao and Livingstone, 2008), while posterior regions (PIT), like V4, may hold more intermediate representations of shape. PIT neurons respond to 2D silhouette patterns and appear to integrate multiple curved contour segments (Brincat and Connor, 2004; Brincat and Connor, 2006). Neurons in IT demonstrate the *size-invariance property* whereby the spike rate of neurons is not significantly modulated by an object’s projected size on the retina (Logothetis et al., 1995; Ito et al., 1995). Together, electrophysiological data suggest that the primate ventral stream constructs progressively more complex representations of shape, with the development of curvature and other early prototypical partial shape representations in V4 or PIT.

The progression along the ventral stream from curved contour representations to those of complex objects raises important mechanistic questions we sought to address in a computational model of the primate ventral stream.

First, if V4 represents curvature as a population code, how are representations of curved contours coherently associated to yield highly selective responses to objects in IT cortex? IT neurons demonstrate selectivity to holistic shapes, such as faces, but the response may diminish when lower-order caricatures are presented. For instance, an IT neuron may respond to a monkey face, but not to certain line drawing representations composed of simpler shapes (Kobatake and Tanaka, 1994). Nonlinear feedforward integration appears to constitute a general-purpose strategy used by the visual system to construct complex representations from those that are simpler (Hubel and Wiesel, 1962; Pasupathy et al., 2006). For example, neurons in the dorsal medial superior temporal area (MSTd) demonstrate selectivity to large radially expanding and contracting motion patterns, which is thought to occur through the structured integration of medial temporal (MT) cells tuned to different motion directions that
match a *template* (Layton et al., 2012a; Mineault et al., 2012). Our model proposes biological mechanisms by which IT neurons may discriminate a holistic object from an disorganized array of luminance contrast.

Second, how do shape selective neurons, with limited RF sizes, determine whether the interior or exterior of a shape appears within the RF (in versus out)? Physiological data indicate that V4 neurons preferentially respond to the interior (figure) rather than the exterior (ground) of a shape (Roe et al., 2012). Bushnell and colleagues found that occlusion, which did not impact local boundary contour curvature, strongly suppressed V4 neuron responses compared to when the shape was unoccluded (Bushnell et al., 2011). The side-of-figure signals of border-ownership cells in V4 to a particular shape reverses in direction when the shape is occluded by another (Zhou et al., 2000). Shape selective V4 neurons therefore appear to exhibit modulation based on figure-ground segregation, a process that requires consideration of large parts of the visual scene. In the present article, we introduce a neural coding strategy that explains how IT neurons can detect the interior of a shape when it appears within the neuron’s RF.

Shape convexity is intimately related to the process of distinguishing between in versus out—both require global information about the visual scene. Shapes similar to those to which neurons are selective may appear within the RF, yet the perceived shape in the visual scene may be different. For example, a response elicited by a square selective cell to the concavity of a ‘C-shape’ would represent a false-positive. This ambiguity exists whenever a shape is viewed through an aperture because the curvature of a local contour does not determine whether that contour defines a convex or concave portion of a shape (Koenderink, 1984). V4 neurons must therefore receive information from outside the classical RF. Our model demonstrates that feedback to shape-selective V4 neurons enhances shape tuning, helps resolve ambiguities related to
determining convexity from local curvature, and suppresses responses to the exterior of shapes.

Third, how do IT neurons multiplex the spectrum of shape representations at different spatial scales distributed across the visual system? All parts of V4 are connected via feedforward connections to PIT and AIT (Baizer et al., 1991; Ungerleider et al., 2007). Neurons with small RFs in V1 and V2 also directly project to both V4 and IT cortex (Nakamura et al., 1993). IT neurons receive feedforward projections from V4 neurons that possess four to seven times larger RFs than V1, but smaller RFs than those in IT. Multi-unit and local field potentials in V4 support the existence of multiple spatial scale representations, which may contribute to size-invariant objects responses in IT (Ghose and Ts’o, 1997; Mineault et al., 2013). IT likely receives input from neurons with a variety of RF sizes, ranging from moderate size from V4 to small sizes from V1 and V2, and therefore represents a candidate area wherein multiple resolution representations of shape converge. We propose that the scatter in RF size that exists in V4 (Gattass et al., 1988; Piñon et al., 1998) and IT (Ito et al., 1995) serves a crucial role for shape processing and figure-ground segregation. In our model, shape representations at different resolutions are coded by units with different RF sizes, which are multiplexed to detect the presence and size of shapes.

Our objective was to create a biologically-plausible model to demonstrate how the primate ventral stream detects the presence and size of shapes using dynamically interacting populations of neurons that process curved contours. The model addresses the three mechanistic points raised above: how sensitivity to complex shapes may be obtained through the combination of curved contour representations, how the inside versus outside of shapes is determined, and how multi-resolution shape representations may be multiplexed. We test the following hypotheses: 1) Scatter in receptive field (RF) location and size can be used to detect shapes in neural populations
connected by fast inter-areal connections. 2) Feedback enhances shape sensitivity by providing contextual information about the global scene layout from outside the classical RF, suppressing neural activity to the exterior of shapes, and enhancing neural activity to the interior. Our simulation results demonstrate that model single-scale units detect the spatial scale of shapes in the visual scene through an activity peak at the appropriate scale, and the RF location of the peak signals the shape’s location. When shapes appeared within a simulated IT cell’s receptive field, feedback both increased the response gain and suppressed activity to the exterior of shapes.

7.2 Methods

Model of shape and figure-ground segregation

In the present article, we describe a set of quantitative experiments using a computational model of the ventral stream that consists of three main stages, which correspond to respective neurons in V4, PIT, and AIT: curved contour units, single-scale units, and multi-scale units (Figure 15-1). Computational properties of model units in each stage are supported by physiological evidence from corresponding cortical areas. The first dynamical stage of the model contains curved contour units, which exhibit sensitivity to contrast segments with different curvatures. Single-scale units group curved contours at a common spatial scale thereby achieving sensitivity to partial shapes. Multi-scale units integrate the partial shape representations garnered by the single-scale units at multiple spatial scales. Model details follow below.

The model has five overarching properties. First, the model consists of systems of shunting first-order differential equations, each of which describes the activity of neurons in cortex (Grossberg, 1973). This architecture affords the simulation of realistic temporal dynamics and competitive interactions of neuronal populations, including recurrent competition and feedback (Layton and Browning, 2012), gain control, and
model contains two types of feedback: multi-scale radial feedback, whereby units that perform multi-scale integration in the final model layer send feedback signals to single-scale units from outside of the classical RF, and single-scale feedback, whereby single-scale units with larger RFs send feedback projections to those with smaller RFs to resolve ambiguities related to figure-ground assignment and the convexity of boundary contours.

Model overview

An overview of the model architecture is schematically depicted in Figure 15-1. Feedforward pathways in the model begin with the input image, followed by complex cell processing, and three dynamical stages containing layers of curved contour units, single-scale units, and multi-scale units. The model first computes complex cell ‘edge map’ representations of input shapes.
Curved contour units

The first dynamical stage of the model, curved contour units, correspond to neurons in V4 (Pasupathy and Connor, 2002). Curved contour units are sensitive to the curvature of contrast segments at different spatial scales (Figure 15.2a). We simulate units that have selectivities to one of eight curvatures and have one of seven spatial scales. Curved contour unit templates resemble quarter circle arcs with eight different phases. The units emit the most activity when a curved contour enters the RF
Figure 7·1: Diagram of our model of shape detection and figure-ground segregation. The model performs complex cell filtering of the image to create an ‘edge map’, followed by three dynamical stages containing curved contour units, single-scale units, and multi-scale units corresponding to primate visual areas areas V4, PIT, and AIT, respectively. There are both feedforward and feedback connections. The response in each successive stage is driven by nonlinear feedforward integration of units in the previous stage. Curved contour units (model V4) detect curved contours in the V1 edge map. Single-scale units (model PIT) perform a nonlinear integration of curved contour units in an annular spatial configuration. The coactivation of multiple curved contour units results in nonlinear increase in the postsynaptic single-scale unit. Multi-scale units (model AIT) integrate single-scale units with scattered RF sizes and locations to multiplex shape part representations at difference scales extended in particular directions across visuotopic space (bottom-right of top panel). Single-scale units are integrated in eight integration directions (four shown in top-right of top panel) that extend radially from the multi-scale RF center (blue circle). Farther away from the multi-scale RF center, the RF size and feedforward weight of the single-scale units decrease. The model contains two types of feedback. Multi-scale units send feedback projections (Multi-scale radial feedback) from all eight radial integration directions to enhance or suppress single-scale unit activity from outside the classical RF. The multi-scale radial feedback signal pools over multi-scale units with different integration directions and thereby signals evidence for the presence of a shape along the radial directions. Single-scale units with large RFs send feedback projections to those that possess small RFs and overlapping curved contour subunits (single-scale feedback). The single-scale feedback mechanism reduces activation due to concave regions that reside outside of figures (Figure 15·6).

that matches the unit’s preferred scale and curvature. Similar to cells in V4 (Yau et al., 2012), strong firing occurs when junctions of straight contours intersect within the RF, and the units are weakly active if contours enter the RF at oblique angles (Figure 15·2b).
Figure 7·2: Curved contour units are sensitive to a particular curvature and orientation of contrast. (a) Example curved contour units. The spatial scale is determined by the RF size. (b) Curved contour units emit the most activity when contrast appears within the RF that match the preferred curvature, and they emit strong activation when junctions of lines provide a close approximation (green). Activation is weak when there is no contrast in the RF or when contrast intersects the RF at oblique angles (gray).

Single-scale units

Single-scale units (model PIT) perform a nonlinear feedforward integration over multiple curved contour units (Brincat and Connor, 2004). Each afferent connection comes from a curved contour unit that possesses a different preferred orientation, but the same RF size (single scale) (Figure 15·3a). Excitatory subunits are configured in a convex, annular, on-surround arrangement. This introduces a preferential response bias in single-scale units toward convexity and closure. Single-scale units have different RF sizes (Figure 15·3a), which are determined by the spatial scale of the afferent curved contour units. Model single-scale units with different RF sizes compete in a recurrent network (cross-scale competition) to sharpen the sensitivity to shapes that fall within the RF (Brincat and Connor, 2006). The nonlinear integration of curved contour units is followed by a logarithmic compressive nonlinearity that saturates responses after a fraction of subunits are concurrently active (Zhang and von der Heydt, 2010). Figure 15·3b illustrates that having a quarter, half, and all of the single-scale subunits coactive accompanies a nonlinear increase in the unit’s activation that rapidly approaches saturation. Note that we depict four of eight subunits
Figure 7-3: Single-scale units form partial shape representations in the model. (a) Single-scale units integrate curved contour units in annular configurations at a common scale. The annular receptive field biases responses toward convexity. Single-scale units possess different RF sizes. (b) Single-scale units perform a nonlinear integration of curved contour units, followed by a compressive nonlinearity. As one quarter, one half, and all subunits become active, the single-scale unit activity will quickly saturate (right panel). Note that simulations included eight orientations of curved contour units, and four are shown for illustrative purposes.

In sum, single-scale units have annular RFs and form partial shape representations in the model at a particular spatial scale.

Multi-scale units

We claim that the scattering of RF sizes (Gattass et al., 1988; Ito et al., 1995) in ventral visual areas is crucial for shape processing and figure-ground segregation.
Shapes and their relevant sizes may be detected by neurons in the visual system that integrate shape-related neural signals at a variety of spatial scales in multiple early visual areas. We propose a neural coding strategy, implemented by multi-scale units, whereby partial shape representations at multiple resolutions are integrated over an extended area to detect the presence and size of nearby shapes.

Consider a single-scale unit that has part of a shape within its RF (e.g. a teardrop), as shown in the second panel from the left of Figure 15.4a. The blue circle indicates
Figure 7-4: Multi-scale units multiplex partial shape representations at multiple spatial resolutions across space. (a) Consider a single-scale unit (orange annulus, second panel from left) whose RF is centered on the round portion of a teardrop shape (blue annulus). Due to the unit’s large scale, it cannot integrate the fine resolution of the teardrop apex outside its classical RF, which provides important shape information. However, a collection of single-scale units with decreasing RF sizes scattered along the rightward direction relative to the large single-scale RF center can integrate the teardrop boundary contours at multiple spatial resolutions. Multi-scale units multiplex these representations distributed across space to become sensitive to the teardrop shape. (b) Multi-scale units nonlinearly combine single-scale units along radial integration directions, including the rightward direction shown in (a). The RF size and feedforward input weight (indicated by opacity) of single-scale units systematically decreases with distance from the multi-scale RF center. The multi-scale unit shown has a rightward integration direction. (c) Sample multi-scale units with four integration directions. Diagonal directions are also simulated in the model. (d) The coactivation of multiple subunits results in non-linear increases in the multi-scale unit’s activation, which is ultimately passed through a compressive nonlinearity. The feedforward integration combines the input from subunit singletons, pairs, 3-tuples, etc. Multi-scale units are most active when there is continuity in the boundary contours along the integration direction.

the RF center. Information about the shape’s form is given by the boundary contours outside the single-scale unit’s classical RF. A single-scale unit that has a nearby visuotopic RF center and the same RF size may not provide additional information about the global shape. In the teardrop example shown in Figure 15-4a, another single-scale unit with the same RF size could not capture the convergence of the boundary contours at the teardrop apex outside the first unit’s RF. However, a unit that integrates the single-scale unit and others with visuotopically proximal RF centers, but decreasing RF sizes, could adequately signal shape properties in the leftward direction, relative to the first single-scale unit’s RF. In the model, we name units that systematically integrate single-scale units with deceasing RF sizes in different spatial
directions multi-scale units.

As shown in Figure 15-4b, the multi-scale RF center coincides with the center of the largest single-scale unit from which it receives feedforward input. Along the integration direction, the feedforward weights of afferent single-scale units decrease with distance from the RF center, consistent with physiological data (Pollen et al., 2002). In our simulations, we have multi-scale units with eight integration directions (horizontal, vertical, diagonal), four of which are shown in Figure 15-4c. Along a multi-scale unit’s integration direction, the unit receives feedforward input from single-scale units according to a summation process whereby individual active subunits, pairs, triplets, and so forth nonlinearly contribute to the activation level. The greater the boundary contour continuity across scale in the integration direction, the greater the multi-scale unit activation. Because all subunit permutations are considered, multi-scale units still elicit strong activity when small parts of the boundary contours are missing. As with the single-scale unit feedforward process, a compressive nonlinearity is applied to the input signal to multi-scale units.

In summary, multi-scale units, which have large RF sizes, adopt a coding strategy whereby single-scale units with scatter in the RF size and location are integrated to obtain multi-resolution information about the shape of boundary contours over extended regions of space.

Through their dynamical interactions with multi-scale units, single-scale units activity peaks indicate the presence of a shape: the spatial scale of the unit codes the shape’s size and the visuotopic position of the RF codes the shape center. A peak produced by single-scale units on the interior of a shape indicates correct figure-ground assignment in the model, while a peak produced to the exterior of a shape indicates incorrect figure-ground assignment. The multi-scale radial feedback and single-scale feedback in the model describe mechanisms by which information about
global scene context is propagated to single-scale units with smaller RFs to facilitate figure-ground segregation.

Multi-scale radial feedback

Our model contains two types of inter-areal feedback connections, the first of which is named multi-scale radial feedback and propagates multi-resolution shape information to single-scale units from outside the classical RF via multi-scale units. Figure 15.5a illustrates that multi-scale units have a number of RF sizes from large...
Figure 7·5: The multi-scale radial feedback mechanism suppresses activity on the shape exterior. (a) Multi-scale units have different RF sizes—larger ones are shown to the left, and smaller ones are shown to the right. The multi-scale unit RF center (blue annulus) is aligned with the largest subunit. Feedback is sent from the multi-scale unit to the afferent single-scale unit that has the largest RF. (b) Consider a single-scale unit whose RF center coincides with the center of a square. This single-scale unit receives multi-scale feedback from multi-scale units shown to the left of the equals sign, which integrate along radial directions relative to the single-scale unit RF center. The feedback comes from all radial directions, resembles a radial pattern, and provides contextual information about the boundary contours (e.g. corners) outside the classical RF. (c) Multi-scale feedback sharpens and enhances single-scale unit activity on the interior of a shape. The left panel indicates that there is a strong multi-scale feedback signal at the center of the square (starred location) because the boundary contours intersect the multi-scale unit RFs along many integration directions. Strong radial signals result in excitatory modulation (red part of sigmoid) of single-scale unit activity at the starred location (middle panel), while weak radial signals result in suppressive modulation (blue part of sigmoid). (d) There is weak multi-scale unit activity on the shape exterior because little or no contrast is present within the RF. This results in suppressive modulation (blue part of sigmoid) of single-scale units that have receptive fields on exterior of the shape (starred locations).

(Left) to small (right). Feedback from multi-scale units to single-scale units is sent to the largest afferent single-scale unit, which coincides with the center of the multi-scale unit’s RF. The reason for this connectivity is depicted in Figure 15·5b. Consider multi-scale units that have RFs centered on a square shape (blue circle). A single-scale unit that has a RF centered on the square (largest multi-scale subunit in Figure 15·5b) cannot integrate information about the corners of the square, which are important junctions of the shape. The corners appear at a fine resolution relative to the large single-scale RF shown in Figure 15·5b and are outside the classical RF. As illustrated in each panel of Figure 15·5b, multi-scale units exploit RF scatter to obtain sensitivity to both the coarse resolution of the global square shape and the high resolution
of the local corners. Multi-scale units multiplex multi-resolution information about the square boundary contours in radial directions relative to the largest multi-scale subunit. Feedback sent from multi-scale units, which share the same RF center but possess different integration directions, to the largest afferent single-scale unit is what we call multi-scale radial feedback. This type of feedback propagates multi-resolution information about the shape boundary to single-scale units from radial directions outside the classical RF.

The feedback signal sent from multi-scale units at each visuotopic location is passed through a sigmoidal function that distributes the activity between -1 and 1. Positive and negative values correspond to facilitatory and suppressive effects on the postsynaptic single-scale unit, respectively. Consider multi-scale units that share a common RF size and are highly active in all integration directions. This is the case for the units with RFs centered on the square (horizontal position indicated by the star) shown in the left panel of Figure 15·5c. Highly active multi-scale units yield a facilitatory signal that targets the postsynaptic single-scale unit, according to the sigmoidal transfer function. Feedback to the single-scale unit centered on the square (starred location) is thereby enhanced and the overall spatial distribution of single-scale unit activity at different positions across the square is sharpened due to the combination of facilitation and suppression. Multi-scale units that have RFs centered outside of shapes are weakly active because there are few or no boundary conditions along most of the integration directions that would drive the units (Figure 15·5d). Due to the sigmoid transfer function, weak multi-scale unit activity yields a strongly suppressive multi-scale radial feedback signal to single-scale units with RFs outside the shape. The activity of single-scale units with RFs outside the shape therefore are suppressed.

In summary, multi-scale radial feedback projects to single-scale units that share
the same RF center, where it has two effects. If multi-scale units are active due to a boundary contour configuration that suggests a single-scale unit is activated within in a shape (part of the figure), the feedback signal is facilitatory. Multi-scale radial feedback also suppresses the postsynaptic single-scale unit when multi-scale unit activity is weak, which occurs when shapes do not appear within the RF.

Single-scale feedback

The second type of feedback in the model is named *single-scale feedback*, which facilitates single-scale unit responses to convex regions of shapes (interior) and suppresses responses to concave regions (exterior). Due to the inherent locality of their RFs, single-scale units suffer from the aperture problem. Responses may be com-
Figure 7.6: The single-scale feedback mechanism suppresses activity in exterior concavities. Consider a single-scale unit that has a large enough RF such that a C-shape falls within the receptive field (largest orange). This single-scale unit will be strong active because it receives bottom-up contrast signals from 3 of the boundary contours. However, single-scale units with smaller RFs centered along the exterior concavity (second largest orange annulus) also integrate the contrast from 3 boundary contours and are therefore equally active. Single-scale units that have RF centers on the interior of the C-shape (smallest orange annuli) are even less active due to the integration of two boundary contour signals. Figure-ground signals are therefore ambiguous. Single-scale feedback resolves figure-ground ambiguity in the presence of exterior concavities. Feedback is sent from single-scale units to those with smaller RF sizes that are approximately tangent and contained within the larger RF. When the small RF single-scale unit spatially coincides with an active afferent curved contour unit (green arcs), feedback from the large RF unit to the small RF unit facilitates the activity of the small RF unit. When the small RF single-scale unit spatially coincides with an inactive afferent curved contour unit (blue arc), feedback from the large RF unit to the small RF unit suppresses the activity of the small RF unit. In displays with exterior concavities, responses inside the shape are facilitated and those outside the shape are suppressed. A neural circuit that implements single-scale feedback is depicted in Figure 7.14c.

parable to convex and concave parts of a shape. This is a problem because units in the model may respond to shapes when in fact none would be perceived. In the C-shape example depicted in Figure 15.6, single-scale units with RFs located on the C-shape interior (Figure 15.6, “Before Feedback”, top-left panel) would be less active than those located on the C-shape exterior (Figure 15.6, “Before Feedback”, top-right panel). This occurs because the single-scale unit that has its RF shown centered on the C-shape interior integrates two boundary contours, whereas the unit that has its RF centered on the concavity integrates three. In this case, the model would incorrectly signal the presence of a shape in the C-shape concavity. Figure-ground assignment would be erroneous.
Single-scale feedback provides a mechanism by which this problem is resolved. Ambiguity about convexity stems from the limited extent of the RF relative to that of the shape. Consider a single-scale unit with a large enough receptive field to integrate the C-shape shown in Figure 15·6 (largest orange annulus). This unit is active due to the excitatory input it receives from curved contour units along the boundary contour of the C-shape (green arcs). However, the unit does not receive input from curved contour units in the concave region, due to lack of contrast signals (blue arc). The large single-scale unit sends feedback to other single-scale units with smaller receptive sizes that also receive input from the same curved contour units to propagate contextual information about the global properties of the visual scene. The single-scale units that are targeted by feedback have RFs that are tangent to and fall within the large single-scale unit’s RF. If the curved contour unit that projects to both the large and small RF single-scale units is active, as is the case in convex regions of shapes, the feedback signal is facilitatory. Conversely, if the curved contour unit that projects to both the large and small RF single-scale units is inactive, as is the case in concave regions, the feedback signal is suppressive. In effect, single-scale units that have RFs that are smaller than the extent the shape and are centered on the interior, receive facilitatory modulation from larger RF single-scale units (Figure 15·6, “After Feedback”, top-left panel). Single-scale units that have RFs that are smaller than the extent the shape and are centered on the exterior, receive suppressive modulation from larger RF single-scale units (Figure 15·6, “After Feedback”, top-right panel). In sum, the single-scale feedback mechanism diminishes model responses in concavities of the C-shape and other shapes and activity is enhanced on the interior. A neural circuit that implements single-scale feedback is depicted in Figure 7·14c.
Model Equations

Feedforward nonlinear integration

Single-scale units perform a nonlinear feedforward grouping of curved contour units into convex, annulus-shaped templates (Figure 15·3a). Each single-scale unit receives input from eight \( n = 8 \) curved contour units. The grouping obeys the following sum, which resembles an analysis of variance decomposition:

\[
I^s = \rho(s) \log \left( \sum_{i=1}^{n} C^s_{\theta_i} + \sum_{i=1}^{n} \sum_{j>i}^{n} C^s_{\theta_i} C^s_{\theta_j} + \sum_{i=1}^{n} \sum_{j>i}^{n} \sum_{k>j}^{n} C^s_{\theta_i} C^s_{\theta_j} C^s_{\theta_k} + \ldots + 1 \right) . \quad (7.1)
\]

Eq. 7.1 computes the feedforward input \( I^s \) each single-scale unit \( S^s \) receives from curved contour units \( C^s_{\theta_i} \), where \( s \) indexes the RF size (Figure 15·3b). The sum is computed over all \( k \)-tuples of curved contour units, where \( k \) ranges from 1 to \( n \) corresponding to the eight curved contour unit orientations \( \theta \). Eq. 7.1 yields multiplicative increases in afferent activity proportional to the number of contours are co-active within the template, which results in large inputs to single-scale units when the input display contains a closed shape. The result is passed through a logarithmic compressive nonlinearity, and one is added to prevent negative values. The function \( \rho(s) = 1 + 5s \) biases the sum toward contributions from larger RF units. The bold symbols signify a 2D regular grid of units each position of the visual field.

The second main feedforward mechanism (Figure 15·4) defines the grouping of single-scale units by multi-scale units. Single-scale units are grouped across visuotopic position and scale (Figure 15·4a). As depicted in Figure 15·4c, multi-scale units possess one of eight ‘integration directions’ \( d \), corresponding to the vertical, horizontal, and diagonal directions. A multi-scale unit’s integration direction specifies the pattern of input it receives from spatially-offset single-scale units. For instance, a multi-scale unit with a rightward integration direction \( (d = 1) \), scale \( s = 3 \), and RF position of \((x, y)\) in the visual field receives input from single-scale units whose RFs are centered
on \((x, y), (x + 1, y),\) and \((x + 2, y)\) (Figure 15·4a). In addition, the RF size of single-scale units over which a multi-scale unit integrates decreases with distance along the integration direction (Figure 15·4b). In the rightward integration direction example, single-scale unit at \((x, y)\) has scale \(s = 3\), the one at \((x + 1, y)\) has scale \(s = 2\), and \((x + 2, y)\) has scale \(s = 1\). Hence, the multi-scale integration occurs radially in the eight cardinal directions about the RF center, and groups single-scale units that vary scale and RF location. The input \(J^*_d\) to multi-scale units with scale \(s\) and integration direction \(d\) obeys the following nonlinear integration:

\[
J^*_d = \nu(s) \log \left( \frac{1}{B_1} \sum_u (K^u_d * S^u) + \frac{1}{B_2} \sum_u \sum_v (K^u_d * S^u) (K^v_d * S^v) + \frac{1}{B_3} \sum_u \sum_v \sum_w (K^u_d * S^u) (K^v_d * S^v) (K^w_d * S^w) + \ldots + 1 \right)
\]

In Eq. 7.2, each product corresponds to the \(k\)-tuple multiplicative combination of possible single-scale inputs, where \(k\) ranges from 1 to the largest scale \(s^*\) (Figure 15·4d). The \(*\) operator signifies convolution. When \(s^* = 2\), Eq. 7.2 only has the first multiplicative term \((k = 2)\), and \(u\) and \(v\) take on values such to sum 3 input pairs along each respective integration direction \(d\) (Figure 15·4d, right panel). In general, the \(n^{th}\) term in Eq. 7.2 will have a total of \(B_n\) nonlinear combinations, where \(B_n\) is the \(n^{th}\) Bell Number, which specifies the number of ways to partition a set of size \(n\) into \(k\) nonempty subsets. In other words, it is the total number of combinations \(nC_k\) for all \(k \leq n\). The notation \(nC_k\) means \(n\) “choose” \(k\). We normalize the \(n^{th}\) term by \(B_n\), the number of combinations. \(S^i\) corresponds to the activity of single-scale units at scale \(i\) and \(K^i_d\) is a binary indicator kernel that picks out the \(j^{th}\) single-scale input along the integration direction \(d\), \(j \leq i\). When the integration direction is rightward and \(s = 2\), \(u\) and \(v\) take on values to select the single-scale unit at \((x, y)\) with \(s = 2\).
at \((x+1, y)\) with \(s = 1\), or both. In Eq. 7.2, \(\nu(s) = \log(s)\) such to give greater weight to input from units with larger RFs.

**Multi-scale radial feedback**

Figure 15-5 schematically depicts how multi-scale units feedback to single-scale units. Multi-scale units are visuotopically aligned in the model such that the multi-scale unit \(M_d^s\) at position \((x, y)\) with scale \(s\) receives its largest feedforward input from the single-scale unit with scale \(s\) at position \((x, y)\) irrespective of the integration direction \(d\) (Figure 15-5a). Multi-scale units at the same visuotopic position perform a radial integration across space and scale in the eight integration directions (Figure 15-5b, right panel).

The feedback signal to single-scale units is determined by summing over all eight integration directions (Figure 15-5b),

\[
R^s = f \left( \sum_k M_{k}^s; \kappa_M, \Gamma_M \right).
\]  

(7.3)

In Eq. 7.3 the function \(f(\cdot)\) is the sigmoidal function,

\[
f(w; \kappa, \Gamma) = \tanh (\kappa w - \Gamma).
\]  

(7.4)

The parameters \(\kappa_M\) and \(\Gamma_M\) adjust the sigmoid steepness and horizontally shift the linear portion of the sigmoid, respectively. The feedback signal from multi-scale units passes through the sigmoid function to either facilitate or suppress the activity of targeted single-scale units (Figure 15-5). When multi-scale units do not receive strong contrast signals in the radial integrations, the radial sum in Eq. 7.3 will be small. Modulation by feedback transfer function (Eq. 7.4 ) yields strong suppression (blue portion of sigmoid in Figure 15-5c, center panel). Conversely, when strong contrast signals are integrated in the radial directions, the sum will be large (red
portion of sigmoid in Figure 15·5d, center panel). The transfer function will preserve the sign and magnitude of the signal, and have a facilitatory effect on the postsynaptic single-scale unit (Figure 15·5d, right panel).

**Single-scale feedback**

Single-scale units receive feedback signals from others with larger RFs that receive input from spatially-overlapping curved contour units (Figure 15·6). The following equation describes the feedback signal $T^s$ a single-scale unit with scale $s$ receives from larger RF units. Note that the receptive field centers are at different visuotopic locations.

$$T^s = \frac{1}{n \cdot s} \sum_{r=1}^{n} \sum_{k>s} f(C^s_{\theta_r} ; \kappa_S, \Gamma_S) \cdot (K^k_r \ast S^k)$$

(7.5)

In Eq. 7.5, the sigmoid $f$ (Eq. 7.4) acts on each curved contour within the annular RF. If the $r^{th}$ curved contour unit is inactive, $f(C^s_{\theta_r} ; \kappa_S, \Gamma_S)$ will contribute a suppressive input to the overall feedback signal $T^s$. Conversely, when the $r^{th}$ curved contour subunit is active, $f(C^s_{\theta_r} ; \kappa_S, \Gamma_S)$ will contribute a facilitatory input to the signal. In Eq. 7.5, convolution with the binary indicator kernel $K^k_r$, as in Eq. 7.2, picks out larger RF single-scale units $k, k > s$ that have tangent RFs (see Figure 15·6).

**Model unit dynamics**

The radii of the annular templates range from 2 pixel to some upper level, generally 10 pixels, in 1 pixel increments. Annular templates are normalized to sum to unity. To construct curved contour templates, we first used the following equation that takes $(s, \theta) \mapsto (x_a, y_a, z_a)$ values to form an annulus of radius $s$:

$$(x_a, y_a, z_a) = (s \cdot \sin(\theta), s \cdot \cos(\theta), \frac{1}{2} \left(1 - \cos \left(\frac{2\pi}{s}\right)\right)).$$

(7.6)
In Eq. 7.7, $\theta \in [0, 2\pi)$, and $(x_a, y_a, z_a)$ are the spatial coordinates of the annular filter. To produce the eight curved contour templates $Q_{\theta}^s$, we extracted curved contours from eight equally sized radial sectors of the annular kernel:

$$Q_{\theta}^s = \{(x_a, y_a, z_a)|\theta_i \in \left\{-\frac{3\pi}{4}, \ldots, \frac{3\pi}{4}\right\}, \theta_{i-1} \leq \tan^{-1}\left(\frac{y_a}{x_a}\right) \leq \theta_i\}. \quad (7.7)$$

The following leaky integrator model describes the dynamics of curved contour units.

$$\frac{d}{dt}C_{\theta}^s = -\alpha C_{\theta}^s + (\beta - C_{\theta}^s)Q_{\theta}^s * A \quad (7.8)$$

In Eq. 7.8, $A$ signifies bottom-up input from a complex cell (edge) map representation of the visual display, $Q_{\theta}^s$ is the curved contour template, $\alpha$ is the passive decay rate, $\beta$ is the saturation upper bound. Parameter values remained fixed throughout all simulations and are listed in Table 15.1.

Curved contour units supply feedforward input to single-scale units in an annular template. Single-scale dynamics obey the following shunting equation with recurrent competition across scale:

$$\frac{d}{dt}S^s = -\alpha S^s + (\beta - S^s)((S^s)^2 + I^s) - (S^s + \gamma_S)\left(\sum_{s=1\ldots s \neq k \neq s} (S^k)^2 + R^s + T^s\right) \quad (7.9)$$

In Eq. 7.9, $I^s$ specifies the feedforward integration over curved contour units (Eq. 7.1), $\gamma_S$ is the hyperpolarizing lower bound of the model cell, $R^s$ is the single-scale feedback signal (Eq. 7.3), and $T^s$ is the multi-scale feedback signal (Eq. 7.5). The variables $S^s$ and $S^k$ are squared to produce a winner-take-all contrast-enhancing network with respect to spatial scale (Grossberg, 1973).

Multi-scale units perform a nonlinear feedforward integration of single-scale units across visuo-optic space and scale in the eight cardinal directions. Multi-scale dynamics
obey the following shunting equation with recurrent competition across integration direction:

\[
\frac{d}{dt} M^s_d = -\alpha M^s_d + (\beta - M^s_d)(g(S^s) + J^s_d) - M^s_d \sum_{k \neq d} g(\lambda M^s_d), \tag{7.10}
\]

where \( g(\cdot) \) is the sigmoidal function:

\[
g(w; \zeta_g, \Gamma_g) = \frac{(1 + (w - \Gamma_g)^{\zeta_g})}{\zeta_g + (1 + (w - \Gamma_g)^{\zeta_g})\zeta_g}. \tag{7.11}
\]

In Eq. 7.10, \( J^s_d \) represents the feedforward input from single-scale units (Eq. 7.2). The sigmoidal function \( g \) (Eq. 7.11), induces winner-take-all, pattern-preserving, and uniformizing behavior when the activation of units in the model falls in the faster-than-linear, linear, and slower-than-linear regions of the signal function, respectively. The parameters \( \Gamma_g \) and \( \zeta_g \) adjust the threshold and slope of the sigmoid, respectively. The symbol \( [\cdot]^+ \) denotes half wave rectification.

Simulations of the model were performed in Wolfram Mathematica 9 on a 2.66 Ghz Mac Pro with 64 GB of memory. Numerical integration of the model dynamics was implemented in C++ using Euler’s method with a temporal step size of 0.01 sec.
Visual Displays

Figure 15.7 shows the parametrically-varied visual display sets used in model simulations. We tested the model on rectangular (Figure 15.7a), square texture (Figure 15.7b), occlusion (Figure 15.7c), cross (Figure 15.7d), C-shape (Figure 15.7e), and randomly-generated block shapes with varying complexities (Figure 15.7f–h). The rectangular, occlusion, and C-shape visual displays resemble those used in electrophysiological studies of figure-ground segregation (Zhou et al., 2000; Qiu and von der Heydt, 2007). The rectangular display set presents a simple test that should yield good model performance. Single-scale units should respond when the rectangle interiors, not exteriors, are centered within their RFs. The square texture display set tests model performance when there are multiple texture elements with various sizes and displacements. The cross and C-shape (Zhou et al., 2000) display sets test the model’s ability to produce shape-selective responses in the presence of local concavities. The block displays (Sakai et al., 2012) test the model’s ability to detect the interior of complicated shapes despite the presence of many local concavities along the irregular boundaries. We produced 500 low (LC), medium (MC), and high (HC) complexity random block displays, and 100 of each type are depicted in Figure 15.7f–h, respectively.

We parametrically varied the aspect ratio of the rectangular displays in the range $\frac{1}{8}$ to 8, yielding 64 shapes. The aspect ratio of the C-shape was adjusted in equally-spaced increments the range $\frac{1}{4}$–4 and the C-shape was 1–6 px thick to yield 96 shapes. The occlusion set consisted of two perpendicular rectangles arranged in a ‘cross-bar’ configuration, each of which took on one of two luminance values. The foreground rectangle took on two thicknesses and the background rectangle took on three thicknesses to yield 72 shapes. Finally, we generated 36 parametrically-varied cross (6 thicknesses×6 sizes) and square-texture displays (6 texture element displacements×6
element sizes).

The random block displays were generated using a modified version of a random block generation algorithm (Sakai et al., 2012). The block algorithm begins with a base rectangle and iteratively adds an adjacent block to a random location along the rectangle boundary. In the iteration following the addition of a block, locations bordering either the rectangle or newly added block may be randomly selected for the next block addition. We construct our random block displays by adding 4, 16, and 32 blocks, respectively. Greater numbers of blocks permit greater complexity in the displays by forming irregular shape boundaries with many local convex and concave regions. Five hundred unique blocks of each type were constructed in each condition.

**Figure-ground performance indices**

In order to quantify model performance across the visual display sets, we define several indices that assess properties of model single-scale unit responses. The indices measure single-cell unit (S unit) activity as a readout because single-cell units selectively respond to shapes at a particular size. Larger index scores indicate better performance. The total activity In-Out-Index (IOI) computes the ratio between the total single-scale activity on the interior of the shape and the overall total activity to anywhere in the visual display:

$$\text{IOI} = \frac{\text{Total S unit activity inside shape}}{\text{Total S unit activity anywhere}} \quad \text{(7.12)}$$

To measure the extent to which single-scale unit activity is distributed within the “surface” of shapes, compared to along the contours, we define two additional indices: surface index (SI) (Eq. 7.13) and contour index (CI) (Eq. 7.14).

$$\text{SI} = \frac{\text{Total S activity } \eta \text{ pixels away from shape border in the shape interior}}{\text{Total S activity } \eta \text{ pixels away of shape border outside the shape}} \quad \text{(7.13)}$$
\[ CI = \frac{\text{Total S activity within } \eta \text{ pixels of shape border}}{\text{Total S activity more than } \eta \text{ pixels away from shape border}} \]  
(7.14)

We set \( \eta = 1 \text{ px}. \) Larger SI scores indicate greater concentration of single-scale unit activity in the shape interior, and larger CI scores signify greater concentration of single-scale unit activity near the shape border.

### 7.3 Results

We use our model to test the following two hypotheses. Hypothesis 1: scatter in RF location and size can be used to detect shapes in neural populations connected by fast inter-areal connections. Hypothesis 2: Feedback enhances shape sensitivity by providing contextual information about the global scene layout from outside the classical RF, suppressing neural activity at the exterior of shapes, and enhancing neural activity in the interior. First, we present results that test Hypothesis 1.

**Model detection of shape exemplars**
Figure 7: Visual display sets simulated to test the model. (a) rectangular, (b) square texture, (c) crossbar occlusion, (d) cross, (e) C-shape, (f) low-complexity (LC) random blocks, (g) medium-complexity (MC) random blocks, (h) high-complexity (HC) random blocks visual displays. The displays shown in (a)–(e) varied in aspect ratio, thickness, and size. A sample of 100 of 500 total random blocks in each display set are shown in (f)–(h).
Figure 7.8: Single-scale units produce activity peaks inside shapes at the correct scale. Simulations are of shape exemplars from the visual display sets. The top row depicts single-scale unit responses with different RF sizes, and the bottom row shows the multi-scale radial feedback signals. Responses of small RF units (smaller spatial scale) are shown on the left and those of larger RF size units (larger spatial scale) are shown to the right. Blue and orange colors in multi-scale panels indicate suppressive and facilitatory feedback, respectively, that is transmitted to single-scale units that have RF centers located at each visuotopic position. The single-scale unit activity at many different scales corresponds to partial shape representations at different resolutions, and the activity peak across scale signals the detected size of the shape. The peak occurs in the interior of the shape. (a) RF sizes of single-scale units are superimposed on the square exemplar. Single-scale units with a RF size comparable to that of the square (Scale 4) elicit the maximal activity peak across scales, and the peak is positioned at the center of the square. The multi-scale units that have RFs centered on the square are most active at Scale 4 because they integrate the boundary contours. The multi-scale radial feedback signal facilitates single-scale units at the center of the square. Suppression of single-scale unit activity occurs outside the square and at smaller and larger scales than the optimal size. (b) Single-scale units that have receptive field sizes comparable to the width of the C interior (Scale 3) elicit the greatest activity. Peaks are distributed along the medial axis. Multi-scale feedback signals followed a bimodal distribution: small scales facilitate the C-shape interior and large scales facilitate the concavity. Single-scale feedback facilitates responses along medial axis and suppresses those in the concavity. (c) The cross display has arms $3px$ wide and a radius of $6px$. The largest single-scale activity peaks occur at Scale 2 along the medial axis of the cross, and the second largest occur at Scale 6 at the center. Multi-scale units produce the largest activity peaks along the medial axis at Scale 2. (d) The exemplar square texture elements have radii of $3px$. Single-scale and multi-scale unit activity peaks are concentrated at the center of the squares at Scale 3. The second highest peak occurs at the center of the display when the RF size of single-scale units equals the size of square element separation, which may indicate an grouped interpretation of the scene, rather than isolated squares. (e) The cross bars shown have the same size (width $2px$). The single-scale unit peak occurs at Scale 3 at the center of the occluding crossbar. The next highest activity peaks occur due to the occluded flanks (Scale 2).
Figure 15.7 shows the visual display sets tested by model simulations and Figure 13.8 depicts model responses to exemplars. The top panels of Figure 13.8 depict the response of single-scale units, which hold representations of grouped curved contours at a particular scale (partial shape representations). The RF sizes of single-scale units relative to the visual display are superimposed onto the top panels of Figure 13.8a. The '*' symbols are placed next to the overall single-scale activity peaks across all scales. The bottom set of panels show the activity of multi-scale units, which multiplex curved contour representations at different spatial scales. Multi-scale units integrate scatter in single-scale unit RFs radially with respect to the RF center. The spatial scale of inputs decreases with distance from the center. Units with smaller RFs are shown on the left hand side (e.g. 'Scale = 1'), and those with larger RFs are shown on the right hand side (e.g. 'Scale = 7').

Figure 13.8a shows model responses to a radius 4 px square. Single-scale units of Scale 4 elicit the maximal activity compared to those with different RF sizes. The activity peak is concentrated in the center of the square’s interior surface. Single-scale units with smaller RF sizes yield activity peaks located along the square diagonals, due to contrast signals from the boundary contours. The smaller the single-scale unit RF size, the closer the activity peaks appear to the corners. All single-scale unit activity is suppressed for units that have larger RFs than Scale 4.

The bottom rows in Figure 13.8a show feedback signals from multi-scale units. The colors correspond to the strength and facilitatory/suppressive effect of the feedback signal to single-scale units: blue indicates suppression and orange facilitation. Single-scale units receive suppressive multi-scale radial feedback at all scales other than Scale 4, in which case single-scale units that have their RFs centered on the square interior receive facilitation. The facilitatory signal is generated because Scale 4 multi-scale units that have RFs centered on the square integrate the boundary contours
Radial feedback from multi-scale units outside the square suppresses ‘false positive’ single-scale unit activity that arises due to feedforward processing. This suppression is caused by the lack of contrast along many integration directions, which results in weak multi-scale activation (Figure 15·5d).

Figure 13·8b depicts the model response to C-shape display that has a radius of 3 \( \text{px} \). The maximally active single-scale unit has a radius 3 \( \text{px} \) RF (Scale 3), and the peak response is distributed along the medial axis of the shape. Therefore, the model correctly performs figure-ground segregation. Multi-scale units with smaller RFs send excitatory facilitatory to single-scale units that have RFs located on the interior, and multi-scale units with larger RFs send facilitatory feedback to single-scale units that have RFs located on the concavity of the C-shape. Thus, the multi-scale radial feedback alone does result in correct figure-ground assignment. Single-scale units yield the maximal activity peak at the correct scale due to the single-scale feedback mechanism (Figure 15·6). Single-scale units that have RFs distributed along the medial axis of the C-shape interior receive facilitatory feedback signals from those that have larger RFs. Both large and small RF single-scale units receive input from the same curved contour units that are active due to contrast signals from C-shape boundary contours. Single-scale units that integrate the concavity are suppressed by the larger RF units because the shared curved contour unit in the concavity is inactive. Single-scale activity due to the concavity is not completely abolished due to facilitation from multi-scale radial feedback.

Figure 13·8c depicts the model response to a cross. The arms of the cross in the simulation are 2 \( \text{px} \) wide and 6 \( \text{px} \) long. Single-scale units with radius 2 \( \text{px} \) RFs demonstrate the largest level of activity, and the activity peaks are distributed along the medial axis. Large activity peaks also occur at Scales 5 and 6 at the center of the cross due to multi-scale feedback. As shown in the bottom panels, multi-scale activity
is weak outside of the cross, resulting in suppression of single-scale units. There is facilitation at the interior—particularly at scales comparable to that of the cross (Scales 5 and 6). The single-scale activity peaks at Scale 2 occur due to single-scale feedback. Notice that the Scale 2 activity peaks are located near the distal parts of the arms. Single-scale units with RF sizes comparable to the cross arm width (Scale 2) and to that of the cross (Scale 6) both receive input from curved contour units, which are active at the distal ends of the arms. This results in facilitation of the small RF single-scale units on the cross interior via single-scale feedback.

Figure 13·8d shows representative model responses to the square texture visual display, which has radius $3 \text{ px}$ square elements. The greatest magnitude single-scale unit activity peaks across scale occur at Scale 3. There are four distinct activity peaks that are located on the center of the squares. The second greatest magnitude peak occurs at Scale 7 because at that scale single-scale units have sufficiently large RFs to bridge the gap between the square elements. Multi-scale feedback is most facilitatory at Scale 3 to single-scale units with RFs centered on the square interiors.

Figure 13·8e depicts the results from model simulations to an occlusion display exemplar. The greatest magnitude single-scale unit activity peak occurs at Scale 3 on the occluding crossbar. At the scale and position of the peak, there are facilitatory multi-scale radial feedback signals due to the integration of boundary contours from the occluded and occluding crossbars. The second greatest magnitude activity peaks occur at Scale 2 on the visible portions of the occluded crossbar due to single-scale feedback.

Natural scenes and bar
Figure 7-9: Single-scale units detect peppers in a natural image and a thin bar. (a) An image of peppers from the Berkeley Segmentation Dataset (left) and single-scale unit model responses (right). The cross-scale single-scale unit peaks coincide with the interior surfaces of the peppers at different scales (Scales 4 and 5). The activity peaks indicate that the model detects the presence of the pepper shapes. When the RF size exceeds that of the peppers, the multi-scale response decreases. (b) The model detects the fine spatial scale of the bar. Single-scale activity is spatially confined to the bar, and activity does not spread over space at larger scales. Because a thin bar yields weak multi-scale unit activation (bottom panel) for all but the three smallest scales, the multi-scale radial feedback completely suppresses single-scale activity (top panel) at far distances from the bar and at scales larger than Scale 2. This is due to the same mechanism by which single-scale activity is suppressed outside of other shapes (e.g. Figure 15-5d). If the RF size far exceeds the bar extent, single-scale units do not respond. When the RF sizes are comparable in size with the bar (Scales 1 and 2), multi-scale radial feedback is weakly facilitatory and facilitates single-scale activity close to the bar.

We tested our model on more complex and simpler visual scenes to better understand the model behavior. In Figure 13-9a, we consider the model response to
a natural scene, and in Figure 13.9b we show the results of a simulation of a thin bar, which represents a limiting simple case. We wanted to verify that the spatial distribution of activity remained close to the bar and did not spread far away.

Figure 13.9a shows the activity of single-scale units with different RF sizes to a natural scene image of peppers taken from the Berkeley Segmentation Dataset (left panel). At small scales, single-scale activity is distributed close to boundary contours. The overall peak activity across scale occurs at Scales 4 and 5, and occurs when the peppers appear within the RF. The bottom row shows the multi-scale unit activity, and indicates clusters of facilitatory feedback that are spatially coincident with the pepper shapes. The greatest magnitude multi-scale unit activity occurs at Scales 4 and 5. Therefore, multi-scale radial feedback appears to support the single-scale unit activity at Scales 4 and 5. At larger scales, single-scale units elicit large magnitude responses to the wood panel below the peppers. Multi-scale unit signals that result in facilitation diminish at larger scales, indicating that the resolution of shapes in the image are too fine relative to the larger multi-scale RF sizes.

Figure 13.9b shows simulation results to a thin bar that is thinner than the smallest single-scale RF size. The top row shows that single-scale units with RFs nearby the bar are active at the two smallest scales. Most activity is concentrated at the smallest scale, closest to the bar. At larger scales, single-scale units are not sufficiently active to overcome the suppression from multi-scale feedback and are completely inhibited. The bottom row depicts multi-scale unit activity, which is facilitatory nearby the line for Scales 1–3 and suppressive at larger scales. Integration in radial directions only weakly activates multi-scale units due to the narrow width and small scale of the bar relative to the scale the larger RFs.
**The structure of single-scale unit RFs**

In order to focus on the computational mechanisms on which the visual system may depend to integrate multiple spatial scale representations of shape, single-scale units have an annular off-center/on-surround RF arrangement. This assumes that single-scale units receive afferent input from curved contour units in convex configurations. However, the visual system likely combines curved contour representations in multiple ways to construct partial shape representations. We tested how curved contour subunit orientation impacted single-scale unit activity and shape detection when a number of subunits had random orientations, which results in non-annular single-scale RFs.

In order to assess shape and spatial scale detection on the large number of visual displays, we define three performance indices. The in-out index (IOI) computes the ratios between total single-scale unit activity on to the interior of the shape compared to anywhere in the visual display. The contour (CI) and surface (SI) indices compute the ratio between the total single-scale unit activity on to the boundary contours and interior surface, respectively, and other regions of visual space.

Figure 13·10 shows how performance indices change relative to the annular single-scale unit RF configuration when afferent curved contour unit orientations are scrambled. We ran model simulations with the 500 visual displays in the HC block display set when curved contour subunits were scrambled in orientation. The locations of the subunits remained the same and only the orientation varied between 0–7. Simulations were run 20 times for each fixed number of subunits that took on random orientations. The error bars correspond to SEM. The x-axis corresponds to the number of curved contour subunits are scrambled, and the y-axis plots the percent change in each respective index score relative to the annular single-scale unit RF case.

The IOI and SI decreased in tandem with the number of subunits that were scram-
Figure 7.10: Model figure-ground performance decreases when the orientation of single-scale unit subunits are scrambled. The figure-ground performance (y axis) is plotted when the orientation of the indicated number (0–7) of single-scale unit subunits (x axis) are randomly scrambled. Simulations were run twenty times and averaged for each number of subunits that were scrambled. Performance is relative to the annular single-scale unit RF configuration (dashed line). Simulations were performed on the HC block display set. Error bars correspond to SEM. The total activity inside the HC block shapes, as indicated by the IOI and SI, decreased as the number of curved contour subunits whose orientation was scrambled increased, indicating decreased shape selectivity. The improvement in the CI performance relative to the unscrambled case indicates an increased concentration of single-scale unit activity along the shape boundary. Annular single-scale unit RFs yielded the best figure-ground and shape selectivity performance.

bled, and the CI increased. Decreases in the mean SI and increases of the CI suggest that the scrambled subunits resulted in single-scale unit activity that was more uniformly distributed across the shape surface, and was more concentrated along the boundary than in the annular RF configuration (Figure 13.8). This indicates reduced
shape selectivity compared to the annular RF case, which yielded more concentrated activity peaks. These results suggest that convex annular RF configurations may be important for figure-ground segregation and shape detection.

**Visual display set performance**

![Graphs showing visual display set performance](image)

The next set of results are from tests of Hypothesis 2: that feedback enhances shape detection.

Figure 13·11 shows model performance according to the performance indices to the low (LC, Figure 13·11a), medium (MC, Figure 13·11b) and high (HC, Figure 13·11c) complexity random block displays. We tested model performance when both multi-scale and single-scale mechanisms were intact (red), when the single-scale feedback...
Figure 7·11: Feedback greatly improved shape selectivity in the LC, MC, and HC block display sets. Performance is assessed according to the in-out index (IOI), the proportion of total single-scale unit activity on the shape, contour index (CI), the proportion of single-scale unit activity distributed along the contour, and surface index (SI), the proportion of single-scale unit activity distributed on the interior of the shape. Larger values indicate better performance, and values were computed for the optimal scale for each shape. Performance was assessed when both multi- and single-scale feedback types are included in the model, when single-scale feedback projections are lesioned, when multi-scale feedback projections are lesioned, and when no feedback is active. The presence of both feedback types greatly augmented the response to the interior of the shapes compared to exteriors, compared to when there was only feedforward processing. The multi-scale feedback only condition produced the next best performance, followed by single-scale feedback only, and finally no feedback. In displays with irregular boundaries and many concavities, such as the HC block set, the single-scale mechanism performs the best. Because the IOI is large and the SI is small in the HC block set, single-scale feedback increases single-scale unit activity on the shape interior, proximal to the boundary contours.

projections were lesioned (blue), when the multi-scale feedback projections were lesioned (yellow), and when both types of feedback projections were lesioned (green). The simulations shown in Figure 13·8 had both feedback mechanisms intact. All performance indices are normalized such that 1 indicates the best performance and 0 the worst. Each bar in Figure 13·11 represents the mean index score across the entire visual display set, and the error bars correspond to SEM. The indices are computed for the single-scale unit activity.

Overall, model simulations with both types of feedback (red) improved the interior-to-exterior response ratios the most compared to when feedback was absent (green, feedforward mechanisms only) according to all indices. Single-scale feedback (yellow) slightly improved performance compared to the feedforward-only processing condition (green). The multi-scale feedback mechanism usually improved performance to
a much larger extent (blue), and having both types of feedback connections intact (red) improved model shape sensitivity the most by a margin that often exceeded the combined performance gains obtained from having individual feedback mechanisms intact. The single-scale feedback only condition often yielded the best performance in the HC block set according to the contour index, which indicates that the single-scale feedback mechanism concentrated more single-scale unit activity near the boundary contour than with feedforward processing alone. Multi-scale feedback greatly improved the mean surface index score, which indicates that the mechanism increases unit responses to the shape interior.

When the block complexity was low both types of feedback yielded the largest scores. However, in the high complexity condition, the single-scale feedback only condition performed the best, and the relative performance of the two feedback types decreased. The higher relative performance of the single-scale only condition indicates that single-scale feedback increases the mean response inside the shape when shapes have irregular boundaries with many concavities. The increased contour index scores indicate that the activity is distributed closer to the boundary contours, but within the interior of the shape.

**Figure-ground segregation performance**

We analyzed how well our model performed at figure-ground segregation. We define ‘correct’ figure-ground assignment as when the greatest magnitude single-scale unit activity peak is located on the interior of the shape, and the spatial scale of units eliciting that peak matches that of the shape. Otherwise, we classify the model response as ‘incorrect’. Figure 13-13 shows a summary of percent correct figure-ground segregation for all visual displays as a function of feedback condition. Overall, the highest figure-ground classification accuracy was obtained when both types of feedback were included in the model dynamics. Except for the rectangular, square
Multi- and Single-scale feedback obtain the best figure-ground percent correct for all visual displays. The figure-ground assignment is ‘correct’ when the largest magnitude single-scale unit activity peak is located on the interior of the shape, and the spatial scale of units eliciting that peak matches that of the shape. Otherwise, the model response is ‘incorrect’. The inclusion of multi- and single-scale feedback types yielded the best figure-ground performance. Except for the rectangular, square texture, and LC block sets, performance decreased in the multi-scale feedback only condition. The LC and MC block displays resulted in good model performance independent of feedback, but percent correct dropped in the HC set. Both single- and multi-scale feedback and single-scale feedback only conditions increased performance in displays with concavities, such as for the cross and C-shape.

texture, and LC block displays, performance decreased when multi-scale feedback was the only type included in the model. This decrease occurred most markedly for the cross displays because the multi-scale feedback alone did not propagate a sufficiently strong facilitatory signal to single-scale units that have RFs centered on the cross interior. Single-scale feedback in combination with the multi-scale feedback yielded
the best performance.

Due to the simplicity of the LC block shapes, any feedback configuration resulted in 500/500 shapes correctly classified according to all performance indices. When no feedback was present, all indices resulted in classification errors in 3/500 (0.6%) of the displays. Overall performance is good for the low-complexity blocks. For the MC blocks, the best performance occurred in the single-scale only condition (493/500, 98.6%), followed by when both feedback projection types were intact in the model (492/500, 98.4%), the multi-scale only condition (487/500, 97.4%), and no feedback (484/500, 96.8%). Error rates increased in all feedback conditions for the HC blocks. The highest error rates were garnered in the multi-scale feedback only condition (89/500, 15.1%).

**Shape response gain due to feedback**

To quantify how feedback affected the model sensitivity to a shape’s scale, we computed the kurtosis for the distribution of single-scale unit activity peaks garnered
Figure 7·13: Feedback improves the detection of and selectivity to shape size. The mean kurtosis is plotted for the distribution of peak single-scale unit responses at each scale to shapes in each set of visual displays. Smaller values result from a broad spread in the peak single-scale unit responses across scales (e.g. “No Feedback” plot in top-right panel), indicating low confidence in the detected scale of shapes. Larger values result from greater “peakedness” in the maximal single-scale unit responses (e.g. “Multi- and Single-Scale Feedback” plot in top-right panel), indicating high confidence in the detected scale of shapes. Multi- and single-scale feedback (5/8 visual display sets) and multi-scale feedback (3/8 visual display sets) garnered the largest mean kurtosis, which was considerably larger than that obtained without feedback.

at each scale. A larger kurtosis indicates that the peak magnitudes at a small number of scales are much larger than those produced at other scales (Figure 7·13, right panel). The concentration of single-scale unit activity in few large magnitude peaks across scales demonstrates a high degree of confidence about the detected scale of a shape. A lower kurtosis indicates that the peak magnitudes did not vary much across scale (Figure 7·13, top-right panel). A broad distribution of peak magnitudes across scale demonstrates a lack of confidence in the detected scale of a shape.

In most visual display sets (5/8), the combination of multi- and single-scale feedback yielded the largest kurtosis, and there best sensitivity to the scale of shapes. The multi-scale feedback only condition yielded the next best sensitivity, and the largest kurtosis is 3/8 visual display sets. Single-scale feedback often did not yield a much larger kurtosis than the no feedback condition. This indicates that the single-scale feedback mechanism, as presently configured, did not increase the response gain across scales. Although the single-scale mechanism did not demonstrate an increased peak ratio response gain across scales, it did improve figure-ground segregation performance in displays with concavities (Figure 13·13).

In sum, the considerably larger kurtosis in the multi- and single-scale feedback
and milt-scale feedback only conditions suggests that feedback plays a crucial role in increasing the in-vs-out response gain. Feedback increased the model confidence about estimates of a shape’s scale.

7.4 Discussion

We presented a model of shape processing and figure-ground segregation in the primate ventral stream. In the present article, we tested the following two hypotheses. 1) The detection of shapes can be achieved through the structured nonlinear integration of curved contour representations at multiple spatial scales across visuotopic space. 2) Feedback enhances shape sensitivity by providing contextual information about the global scene layout from outside the classical RF, suppressing neural activity at the exterior of shapes, and enhancing neural activity to the interior. Key mechanisms in the model build on established trends in physiology of the visual system. We propose that scatter in RF locations within a visual area, as well as the large variance in RF size at any eccentricity (Piñon et al., 1998; Ungerleider et al., 2007), serves a crucial role for figure-ground segregation and shape processing. Model multi-scale units exploit RF scatter in location and size to multiplex shape representations. Our model also extensively uses inter-areal connections, which have fast conduction velocities (Hupé et al., 1998; Girard et al., 2001), for feedforward and feedback processing. Feedforward inter-areal connections in the model rapidly assemble partial shape representations, and exploit the increase in RF size with areas higher up in the hierarchy of visual areas. Quick inter-areal feedback is essential in the model to dynamically propagate contextual information about the visual scene from later to early visual areas.

Results from tests of Hypothesis 1 are shown in Figures 13.8–13.9, which indicate that the model detects the location and scale of shapes in the visual scene. Model
units are sensitive to when the interior, rather than the exterior, of perceived shapes are positioned within the RF. The location is coded by response to the shape when it appears within the RF, and the scale is coded by the RF size of model units. We also tested the shape sensitivity of single-scale units with non-annular RFs by scrambling a number of the curved contour subunits orientations (Figure 13·10). We found that the annular configuration on the high complexity random block display set yielded the best performance according to performance indices that assessed the single-scale unit response to the shape interior compared to anywhere else in the visual scene. Upon scrambling subunits such that their receptive fields no longer appeared annular, single-scale unit activity was distributed closer to the boundary contours and on the shape exterior. This suggests that RFs that have a convex shape may play an important role in figure-ground segregation.

We evaluated Hypothesis 2 by analyzing model performance on simulations of large sets of parametrically-varied shapes with lesioned or intact feedback projections. The set of visual displays we tested included shapes used in electrophysiological studies as well as randomly generated block displays in which the number of local concavities and irregularity of boundary contours increased with the shape complexity. Model shape detection and figure-ground segregation performance was gauged using a number of indices that emphasized the response gain at the shape’s interior compared to its exterior. The larger the gain, the greater the model’s shape sensitivity. Consistent with our hypothesis, feedback, whether from multi-scale units that jointly integrated visual representations across visuotopic space and scale, or from single-scale units with larger RFs, increased the peak-to-peak, mean, and total activity response gains to the interior compared to the exterior of shapes. In most cases, both types of feedback in the model yielded the greatest increase in the shape response signals. Multi-scale feedback provided a mechanism by which units with lim-
ited RF sizes may obtain contextual information about figure-ground relations that have been shown to modulate V4 neuron activity (Zhou et al., 2000; Bushnell et al., 2011; Roelfsema et al., 2012). The single-scale feedback mechanism yielded the best shape sensitivity according to a number of performance indices, particularly when the prevalence of local shape concavities increased (Figure 13·11c), which indicates that the mechanism helps mitigate responses to the exterior of figures. Asymmetric surround suppression (Walker et al., 1999; Jones et al., 2002) and the response bias of border-ownership neurons toward figures in neurons within the early visual system (Zhou et al., 2000; Zhang and von der Heydt, 2010) are consistent with responses to shapes, and not their exterior concavities.

**Comparison with existing models**

Our model is consistent with others of the primate ventral stream that employ non-linear feedforward grouping to transform simpler visual representations to those that are progressively more complex (Riesenhuber and Poggio, 2002; Pack et al., 2006; Cadieu et al., 2007; Willmore et al., 2010; Yau et al., 2012). Models of ventral areas employ feedforward processing stages, followed by a nonlinear transformation, such as a compressive nonlinearity, and normalization (Heeger, 1992; Mante et al., 2008; Willmore et al., 2010). For instance, the HMAX model of object recognition performs a series of simple and complex cell filtering operations followed by the nonlinear Max operation (Riesenhuber and Poggio, 1999; Serre et al., 2007; Riesenhuber and Poggio, 2002). Our model mechanisms are consistent with data that V4 and IT may too perform a nonlinear feedforward integration of their subunits, possibly involving a compressive nonlinearity (Pollen et al., 2002; Brincat and Connor, 2004; Mineault et al., 2012). Models that employ recurrent on-center/off-surround model networks, similar to those employed in the present model, have successfully explained V2-MT form-motion interactions (Berzhanskaya et al., 2007), attentional modulation of V4
activity (Foley et al., 2012), and neural representations of three-dimensional surfaces and boundaries (Kelly and Grossberg, 2000). The convex, annular arrangement of single-scale units is consistent with other computational (Pizer et al., 1992; Pizer et al., 1998) and biological (Craft et al., 2007; Mihalas et al., 2011; Layton et al., 2012b) models of figure-ground segregation and border-ownership. Our model extends existing descriptions of shape processing and figure-ground segregation in the ventral stream in several ways. First, unlike present hierarchical models that depend on optimization routines and training to learn synaptic weights (Cadieu et al., 2007), our model uses a template matching approach to define feedforward and feedback interactions in the network. Our model also includes dynamical interactions between model areas. Therefore, network parameters and model dynamics may readily be interpreted with respect to network connectivity and underlying mechanisms. Second, unlike existing models that exclusively rely on feedforward processing, our model contains extensive feedback, which has been shown in physiological studies to subserve critical contextual functions (Hupé et al., 1998). Third, our model defines mechanisms and dynamics by which neurons may be grouped across RF size and visuotopic location to multiplex shape representations across spatial resolution. Fourth, Cadieu and colleagues garnered shape selective responses to the visual displays used by Pasupathy & Connor (Pasupathy and Connor, 2002) in a nonlinear feedforward model based on HMAX (Cadieu et al., 2007). The model uses a cascade of simple and complex cell processes stages, and does not have units that are selective to curved contours. Our model incorporates curved contour units and provides a mechanistic explanation of how they may be associated and combined to yield more complex shape representations.

Our model provides a mechanistic explanation of how curved contour representations may be integrated in the visual system to yield selectivity to shapes and
how ambiguities related to convexities/concavities and figure-ground segregation are resolved. Consistently with a number of other models (Craft et al., 2007; Miha-лас et al., 2011; Layton et al., 2012b), we grouped curved contours representations through a nonlinear feedforward process into an annulus-shaped RF configuration. Simulations demonstrated that grouping curved contour units into an annular spatial arrangement, as opposed to other possible configurations, yielded the most sensitive shape responses (Figure 13-10). More scrambled single-scale subunits accompanied a shift in the distribution of single-scale unit spatial activity from shape interiors to the boundary. Our use of a compressive nonlinearity in the feedforward signal from curved contour units to single-scale units, and from single-scale units to multi-scale units, is consistent with physiological data. Zhang & von der Heydt showed that the contributions of individual contours on a border-ownership cell’s firing rate to a square display quickly saturated, supporting a compressive feedforward mechanism (Zhang and von der Heydt, 2010). Our nonlinear feedforward integration is also consistent with Gaussian subunit fits obtained to PIT shape-selective neurons (Brincat and Connor, 2004).

Some models of figure-ground segregation create a dichotomy between contour and surface processing (Grossberg and Mingolla, 1985a; Kelly and Grossberg, 2000). In particular, it is hypothesized that the perception of a surface arises from the activation of its boundary contours, which spreads via a diffusive filling-in process. If such process existed within the visual system that mapped onto the filling-in processes used in existing models, waves of activity are predicted to propagate between neurons in the same visual area with visuotopically adjacent RFs. Although there is support of perceptual filling-in (Pessoa et al., 1998), the existence of coherent oscillations within visual cortex, as predicted by a physiological filling-in process, has been debated (Craft et al., 2007). Cells have been identified in V1 and V2 that responded
to colored and monochrome surfaces, which is not consistent with the filling-in approach. Results from our model suggest that surface perception may emerge through the multi-scale integration of units that respond to shape parts. As shown in Figure 13-8, single-scale units elicit concentrated activity across scales on different parts of the interior of shapes, where surfaces are perceived. For example, in Figure 13-8c, single-scale units respond to the cross in different locations, ranging from the medial axis to the boundary, depending on the spatial scale. The collection of single-scale unit responses span the interior of the shape. Surfaces may be perceived through communication across inter-areal networks of neurons sensitive to shape parts at different spatial resolutions.

**Feedforward and feedback processing in the primate ventral stream**

Template matching, as used in our model, may represent a general-purpose strategy in the visual system to construct more complex representations from those that are simpler. Cells that are selective to curved contours may obtain their selectivity through a feedforward template involving end-stopped and non-end-stopped cells in V1 (Pasupathy et al., 2006), complex cell selectivity may emerge through the structured integration of simple cells (Hubel and Wiesel, 1962), and so forth. Our model makes use of templates to integrate the partial shape representations of single-scale units in radial directions across space relative to the RF center. The template structure predicts how neurons in the ventral stream may exploit RF scatter for figure-ground segregation and integrate shape representations at different resolutions to obtain scale-invariant responses. The use of radially symmetric templates have also been used to describe how area MSTd integrates feedforward inputs from MT (Layton et al., 2012a). Although the radial expansion drives many MSTd neurons (Duffy and Wurtz, 1991a), the actual templates may be far more complex (Mineault et al., 2012). The same is likely to be true of IT, if the area makes use of radial subunit
structure, as predicted by the model. Although templates define spatial feedforward and feedback interactions in the model, populations obey nonlinear temporal dynamics. Spatio-temporal dynamics may therefore represent an integral part of a neural template matching strategy to build complex shape representations.

Feedback further complicates the structure of V4 RFs. Shape-selectivity in our model is largely driven by feedforward processes, but feedback plays an important role in garnering responses to the interior rather than the exterior of shapes. Brincat & Connor report a temporal gradient in PIT neuron shape selectivity (Brincat and Connor, 2006). After the visual display onset, PIT neurons demonstrated selectivity to simple contours. Over the course of approximately 60 msec, response selectivity increased and multipart shape selectivity emerged. We estimate cortical distance between V2 and V4 to be \( \sim 20 \) mm and between V4 and IT to be \( \sim 15 \) mm (Gattass et al., 1988). A signal traveling at inter-areal conduction velocities of \( \sim 3.5 \) m/sec (Girard et al., 2001) would take at least \( \sim 11.5 \) msec and \( \sim 8.5 \) msec to complete a round trip between each respective pair of areas. 60 msec affords \( \sim 6 \) round trips for feedback signals to propagate between lower and higher ventral areas. A temporal gradient of selectivity due to feedback is also consistent with the shape responses obtained by (Hegdée and van Essen, 2006). Tuning to “non-polar” and “polar” displays were broader during onset transients and became sharper over time. Segregation between the two types of displays was maximal in multidimensional scaling plots 80–100 msec from onset. Population level response correlations decreased over time as shape selectivity increased and became more complex. The increase in response complexity may be in part due to feedback, as Hegé & van Essen report rapid firing changes followed by small modulation in the long-term responses (Hegdée and van Essen, 2006). This is consistent with the multi-scale feedback mechanism in the model. This mechanism may be implemented in neural circuits through inhibitory interneurons: when the
multi-scale feedback signal is small (large), a targeted inhibitory interneuron will exert large (small) tonic inhibition on the postsynaptic single-scale unit and therefore will inhibit the single-scale unit or release it from inhibition, respectively.

Multi-scale units nonlinearly integrate annular RF units with different RF sizes over visuotopic space. Anatomical evidence indicates that direct projections exist between V1, V2, V4, and IT cortex, including those that bypass V2 and V4 (Nakamura et al., 1993). V1 neurons possess small (~1°) RF sizes (Zhou et al., 2000). V4 RFs are roughly seven times larger than those in V1 (Kobatake and Tanaka, 1994), and projections from lower visual areas converge in IT cortex. Thus, IT neurons likely multiplex visual representations across visuotopic space and scale, as do model multi-scale units.

The focus of our model was shape detection and not border-ownership. However, the model could be configured to send feedback from model single-scale units that demonstrate scale-selective responses to the interior of figures to border-ownership cells in V1 and V2, which would propagate information about the global scene context. Feedback has been implicated as the likely mechanism by which border-ownership cells demonstrate side-of-figure selectivities, despite their small RF sizes relative to figures in the visual scene, in as short as 20 msec (Zhou et al., 2000). Model single-scale units could be configured to send excitatory feedback to nearby border-ownership cells with radially-inward side-of-figure selectivities and inhibitory feedback to border-ownership cells that are farther way to suppress responses on the exterior of shapes. Therefore, the present model is compatible with existing border-ownership models, such as the RGB model (Layton et al., 2012b), that achieves side-of-figure selectivity through extensive inter-areal connections, and neurophysiological evidence of facilitatory RF near surrounds and suppressive far surrounds (Angelucci et al., 2002; Schwabe et al., 2010).
Simulation results indicate that the single-scale feedback mechanism in our model improves figure-ground segregation in the presence of local concavities by increasing the interior versus exterior response gain and mitigating responses to the shape exterior. The mechanism facilitates or suppresses a smaller RF single-scale unit that receives input from the same curved contour subunit. We are not aware of any electrophysiological study that directly tests for neuronal facilitation or suppression based on spatially coincident subunit activity. Figure 7.14 schematically depicts an electrophysiological experiment that would support or reject the single-scale mechanism.
Figure 7·14: Proposed neural circuit and electrophysiological experiment to test the model’s single-scale feedback mechanism. (a) A neuron is recorded from that is tuned to a certain size annulus (gray). The annulus, along with another larger annulus (black), is flashed within the RF. The size of the larger annulus is varied because it is hypothesized to stimulate a larger RF single-scale cell that projects to the cell that is recorded from. Smaller than optimal annuli are also flashed because it is not clear whether the cell that is being recorded from may send or receive single-scale feedback. Reverse correlation is used to map out the relative response for various relative positions of the annuli with different temporal delays (dt). The contrast of the two annuli are either high/high, high/low, low/high, or low/low. (b) In accord with the proposed single-scale feedback mechanism, we predict that when the annuli overlap and the sides are close to tangent, there will be suppression (blue) when the small annulus contrast is lower than that of the larger annulus. This condition is most similar to the C-shape concavity case shown in Figure 15·6b. Facilitation is predicted when the contrast of both annuli are high and equal, which is similar to the interior C-shape case shown in Figure 15·6a. (c) Single-scale feedback can be implemented in a neural circuit that contains two inhibitory interneurons in the same layer as the single-scale cell with the smaller RF. When a curved contour cell is active and sends an excitatory signal to the large RF single-scale cell (left panel), the interneuron that inhibits the small RF single-scale cell is inhibited, which results in disinhibition of the small RF single-scale cell. This is similar to the case considered in Figure 15·6a. When a curved contour cell that supplies input to the large RF single-scale cell is inactive, but the large RF single-scale cell is active due to input from other curved contour cells (not shown), the interneuron inhibits the small RF single-scale cell, resulting in suppression. This is similar to the case considered in Figure 15·6b.

In the proposed experiment, the reverse correlation technique is used to map out a neuron’s RF. The technique is typically performed to identify the most likely stimulus that elicited spikes after a temporal delay when white noise is presented to a monkey on a computer display. Figure 7·14a shows how the reverse correlation paradigm may
be employed to test for facilitation or suppression when two neurons have spatially coincident subunits. Once a neuron has been located that is tuned to circular shapes of a particular size, smaller and larger annuli are flashed at random positions within the RF. Both larger and smaller size annuli need to be tested because it is unclear whether the neuron that is recorded from sends or receives feedback. The contrast of each annulus are covaried: high/high, high/low, low/high, and low/low contrasts. When the annuli overlap within the RF, the sides are approximately tangent, and the contrast of both annuli is low, we predict the neuron will exhibit response suppression. This is predicted because the presumed activity of shared curved contour cell input at the circular intersection would be small due to the low contrast, and the single-scale feedback mechanism will suppress the neuron. This is similar to the C-shape concavity case shown in Figure 15·6b. When at least one annulus is high contrast, we predict that there will be facilitation because the presumed activity of the overlapping subunit would be greater. This is similar to the C-shape interior case shown in Figure 15·6a. The demonstration of facilitation and suppression in these cases would provide support for the single-scale mechanism.

Figure 7·14c shows a neural circuit that implements single-scale feedback. Curved contour cells project to the larger RF single-scale unit and an inhibitory interneuron in the same layer as the smaller RF single-scale unit. The large RF single-scale unit projects to another inhibitory interneuron that receives an inhibitory connection from the other interneuron and sends an inhibitory connection to the smaller RF single-scale unit. When the curved contour and larger RF single-scale cells are active (Figure 7·14c, left), the inhibitory effects on the small RF single-scale cell are nulled out, which results in facilitation. When the curved contour cell is inactive but the larger RF single-scale cell is active (Figure 7·14c, right), as would occur due to the concavity in the C-shape, feedback from the large RF single-scale cell to the
interneuron results in suppression of the small RF single-scale cell.

**Model limitations**

The manner in which IT neurons combine multiple resolutions may differ from the directional integration templates employed by multi-scale units the model. In particular, the distribution of subunit sizes in each radial direction is unknown, and additional physiological work is required to determine whether regularity in IT subunits exist. *In vivo* neurons likely demonstrate variance in their subunit structure, but we predict that IT neurons, either by being prewired or through development, exploit subunit radial size gradients to produce partial shape representations. Cortical magnification and eccentricity further complicates multi-scale integration in IT. The exact gradient in RF variation in the radial directions was not essential for obtaining our results. Convex, annular RF organization also likely represents a significant simplification of the great diversity of RFs *in vivo*.

We did not directly model inhibitory subunits, though recurrent competition and inter-areal feedback in the model may afford functionally similar behavior. Others have proposed that inhibitory RF surrounds emerge through feedback, rather than feedforward, processes (Hupé et al., 1998). V4 and PIT are functionally diverse areas (Roe et al., 2012; Brincat and Connor, 2006), and the RFs of neurons likely integrate both convex and concave contour representations, which could increase the specificity of shape responses. For instance, shape-selective neurons in V4 exhibit sensitivity to both concave and convex boundary contours. Model simulations demonstrate that the annular template is sufficient to obtain shape-sensitive responses in the displays we tested.
Conclusion

In sum, we introduced a model that explains how shape processing and figure-ground segregation may interact to give rise to the detection of shapes and their size, with representations that become increasingly complex in areas further along the primate ventral stream. The model predicts that nonlinear feedforward template matching between successive visual areas builds up complexity in shape responses, and that feedback enhances shape sensitivity, transmits contextual figure-ground information, and helps resolve local shape interior versus exterior ambiguities due to convexity. Model dynamics between areas played a crucial role in how feedforward and feedback signals interacted to give rise to shape sensitivity and figure-ground assignment. Future physiological work is required both to characterize V4 and IT neuron subunit structure and to better understand how curved contour representations are combined into those of partial shapes.
Table 7.1: Parameter values used in simulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_S$</td>
<td>1.0</td>
<td>Slope of single-scale feedback sigmoidal transfer function</td>
</tr>
<tr>
<td>$\Gamma_S$</td>
<td>0.3</td>
<td>Threshold of single-scale feedback sigmoidal transfer function</td>
</tr>
<tr>
<td>$\kappa_M$</td>
<td>1.0</td>
<td>Slope of multi-scale feedback sigmoidal transfer function</td>
</tr>
<tr>
<td>$\Gamma_M$</td>
<td>2.0</td>
<td>Threshold of multi-scale feedback sigmoidal transfer function</td>
</tr>
<tr>
<td>$\zeta_g$</td>
<td>0.15</td>
<td>Shape parameter of multi-scale sigmoidal transfer function</td>
</tr>
<tr>
<td>$\Gamma_g$</td>
<td>0.01</td>
<td>Threshold parameter of multi-scale sigmoidal transfer function</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.1</td>
<td>Neuron passive decay rate</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.0</td>
<td>Neuron saturation upper bound</td>
</tr>
<tr>
<td>$\gamma_S$</td>
<td>0.4</td>
<td>Neuron inhibitory lower bound</td>
</tr>
</tbody>
</table>
Chapter 8

Independently moving objects and heading perception

8.1 Independently moving objects

Under ordinary ecological conditions, animals rarely navigate in rigid environments. That is, locomotion frequently accompanies motion in the optic flow field not due to observer self-motion. In this chapter, we consider the impact IMOs have on human self-motion perception. We investigate the mechanisms by which humans perceive their self-motion in the presence of IMOs, and present models that fit humans psychophysical data and are consistent with physiology of the primate visual system. First, we discuss the problem of detecting the presence of IMOs from retinal flow.

8.2 Detection of independently moving objects

The presence of IMOs complicates the optic flow experienced by an observer during self-motion. When the environment is rigid, observers who move along straight paths experience a radially-expanding optic flow pattern, and the FoE specifies the heading direction. Unless the depth between an IMO and observer remains constant over time, the IMO will appear to approach or recede relative to the observer and generate its own FoE or FoC, respectively. The pattern of motion produced by an IMO that maintains a fixed depth to the observer appears planar. For example, a leftward moving IMO that preserves the relative depth between it and a translating
observer will result in fixed-length leftward first-order optic flow vectors proportional in magnitude to the lateral speed.

The multitude of patterns of motion that may define an IMO raises the question of how humans detect the presence of IMOs from the retinal flow in the first place. Royden and colleagues studied how humans detect IMOs as a visual search problem (Royden et al., 2001). In the visual search literature, the strategy by which a target is detected amidst a field of distractor items has been classified either as a serial or parallel search on the basis of human reaction time (RT). When the target differs from the distractors by some “basic feature”, such as color or orientation, subjects report that the target “pops out” and the detection RT is constant and independent of the number of distractors, indicating a parallel or “very efficient” search (Treisman and Gelade, 1980; Wolfe, 1998; Wolfe and Horowitz, 2004). A serial or “inefficient” search occurs when the detection RT scales with the number of distractors and the target is more difficult to find. Royden and colleagues conducted an experiment to determine whether humans appear to use a serial or parallel search strategy to detect a stationary target disk amidst moving distractors, termed “dead fly condition” due to the analogy to a dead fly on a car windshield, or to detect a moving target disk amidst stationary distractors, termed “live fly condition” due to the analogy to a fly moving in a rigid environment. Moving items either moved in the same direction (“uniform motion”), random directions (“random motion”), or moved according to a random walk process (“Brownian motion”) whereby the direction of the linear trajectory of the item would periodically be perturbed. Royden and colleagues replicated existing studies that suggested that visual search in the motion domain is asymmetric—humans find the target easier in live fly condition than in the dead fly condition in all item motion conditions (Nakayama and Silverman, 1986; McLeod et al., 1988). In a subsequent experiment, Royden and colleagues tested visual search
of a stationary target (dead fly) amidst a radial expansion, deformation, or random field of distractors. The radial field resembled optic flow during self-motion. The deformation field is mathematically equivalent to the radial field and has the same gradient, except the vertical motion components are negated. Items in the random field moved as in the radial field, but their spatial positions were scrambled. Royden and colleagues found that humans appeared to use a serial search strategy in all cases, with RTs increasing in the deformation, radial, and random flow fields (Royden et al., 2001). Interestingly, RTs in the radial condition were shorter than in the planar cases, suggesting a more efficient search and that radial fields somehow appear ‘special’. Perhaps objects that move in the same coordinate frame as observer (i.e. stationary on retina) are seen in observer coordinates, where the observer position marks the origin. When the object moves, perhaps the visual system employs a different strategy to perceive object motion.

Interestingly, Rushton and colleagues asked subjects to perform a more complex visual search task during self-motion and found the opposite finding of Royden, that subjects employed parallel search and the target “popped out” (Rushton et al., 2007). Observers fixated a computer screen and simulated gaze was on the center of a 3D volume of textured cubes while laterally translating the head and counter rotating the eyes to maintain fixation. The authors state that the optic flow “corresponds to that you might experience if you were sat at your desk and moved sideways in your chair whilst studying a collection of flies resting on (deep) bookshelves beyond your desk.” Independent of whether the object moved in the 3D volume relative to the complex pattern of self-motion and its speed, subject RTs did not scale with the number of items in the scene. These results vis-à-vis those of Royden are puzzling, but suggest subjects were performing a different task because the RTs of Royden’s subjects were approximately double those garnered by Rushton for comparable set
sizes (Royden et al., 2001). Rushton and colleagues also found that the removal of binocular disparity (i.e. depth information) degraded RTs consistent with a serial search strategy, which may represent a difference between the studies. The item speeds are difficult to compare across studies due to differences in the gaze paradigm and may contribute to the discrepancy in the results, but Rushton and colleagues found that item speed affected mean RT but not search strategy (Rushton et al., 2007).

Royden and Connors further investigated the impact speed had on IMO detection and found that fields with radial expansion facilitated the detection of a target compared to fields with deformation (Royden and Connors, 2010). Target disks moved along linear trajectories that deviated by a particular angle from those expected by elements occupying the same position in the background field. Subject detection accuracy improved and RTs were shorter in the radial field condition than the deformation condition. Eye movements or fixation did not affect the results. Fields similar to those experienced during faster translation made IMO detection easier. Detection thresholds increased near the FoE, which suggests that the slower IMO speeds induced greater uncertainty. This is consistent with the notion that IMOs “pop out” during self-motion (Rushton et al., 2007). In a follow-up study, Royden and Moore determined that image speed of the IMO improved detection to a greater extent than depth (Royden and Moore, 2012). Together, these results suggest that the visual system detects IMOs by assessing the degree of deviation of the element from the expected motion at its spatial location in a radial expansion motion template.

8.3 Observer- versus world-centered coordinates

During self-motion, humans may either perceive the motion of IMOs in a coordinate system that moves with the observer (i.e. in observer coordinates), or in an absolute
coordinate system that remains fixed with respect to the rigid environment (i.e. in world coordinates). Our stable view of the environment when we move our heads, despite whether it contains IMOs traversing complex trajectories, suggests IMOs are perceived in world coordinates. The perception of IMO trajectory does not appear biased by movements of the body. The optic flow experienced by a moving observer in the presence of an IMO is equivalent to the sum of the fields that would be produced during self-motion in a rigid environment, and while the observer remains stationary, but the IMO moves along its trajectory. That is, compared to when a stationary observer passively views the trajectory of an IMO, the optic flow produced by self-motion biases the retinal motion pattern of the IMO. Recovering the veridical trajectory of the IMO (i.e. in world coordinates) poses a challenging underdetermined problem because the retinal motion pattern reflects the relative motion between observer and IMO (i.e. in observer coordinates). In fact, there are infinitely many possible combinations of observer and IMO trajectories that given rise to the same local retinal flow pattern proximal to the IMO.

Of course, humans only perceive a single IMO trajectory, so the optic flow component due to object motion is somehow recovered. Fajen and Matthis tested whether humans perceive IMO motion in observer or world coordinates by asking subjects whether they would pass in front of or behind an IMO in an immersive virtual reality environment (Fajen and Matthis, 2011). The authors manipulated the visual gain of motion outside the confines of the IMO—i.e. subjects experienced optic flow that would be viewed when moving 50% faster than their actual translation rate. The visual gain of the IMO motion remained unchanged. If observers perceive IMO trajectories in world coordinates and depend on the global optic flow pattern to recover the IMO motion, modulating the global flow should influence subjects’ passability judgments. Because faster translation rates accompany a greater weight of the self-
motion component in the retinal flow, subjects should indicate that they are less likely to pass the object in the higher visual gain case. Therefore, the perceived bearing angle between the observer and IMO is anticipated to decrease, indicating either a quicker collision or interception with the IMO’s path. Fajen and Matthis found that human passability judgments were consistent with this trend (Fajen and Matthis, 2011). Together, psychophysical experiments indicate that object motion is perceived in world coordinates (Dyde and Harris, 2008; Calabro et al., 2011).

Due to the linearity of the self-motion and IMO components in the composite retinal flow field, in principle, the veridical IMO motion can be recovered by “factoring out” the self-motion component. The proposal that the visual system discounts the self-motion component to recover IMO motion and trajectory has been termed flow parsing by Warren and Rushton. In a series of experiments, Warren and Rushton present psychophysical results consistent with their hypothesis that the primate visual system recovers object motion in world coordinates by “parsing” the retinal flow into self-motion and object motion components (Warren and Rushton, 2007; Warren and Rushton, 2008; Warren and Rushton, 2009a; Warren and Rushton, 2009b; Duke and Rushton, 2012). Although specific mechanisms have not been proposed, Warren and Rushton suggest that cells sensitive to radial expansion may be used to “subtract” the self-motion component out from the retinal flow to obtain the object motion in world coordinates. The flow parsing hypothesis implies that the global rather than local structure of optic flow is important for the perception of IMO motion.

Warren and Rushton tested how the presence and structure of a radially-expansive field affects the perception of object trajectory (Warren and Rushton, 2008). Subjects viewed a computer screen that contained a moving probe that moved along linear trajectories and were asked to physically adjust a paddle to the tilt angle of the probe’s motion. The radial expansion displays had a variable fraction of dots on the
left and right side of the screen. In another condition, a certain number of the dot trajectories were “swapped” or scrambled to introduce noise. When no noise was present, the symmetry of the dots had no impact on the perceived trajectory of the probe. This is consistent with a mechanism that performs a template match with a radial field to discount the self-motion component. Warren and Rushton found slight bias when all the dots were distributed to either the left or right side of the screen. Bias in the direction opposite the dot placement greatly increased in the presence of noise on same side of the screen as the dots, which could result from a low confidence template match. When the dots were symmetrically distributed with noise, there was no bias in tilt judgments, suggesting a global averaging process (Warren and Rushton, 2008). Apertures, independent of size, that locally masked the optic flow around the probe did not change the tilt judgments garnered in the symmetric full field condition (Warren and Rushton, 2009a). However, when the aperture preserved the optic flow around the probe and masked the remainder of the field, tilt judgments more closely matched the on-screen probe velocity. In other words, the recovery of object motion in world coordinates did not appear to occur only when local motion surrounded the probe.

Warren and Rushton also compared tilt judgments when observers performed eye movements to track the probe (rotation) and when observers were simulated to translate sideways (no rotation) (Warren and Rushton, 2007). Extra-retinal signals from the eye movements were hypothesized to subtract a fixed amount from retinal motion signals, independent of depth. Warren and Rushton therefore hypothesized that the signal available to globally parse object and self-motion components has velocity differences in the motion vectors at different depths. Hence, subjects should perceive different IMO trajectories as a function of the probe depth. Because the sideways translation condition does not contain rotation, extra-retinal signal cancellation does
not occur, so the vector velocity differences due to depth should be “parsed out” and probes should be judged to move along the same tilt irrespective of depth. These predictions are consistent with human tilt judgments (Warren and Rushton, 2007).

8.4 Model of heading perception in the presence of independently moving objects

In the presence of large IMOs, heading estimation performance is good, but humans make systematic errors in the judged heading direction and magnitude in some circumstances. This suggests that if flow parsing occurs, the recovery of the self-motion component in optic flow is incomplete. Warren and Saunders assessed human judgments of heading in the presence of an IMO moving perpendicularly to the path of a translating observer (Warren and Saunders, 1995). The optic flow stimuli were generated using randomly positioned dots on planar surfaces, where each dot moved in a manner consistent either with the background or the IMO. The object was always initially located 6° to either side of the center of the display and grew in size as the trial progressed due to the decreasing distance between the observer and the object. Warren and Saunders found that when the IMO obscured the observers FoE for much of the trial, subjects experienced strong heading bias in the direction toward which the IMO originated (Warren and Saunders, 1995).

Royden and Hildreth studied human heading accuracy for objects that maintain a fixed distance from the observer (Royden and Hildreth, 1996). Subjects viewed translational optic flow fields on a computer display represented by dots moving on two fronto-parallel depth planes. One experiment assessed heading judgments to horizontal object motion as a function of initial IMO position. When the object occluded the observers FoE for more than 50% of the trial, Royden and Hildreth found that subject heading biases were in the direction of object motion. Conditions
in which the IMO did not obscure the FoE resulted in virtually no heading bias. This pattern of bias for non-approaching IMOs is opposite of that yielded for approaching IMOs, where the bias is in the direction from which the IMO originates.

In the next section, we propose a neural model that simulates the approaching and non-approaching displays presented to subjects in the experiments of Warren and Saunders and Royden and Hildreth, while closely replicating the experimental conditions (Warren and Saunders, 1995; Royden and Hildreth, 1996). The goal is to identify the key neural mechanisms that give rise to the distinct patterns of bias.

The model will consist of stages corresponding to visual areas in the dorsal pathway: V1 performs local motion detection, MT pools over V1 cell motion representations, and MST will perform a radial optic flow template match on the MT representations to determine heading direction. Cells in MST have been shown to respond to large radial optic flow fields with different FoE positions (Duffy and Wurtz, 1991b; Duffy and Wurtz, 1995), and template matching in the model determines how closely MT motion signals resemble the response that would be garnered to a radially expanding optic flow field with a particular FoE position. In certain circumstances, two peaks form in MST corresponding to the FoE of the observer and IMO. Peaks represent the most active MST units and therefore are used to read out the FoE location or heading direction.

Simulations show that the model is able to produce both approaching and non-approaching patterns of bias (Layton et al., 2012a). Because the approaching IMO generates its own FoE, when the IMO obscures the observer’s FoE the two peaks in MST merge and a peak shift in the MST population peak activity occurs from the observer FoE location toward that of the object. When the non-approaching IMO moves in front of the observer’s path, two ‘false’ FoE candidates emerge at the leading and trailing IMO edges. The model produces heading bias in the direction of the
IMO motion because the two ‘false’ FoE candidates lead to a bimodal MST activity
distribution, and competition in the network suppresses the weaker of the two.

8.5 Contributions

The following section introduces the model in the form of a journal article published
in the Journal of Vision (Layton et al., 2012a). As first author of the study, I devel-
oped the model architecture, implemented and tested it on the set of visual displays,
analyzed the results, and wrote and revised the manuscript. The coauthors of the
manuscript oversaw progress and contributed in the capacity of academic advisors.
Chapter 9

A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects

9.1 Introduction

Navigation is an important activity of many species. Understanding how humans and other animals adeptly move around in their environments has stimulated much research (Warren et al., 2001; Warren, 2010). Humans can accurately estimate their direction of heading in stationary environments, to within 1-2° using visual information alone (Warren et al., 1988; Warren and Hannon, 1990; Hatsopoulos and Warren, 1991; Warren et al., 1991; Warren and Kurtz, 1992; Royden et al., 1994). Animals with eyes sample structured distributions of light over time (called optic flow) to obtain information about heading (Gibson, 1979). During forward locomotion in stationary environments, and in the absence of sources of rotation due to eye, head, or body movements, the observer experiences purely translational optic flow. Under such circumstances, the focus of expansion (FoE), the singularity in the radially expanding optic flow field, uniquely specifies the heading and direction of the linear path traversed by the observer. Although the FoE is only “visible” during forward locomotion, the optic flow field also contains a locus of inflow or focus of contraction (FoC) (i.e. where one came from), which is visible during backward locomotion (Gibson, 1979). Animals with wide fields of view may be able to view both FoE and FoC simul-
taneously. Humans and other animals must also navigate in dynamic environments, often with independently moving objects (IMOs). In the presence of large IMOs, heading estimation performance remains good, but humans make systematic errors in the judged heading direction and magnitude under some circumstances (Fajen and Kim, 2002). We present a computational neural model that explains the patterns of errors humans make in rigid environments to clarify the functional characteristics of brain circuits involved in the estimation of heading. The model performs as well in replicating the human heading estimation data from Royden & Hildreth (1996) and Warren & Saunders (1995) as the biologically-inspired model of Royden (2002), but is more representative of known neurophysiology.

9.1.1 Theories of heading estimation

There are two main classes of human heading estimation models: *motion pooling models* and *differential motion models*. The focus of these heading models is typically to explain how the FoE can be used as an indicator of heading in a general case when rotations may be present in the visual field — as is often the case with animals that can move their eyes. Motion pooling models estimate the observer’s heading by integrating motion signals over large portions of the visual field. The assumption of such models is that either the effects of rotation in the flow field on heading estimates is unimportant, or that rotation has been previously removed from the optic flow field, for example by using vestibular ocularmotor signals (Pack and Born, 2001). Differential motion models exploit the separability of the translation and rotation components of optic flow as demonstrated by the Longuet-Higgins & Prazdny (1980) retinal flow equations to remove the effects of rotation from the motion field.
\[
\begin{pmatrix}
\dot{x} \\
\dot{y}
\end{pmatrix} = \frac{1}{Z} \begin{pmatrix}
-f & 0 & x \\
0 & -f & y
\end{pmatrix} \begin{pmatrix}
t_x \\
t_y
\end{pmatrix} + \frac{1}{f} \begin{pmatrix}
x y & -(f^2 + x^2) & f y \\
(f^2 + y^2) & -x y & -f x
\end{pmatrix} \begin{pmatrix}
r_x \\
r_y
\end{pmatrix}
\] \quad (9.1)

Eq. 9.1 describes the instantaneous (i.e. first-order) optic flow for points in the world \((X, Y, Z)\) projected onto a 2D ‘retinal plane’ with coordinates \((x, y)\), assuming a planar camera model (Longuet-Higgins and Prazdny, 1980). Each point in world coordinates has a depth \(Z\), \(f\) denotes the camera focal length, and \(\vec{t}\) and \(\vec{r}\) represent the translational and rotational velocities of the observer, respectively. Suppose one samples two closely-spaced motion vectors from the global array in a rigid environment at a location at which the depth changes (a ‘depth discontinuity’). Assuming the vectors are sufficiently close in space, they should have similar translational and rotational components, but different depths. Since in Eq. 9.1 the rotational term does not depend on depth \(Z\), but the translational term does, subtracting the two vectors yields the difference in translational components, which is proportional to the size of the depth discontinuity. Supposing one samples and performs vector subtraction at several locations, one can analyze the difference vectors to recover the FoE due to observer translation. Differential motion models perform vector subtraction at depth discontinuities resulting in difference vectors that have the same angle as the original vectors, provided that the scene is rigid and the only motion is due to the observer translation (Perrone and Krauzlis, 2008). All that remains is to triangulate the “difference vectors” to obtain the estimated heading. While the angle of the difference vectors remain the same, the sign may be switched. Hence, one must have some knowledge about the direction of observer translation to differentiate the FoE from the FoC. The differential motion approach critically relies on the separability of translational and rotational components in Eq. 9.1, and on the existence of significant depth discontinuities (not gradients) in the environment. For a comprehensive review
of methods used to analyze Eq. 9.1 and their performance, please see (Raudies and Neumann, 2012).

Figure 9-1: Schematic of approaching and non-approaching IMO conditions. (a) An observer translating toward a dot plane and an approaching IMO (light gray arrow) or a non-approaching IMO (medium-gray arrow), which has a depth component of motion that equals that of the observer. (b) An approaching IMO. The solid and dashed outlines indicate its starting and ending appearances to the observer, respectively. (c) An non-approaching IMO. The solid and dashed outlines indicate its starting and ending appearances to the observer, respectively.

IMO’s may get closer to the observer, either due to observer or object motion (approaching IMO), or maintain a fixed-distance from the observer irrespective of the observer motion (non-approaching IMO). In this article we define approaching and non-approaching IMOs as the stimuli described by Warren & Saunders (1995)
and Royden & Hildreth (1996), respectively. An approaching object is the natural case, when the observer moves towards the object, the distance between the observer and the object gets shorter. This may occur when driving a vehicle on a straight course and a truck enters the vehicle’s future path from a perpendicular side street. A non-approaching object is less likely to occur in a natural setting, the motion of the observer in depth is the same as that of the IMO in that direction, as shown in Figure 9·1a. This may occur when driving a vehicle on a straight course and a car traveling at the same speed changes lanes from an adjacent lane, quickly passing in front of the observer. In both cases the optic flow describing the object motion has an FoE that defines where the object is coming from. During observer translation in the presence of an IMO, the FoE due to observer motion and the FoE due to object motion are both present in the motion field. The goal of heading estimation is to extract the FoE due to observer motion. Humans do this very well unless the IMO crosses the observer FoE and thereby reduces its visibility. Human psychophysics data provide some design constraints that we can use to determine the nature and form of the neural circuits of the primate brain that give rise to heading perception.

9.1.2 Neurophysiological background

Neurons in primate medial temporal area (MT) are functionally tuned to properties including retinal position, direction of motion, and speed (Born and Bradley, 2005). Like primary visual cortex (V1), each hemisphere features a retinotopic organization and nearly complete representation of the contralateral visual field (Gattass and Gross, 1981). V1 provides more inputs than any other area to MT (Nassi and Callaway, 2006; Sincich and Horton, 2005). Neurons in MT tend to possess receptive field sizes up to ten times larger than V1, typically 0.2°-1.2° (Born and Bradley, 2005; Zhou et al., 2000). Early researchers suspected that MT participates in longer-range motion integration than V1 does, however, recent work shows that cells in MT inte-
Figure 9·2: Proposed segregation between MT⁺/MSTD and MT⁻/MSTv motion pathways based on the neurophysiological literature. The top pathway projects to MSTD where heading-sensitive cells are located via motion pooling cells in MT⁺ and V1. The bottom pathway includes cells sensitive to differential motion in MT⁻, with on-center/off-surround receptive field antagonism, and projects to MSTv, not MSTD. Thick arrows indicate visual areas simulated in our model to estimate heading.

Grate motion over a shorter range than previously thought (Born and Bradley, 2005). Churchland et al. (2005) found that of a sample of 100 MT neurons, average second-order maps yielded spatial two-dot flash integrations of 0.62° and 0.73° for 16°/sec and 32°/sec flashes, respectively. These averages fell within 0.05° from those of sampled V1 neurons, showing sampled cells in MT appear to integrate motion over similar regions of space, despite having larger receptive fields.

Primate area MT contains at least two distinct populations of cells — differential motion cells in MT⁻ and additive motion cells in MT⁺ (Berezovskii and Born, 2000). These cells primarily differ based on the functional characteristics of their receptive fields. As depicted in Figure 9·2, differential motion cells in MT⁻ possess surround or side lobes of suppression (Xiao et al., 1997; Born and Bradley, 2005), additive cells have no such antagonism. MT cells are selective to a specific range of speeds, stimuli sizes, and directions of motion. The antagonistic zones of differential motion cells are so named because they possess the same velocity sensitivity as the other portion
of the receptive field and suppress the response of the neuron. Hence, they act like spatial differentiators in the motion domain.

Medial Superior Temporal (MST) in primates represents a functionally heterogeneous area in extrastriate cortex that receives lateral, symmetric, reciprocal projections directly from MT (Maunsell and van Essen, 1983; Boussaoud et al., 1990). Cells in MST exhibit sensitivity to translational, spiral, rotational, expanding, and contracting motion fields, up to 100° in size (Duffy and Wurtz, 1991a; Duffy and Wurtz, 1991b; Duffy and Wurtz, 1995; Nelissen, 2006; Eifuku and Wurtz, 1998). Cells located in the dorsal region of MST exhibit higher sensitivity to wider motion fields, larger areal summation, and have no antagonistic surrounds compared to those located more ventrally (Nelissen, 2006). Differential motion cells from MT primarily feed ventral medial superior temporal (MSTv), while additive motion cells in MT primarily feed dorsal MST (MSTd). MSTd contains cells that are sensitive to motion patterns consistent with the estimation of translational heading whereas MSTv does not (Eifuku and Wurtz, 1998; Berezovskii and Born, 2000; Nelissen, 2006; Orban, 2008).

9.1.3 Human heading estimation models

The potential presence of both an FoE due to observer translation (defining the heading) and an FoE due to object motion (defining the object point of origin) poses challenges to both the motion pooling and differential motion theories. Warren & Saunders (1995) and Royden & Hildreth (1996) provide important data on heading perception in the presence of IMOs that tests the predictions of the two theories. Warren & Saunders (1995) found human heading bias in the direction from which independently object emanate. Royden & Hildreth (1996) found human heading bias for non-approaching IMOs in the direction of motion — opposite of that yielded in the presence of approaching IMOs. Before these data, biologically-motivated models
aimed to explain human heading perception in the static environment. For example, the model of Lappe & Rauschecker (1993) developed the subspace algorithm of Heeger & Jepson (1990) into a neural framework. The algorithm minimizes a residual function of five image velocity sample vectors to recover heading (Heeger and Jepson, 1990). The multi-layer neural network implementation uses the residual function form as synaptic weights between “MT” and “MST” layers (Lappe and Rauschecker, 1993). Recent models have focused on human heading perception in the presence of IMOs. For example, a Bayesian framework has been developed that uses maximum likelihood to estimate observer translation and rotation in the presence of IMOs (Saunders and Niehorster, 2010). The model yields results consistent with heading judgement data (Warren and Saunders, 1995; Royden and Hildreth, 1996; Royden and Conti, 2003), but is not intended to provide an explanation of how the primate brain gives rise to the heading bias.

Warren & Saunders (1995) proposed a template-matching model of human visual areas MT and MST to explain human heading bias in the presence of approaching IMOs that performed motion pooling in MT. The template-matching approach intuitively identifies how well the global pattern of optic flow experienced by the observer matches radially-structured optic flow patterns or templates (Perrone and Stone, 1994). Since each template possesses a FoE from which the radial vectors emanate, and in the rigid environment without rotation the FoE specifies the heading, the matching procedure can be used to estimate the heading. The model defines units in MST that respond to different translational optic flow fields as a function of FoE location. MST performs a Gaussian-weighted match between the velocity-sensitive MT units activity and templates to determine the most likely focus of expansion. Since motion pooling integrates over large portions of the optic flow field, typically without regard to the presence of objects, the model predicts an averaging between
the translational and object FoE that was consistent with the data (Warren and Saunders, 1995).

While the model of Warren & Saunders (1995) explains biases in the presence of approaching IMOs, Royden & Hildreth (1996) showed it does not match the human data in the case of non-approaching IMOs (Royden and Hildreth, 1996; Royden, 2002). Royden & Hildreth (1996) instead argued that differential motion is essential to human heading perception. Rieger and Lawton developed an algorithm (Rieger and Lawton, 1985) that used difference vectors to compute heading in a rigid environment (Longuet-Higgins and Prazdny, 1980). Hildreth extended the difference vector approach by determining the heading of an observer in the presence of an IMO (Hildreth, 1992). After computing difference vectors, Hildreth defines local patches within which the center may serve as candidate FoE. By searching for evidence of vectors that lie on lines extending radially from the patch, the algorithm determines how likely the patch center is the FoE. When successful, the algorithm discounts motion due to the object, by ignoring patches of the visual field that contain inconsistent data, and considering contributions in the optic flow from the translating observer. The algorithm votes across patches and determines the most likely FoE. Note, depending on how this evidence accumulation step is performed, the algorithm can remain ambiguous whether the point is a focus of expansion or contraction along the observer axis of translation (Royden, 1997). Royden (2002) further developed the “difference vector” model of Hildreth (1992) to include differential motion operators, which act similarly to cells found in primate visual area MT with on-center/off-surround direction-of-motion antagonism. The model of Royden (2002) demonstrates human-like heading biases due to approaching and non-approaching objects, using differential motion operators inspired by cells in MT. Numerous authors now consider differential motion as the best explanation for human heading perception (Warren, 1998; Royden and
We demonstrate that differential motion operators are not necessary to explain human heading bias data. The use of differential motion operators is difficult to reconcile with the neurophysiological data indicating that differential motion cells do not appear to project to heading-sensitive area MSTd, but rather to MSTv (Berezovskii and Born, 2000; Nelissen, 2006; Eifuku and Wurtz, 1998; Orban, 2008). Figure 9-2 illustrates the pathways described in the neurophysiological literature. Mineault et al. (2012) solved an optimization problem to identify properties of expansion-selective MSTd cell subunits from MT that maximize the variance accounted for in MST cell data. Adding inhibitory surrounds to MT units do not improve the model’s ability to fit the cell data (Mineault et al., 2012), thereby supporting prior anatomical studies showing that differential motion cells do not appear to feed expansion-selective cells in MSTd. Depth discontinuities in the environment required for differential motion often do not improve heading detection thresholds (Royden and Hildreth, 1999; Britten, 2008).

Our model, based on of the Visually-guided Steering, Tracking, Avoidance, and Route Selection (ViSTARS) model, uses motion pooling in MT+ and template-matching in a competitive network in MSTd to replicate the human heading biases. ViSTARS is a model of primate visual processing describing the retina-V1-MT-MST motion processing pathway. It demonstrates how MT/MST interactions can process video input for the purposes of obstacle detection, goal approach, and the estimation of heading (Browning et al., 2009a; Browning et al., 2009b). ViSTARS is a dynamical model that explains a range of data, including the human bias demonstrated under simulating eye rotation conditions (Royden et al., 1994) and exhibits robustness to noise, but its complexity obscures the necessary conditions to explain the neural mechanisms underlying the human heading bias data (Warren and Kurtz, 1992; War-
ViSTARS unifies a number of prior models that were developed in a variety of contexts for the purposes of human navigation. For example, it integrates the FORMOTION models, which describe how V1, MT, and MST can perform motion integration and segmentation to solve the aperture problem and explain a number of visual displays with planar motion, such as the barberpole and chopsticks illusions (Baloch and Grossberg, 1997; Grossberg et al., 2001). ViSTARS also integrates the models of Chey et al. (1997, 1998), which investigate how speed perception and discrimination are affected by contrast, duration, dot density, and spatial frequency (Chey et al., 1997; Chey et al., 1998). A related model by Pack et al. (2001) shows how areas MT+, MT−, MSTv, and MSTd can interact to produce a gaze counter-flow circuit to stabilize targets while performing a smooth pursuit eye movement (Pack and Born, 2001). The effects that eye movements have on heading perception, are also explained by the precursor to ViSTARS, the STARS model (Elder et al., 2009) which uses gain-fields to compensate for the effects of eye-rotations.

Our present modeling work builds on ViSTARS to localize and explicate the simplest MT/MST neural circuits that explain the human heading bias data in the presence of IMOs. We do not address eye rotations in the model because eye movements did not affect the human heading bias results (Warren and Saunders, 1995; Royden and Hildreth, 1996), and prior models have illustrated how larger circuits may interface with MSTd to deal with rotation (Beintema and van den Berg, 1998; Pack and Born, 2001; Elder et al., 2009). Before describing our model, we first summarize the psychophysical experiments of Warren & Saunders (1995) and Royden & Hildreth (1996), then introduce the respective models that they proposed to explain these data. Our analysis shows that our model correctly replicates the direction and magnitude of heading biases at least as well as other proposed models and that the
model provides a detailed and neurophysiologically consistent explanation for how the primate brain determines heading, which no other model proposed to date can do.

9.1.4 Heading estimation in the presence of an approaching moving object

Warren & Saunders (1995) assessed human judgments of heading in the presence of an IMO moving perpendicularly (Figure 9·1a, light-grey dashed arrow) to the path of a translating observer (Figure 9·1a, black dashed arrow). Subjects viewed translational optic flow fields on a computer monitor. The optic flow stimuli were generated using randomly positioned dots on planar surfaces, where each dot moved in a manner consistent either with the background or the IMO. The object was always initially located 6° on either side of the center of the display and grew in size as the trial progressed due to the decreasing distance between the observer and the object. The object initiated movement from fixed locations and the authors altered a variable across trials called the path angle, reflecting the angular difference between the object and observer FoE. Positive values indicate that the object FoE is positioned closer to the center of the screen, and the observer FoE is further to the periphery. Warren & Saunders studied path angle settings of −6°, 0°, and 6°. Figure 9·3 shows example snapshots during object motion as a function of path angle when the object begins on the right size of the display. Subjects made left-right heading judgements following each trial, relative to a “probe” location indicated by a 1° vertical line. The authors employed a two-alternative fixed choice (2AFC) experimental paradigm, and many different observer heading and object FoE cases were tested with each path angle. Example observer and object FoE configurations are shown in Figure 9·3 (Warren and Saunders, 1995). During a trial, dots appeared in their initial locations for 1 sec to communicate the beginning of the trial to the subject. Dot motion occurred for
1.5 sec and dots lingered in their final positions until the subject responded (Warren and Saunders, 1995).

One of Warren & Saunders (1995)’s experimental conditions constrained the object movement to one side of the computer monitor such that it did not occlude the observer FoE on the opposing side. Under such conditions, subjects generated constant slightly positive heading biases (1.25°) toward the center of the screen for all path angles. Subjects yielded the same bias without an object present, and the authors concluded that the IMO does not impact heading judgments when it does not cross the observer FoE. This conclusion was supported by Royden & Hildreth (1996).

In another experiment, the object always moved on the same side of the display as the direction of observer translation, occluding it for at least some of the trial (Warren and Saunders, 1995). Warren & Saunders postulated that human heading judgments could be impacted either by the object obscuring the observer’s FoE with inconsistent motion or due to the fact that the observer FoE is not visible for some portion of the trial. To disambiguate these possibilities, Warren & Saunders employed three object types that varied in their opacity. In the opaque object case depicted in Figure 9-3, the background dots are suppressed in areas that the object obscures (i.e. the object occludes the background). In the transparent object case, the dots belonging to the object and background field coexist intermingled (i.e. the object does not occlude the background). Finally, the ‘black’ object case features no dots where the object exists. Warren & Saunders (1995) reported strong positive biases under the opaque (6°) and transparent (4°) object conditions when the path angle was set to 6°. In other words, when the object approaches the observer from closer to the center of the screen than the observer FoE and the object occludes the translational FoE, subjects experience strong biases toward the center of the screen. When the path angle was set to –6°, the authors found a stronger negative bias for the opaque
Figure 9·3: Illustrations of approaching IMO optic flow displays used by Warren & Saunders (1995) in the opaque object condition. Each row displays frames 1, 23, and 45 from a psychophysically presented motion sequence for a translational heading of 6.5°. \( \delta = -6°, 0°, \) and 6° in (a), (b), and (c), respectively, where \( \delta \) denotes the path angle defined in the text as the difference between the observer’s and object’s foci of expansion. Background optic flow is represented in red, while the IMO is depicted by blue. In the psychophysical presentations dots were the same color. Although the optic flow due to observer translation and object movement appear to commingle at the object boundary, the dots remained separate in the opaque object simulations.
object \((-2^\circ)\) compared to the transparent object \((-0.5^\circ)\). In other words, when the object approaches the observer from the edge of the screen and occludes the observer FoE, subjects experience heading biases in the direction of the edge, albeit weaker than when it approaches from the center of the screen. When the path angle was set to 0\(^\circ\), both opaque and transparent object conditions produced approximately equivalent positive heading biases of 2\(^\circ\), similar to those generated when the object did not cross the observer FoE. The black object conditions yielded a small positive bias under all path angle conditions approximately equivalent to the bias yielded in the absence of an IMO \((<2^\circ)\) (Warren and Saunders, 1995). The black object results suggest an inability to see the observer FoE alone does not induce a heading bias, but when combined with dot motion from the object error is introduced into human heading judgements.

9.1.5 Heading estimation in the presence of a moving object maintaining a fixed distance

Royden & Hildreth (1996) studied human heading accuracy for objects that maintain a fixed distance from the observer, Figure 9-1a dashed line. Subjects viewed translational optic flow fields on a computer display represented by dots moving on two fronto-parallel depth planes. One experiment assessed heading judgments to horizontal and vertical object motion. Four horizontal heading directions 4\(^\circ\), 5\(^\circ\), 6\(^\circ\), and 7\(^\circ\) were simulated on the right-side of the display, and vertical headings of 0\(^\circ\) and 2\(^\circ\) above and below of the horizontal midline of the display were tested (Royden and Hildreth, 1996). The object was opaque, possessed denser dot motion than that of the surrounding translational field, and moved with a constant speed either right, left, up, or down. For the horizontal movement conditions, the object started at one of six different starting locations. Because the object maintained a fixed-distance with respect to the observer, it only appeared to move horizontally or vertically during...
trials. Figure 9.4 depicts snapshots during different illustrative motion sequences. Subjects viewed the initial frame of the sequence, then initiated each trial via a button press. Dots remained in their final locations until the subject placed the mouse cursor in the perceived direction of motion and clicked to conclude the trial (Royden and Hildreth, 1996).

![Example fixed-distance IMO optic flow display](image)

**Figure 9.4**: Example fixed-distance IMO optic flow display used by Royden & Hildreth (1996). Frames 1, 10, and 20 from the rightward object sequence R3 are displayed (see Model section for details). At the trial outset (left), the object (blue) does not occlude the observer FoE ($9^\circ$) in the background optic flow field (red). By the final frame, the IMO completely occludes the observer FoE.

When the object moved vertically, Royden & Hildreth reported average horizontal heading biases as a function of horizontal starting position, averaged across all subjects. Positive biases and starting positions correspond to the right-side of the screen, whereas negative values correspond to the left-side of the screen. This definition of bias differs from that of Warren & Saunders (1995), who defined subject heading bias relative to the object FoE. In both upward and downward moving cases where object motion occluded the observer FoE for less than 50% of the trial, subjects produced a bias of approximately zero, similar to results garnered by Warren & Saunders (1995) under analogous conditions. However, when the object occluded the observer FoE for at least 50% of the trial, a small negative, leftward bias of approximately $-0.5^\circ$ occurred irrespective of the vertical direction of object motion.
When the object moved horizontally and occluded the observer’s FoE for more than 50% of the trial, Royden & Hildreth (1996) found different directions of bias depending on the direction of object motion. Conditions in which the IMO did not obscure the FoE resulted in virtually no heading bias. Subjects reported negative average biases (approximate maximum magnitude $-1^\circ$) to leftward moving objects that occluded the observer FoE for part of the trial (Royden and Hildreth, 1996). In the rightward moving object condition, there were positive rightward biases (approximate maximum magnitude $0.5^\circ$) when the object crossed the observer FoE for part of the trial. Therefore, when laterally moving objects obscure the observer FoE for at least some portion of the trial, human heading judgments become biased in the direction of the object motion. The direction of heading errors found by Royden & Hildreth (1996) represents the opposite of that found by Warren & Saunders (1995). Objects that maintain a fixed distance from the observer affect heading estimation differently from those that approach.

In this article we present a neural model of primate visual areas MT+ and MSTd. We use this model to unify the psychophysical data on approaching and non-approaching IMOs. As shown in Figure 9.5, our model pools motion over V1 motion representations in model MT+ and performs template-matching in a competitive network in model MSTd, maintaining consistency with the primate neurophysiological data and demonstrates human-like heading bias. Our model explains the psychophysical data through an emergent peak shift in model MSTd.

9.2 The Model

We replicate the displays shown to human subjects participating in the psychophysics studies of Royden & Hildreth (1996) and Warren & Saunders (1995). The subjects viewed optic flow displays on a monitor $p$ pixels and $w$ cm wide at a distance $d$ cm.
Figure 9.5: Model diagram. Analytic representations of the input sequence are computed in model area V1. Model area MT$^+$ pools over V1 cell responses, which feed template-matching in model area MSTd. Model MSTd cell responses are smoothed in a heading matching layer and MST units compete over time. The maximally active unit represents the best match and is taken as the heading.
Figure 9.6: Experimental setup. A model observer viewing $\alpha^\circ$ of optic flow fields from Royden & Hildreth (1996) and Warren & Saunders (1995) at a distance $d$ cm from the monitor $p$ pixels and $w$ centimeters wide (top-down view).

The studies report experimental conditions and results with respect to degrees of visual angle. We convert $\alpha$ degrees to $P$ pixels and vice-versa using

$$ P = \frac{pd}{w} \tan\left(\frac{\alpha \pi}{180}\right) \quad (9.2) $$

$$ \alpha = \frac{180}{\pi} \arctan\left(\frac{wP}{pd}\right). \quad (9.3) $$

The study of Royden & Hildreth (1996) employed an Apple 21” monitor paired with an Apple Quadra 950 workstation. We assume the study used a 21” Macintosh Color Display (19” viewable area) with 1152 x 870 pixel resolution with an approximate physical aspect ratio of 1.06\(^1\), as was standard with this type of Apple computer at that time. The viewing distance of the subjects is 30 cm. Using Eq. 9.3 we find the $30^\circ \times 30^\circ$ viewing window and $10^\circ \times 10^\circ$ IMO are $529 \times 423$ pixels and $173 \times 137$
Royden & Hildreth (1996) generated optic flow stimuli using random dots refreshed on the monitor at 25 frames/sec, each stimulus had a duration of 0.8 sec, for a total of 20 frames per trial. Royden & Hildreth (1996) used dot densities of 0.56 dots/degrees\(^2\) and 0.8 dots/degrees\(^2\) for the background and object, so we generate our backgrounds and objects with 500 dots and 80 dots, respectively. Adopting the convention used in the study, the center of the viewing window represents the origin of the image plane. As such, negative and positive positions reflect those to the left and right of the center, respectively. By Eq. 9.3, the simulated horizontal observer headings of 4°, 5°, 6°, and 7° become 69 px, 86 px, 104 px, and 121 px, respectively. In the leftward motion conditions, the IMO moved from -1.4°, 0.6°, 4.7°, 8.7°, 10.7°, and 12.7°to -7.88°, -5.88°, -1.78°, 2.22°, 4.22°, 6.22°, respectively. In the rightward motion conditions, the IMO moved from -9.9°, -5.9°, -1.9°, 0.2°, 2.2°, and 6.3°to -3.42°, 0.58°, 4.58°, 6.68°, 8.68°, 12.78°, respectively. We designate conditions in which the object moves left and right with a 'L' and 'R', respectively, and append ascending numbers to reflect the relative starting position of the object. For example, in condition L1 the object began further to the left than in L3. The initial and final positions of the IMO replicate those used in Royden & Hildreth (1996). While in motion, the object moves with a constant velocity to the end point, as described in Royden & Hildreth (1996).

Warren & Saunders (1995) used a 1280 x 1024 px monitor with a 60 hz refresh rate. Since the viewing window for the visual displays subtended 40° x 32° and were viewed from a distance of 43 cm, we find according to Eq. 9.3 this is equal to 961 x 757 pixels. Similar to Royden & Hildreth (1996), Warren & Saunders (1995) used 10° x 10° objects which is equivalent to 231 x 131 pixels on the display. Since each stimulus was

---

1Assuming a more typical 1280 x 1024 monitor with a 5:4 physical aspect ratio did not impact our results.

2We assume the objects of Royden & Hildreth (1996) had a constant size in pixels in a particular trial frame, and the object size in degrees was estimated at the center of the screen.
presented for 1.5 sec and was rendered at 30 frames/sec, resulting in a total of 45 frames per trial. We simulate the opaque, transparent, and black approaching IMO cases tested by Warren & Saunders (1995). The background consisted of 300 dots, while the object had 25 dots initially. Because both the observer and object translate at a constant speed, the approaching object grows linearly in time across frames. The object initially subtended $10^\circ \times 10^\circ$ and grew to about $20^\circ \times 20^\circ$ by the end of the trial. As described by Warren & Saunders (1995), we fix the initial position of the object to $\pm 6^\circ (\pm 137 \text{ px})$ from the center and constrained the object movement such that it begins on the same side as the simulated heading. The headings were $0^\circ$, $\pm 2^\circ$, $\pm 3-11^\circ$ in $0.5^\circ$ increments, $\pm 12^\circ$, and $\pm 14^\circ$. The object motion remained fixed, and the path angle $\delta$ varied between $-6^\circ$, $0^\circ$, and $6^\circ$(i.e. the angle between the object FoE and that of the observer). If objects in either set of psychophysics experimental conditions grew or moved beyond the viewing window, we clipped the object at the viewing boundaries.

Our simulations were performed on a 2.66 Ghz 8-core Apple Mac Pro with 16 GB RAM in Wolfram Mathematica 7. We implemented a simplified version of the ViSTARS heading model (Browning et al., 2009b), focused on the core computations involved in heading estimation.

Model V1

In model V1, we analytically compute first-order representations of the optic flow field in the non-approaching and approaching IMO conditions. Using Eq. 9.1, we set $\vec{r} = 0$, and $t_y$ to zero since the observer only translates in depth and there is no rotational optic flow in the displays reported in the psychophysical studies. Hence, Eq. 9.1 reduces to

$$I_1(x, y) := \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} x t_z - t_x \\ y t_z \end{pmatrix}. \quad (9.4)$$
In Eq 11.4, \((\dot{x}, \dot{y})\) represent the horizontal and vertical flow components at the position \((x, y)\), \(t_z\) signifies the depth component of the translational velocity of the observer, \(t_x\) indicates the horizontal component of the observer translation, \(Z\) is the distance from the observer to the point in space represented by the dot, and \(l\) specifies the \(l^{th}\) frame in the motion sequence. Although Eq. 11.4 only considers first-order information (i.e. the velocity vector field over time), the velocity field representation of optic flow can yield the same heading estimates in humans as fields containing higher-order information for dot displays (Warren et al., 1991). While we and many models assume velocity field representations of optic flow, higher-order flow may afford amenable information for navigation to the observer. We use the notation \(I_l(x, y)\) to represent the vector-valued optic flow field \((\dot{x}, \dot{y})\) at frame \(l\) with spatial location \((x, y)\).

In the non-approaching condition, we generate optic flow fields using Eq. 11.4 with \(t_z = 200\text{ cm/sec}\), the translational speed of the observer toward to the background dot planes located at 400 cm and 1000 cm (Royden and Hildreth, 1996). Each object point moved at a constant speed of 8.1 deg/sec, as described by Royden & Hildreth (1996). In the approaching IMO case, we reproduce the 5 sec time-to-contact between the observer and fronto-parallel dot plane by setting \(Z = 1000\text{ cm}, t_z = 200\text{ cm/sec}\).

Where the object exists on the display, we set \(t_z = 300\text{ cm/sec}\) to recreate the 3.33 sec time-to-contact between the observer and the object. We consider the opaque object experimental condition of Warren & Saunders (1995), as it yielded the most pronounced heading biases. In order to generate the opaque object, we replace background points in the object region with those corresponding to the object. Figures 9·3 and 9·4 and exhibit example V1 representations used in the simulations.

After converting the degrees of visual angle subtended by the optic flow displays in each respective study into pixels (Eq. 9.3), we generate uniformly sampled heading templates \(T_i(x, y)\) by substituting pixel locations \((x, y)\) into Eq. 11.4 and normalizing
each vector to unit length, with FoE at each horizontal position $i = t_x$. Figure 9-7 shows example templates used in the simulations.

![Templates](image)

**Figure 9-7:** Example heading templates. From left to right, we display $-8^\circ$, $0^\circ$, and $8^\circ$ normalized templates used in the non-approaching object simulations (30° x 30° display). Note, $0^\circ$ refers to the center of the display and positive and negative angles correspond to the FoE position on the right and left side of the display, respectively.

**Model MT**

Model V1 projects directly to Model MT, where cells have receptive fields that integrate over short-range motion of particular velocities. We define the pooled MT motion $D_l(x, y)$ according to

$$ D_l(x, y) := (I_l * G_{MT})(x, y; \mu_{MT}, \Sigma_{MT}, r_{MT}) $$

(9.5)

where $*$ denotes the 2D convolution operator, $G_{MT}$ is a 2D discrete multivariate Gaussian kernel with mean $\mu_{MT}$ and covariance matrix $\Sigma_{MT}$ normalized such that all points in the kernel’s support sum to unity, and $r_{MT}$ defines the kernel radius. We model MT cells with circular receptive fields, hence we set $\mu = \bar{0}$ and $\Sigma_{MT}$ such that $\sigma_x = \sigma_y = \sigma_{MT}$ and the covariance $\rho$ between $x$ and $y$ is zero. We employ a single parameter set to conservatively simulate MT cell receptive field properties
found in neurophysiological studies (Born and Bradley, 2005; Churchland et al., 2005). Nelissen et al. (2006) found strong fMRI responses to kinetic gratings compared to baseline conditions in area MT for spatial periods of 0.125 deg/cycle and to random textured patterns 3-28° in diameter, with responses increasing with size. We used $\sigma_{MT} = 0.05°$ and $r_{MT} = 3°$ to fit these findings. Figure 9-8 shows example frames from the motion sequences after pooling. All convolutions in MT+ are zero-padded and performed component-wise.

![Figure 9-8: Example MT representations after motion pooling.](image)

(a) and (b) show Frames 1 and 20 of the non-approaching IMO condition L4, respectively.

**Model MSTd**

In model MSTd, we perform a template match between optic flow frames $D_l(x, y)$ and all templates $T_i(x, y)$. That is, for a given $D_l(x, y)$, we match against $T_i(x, y)$ for all horizontal headings $i$. We obtain a scalar value $p^l_i$ for each frame $l$ at the horizontal heading $i$, representing the cosine similarity (i.e. inner product) between distance-weighted vectors in the motion frame and those in the template defined by
In Eq 11.7, $W_i(x, y)$ represents a distance-dependent weighting from horizontal heading $i$. We use inverse 2D Euclidean distance, scaled by a parameter $\lambda$ to adjust the spatial extent of the templates. We selected $\lambda = 300$ for a broad spatial tuning. The inner summation performs component-wise multiplication between vectors in template $T_i(x, y)$ and MT$^+$ output $D_i(x, y)$ for every spatial location and frame. The resulting vector is normalized by the $L^2$ (Euclidean) norm (denoted $||D_i(x, y)||_2$) and then the vector components $\{\dot{x}, \dot{y}\}$ are summed.

We subsequently smoothed the 1D pattern match distribution in MSTd according to

$$P_i^l := (p_i^l \ast G_{MST})(i; \mu_{MST}, \sigma_{MST}, r_{MST})$$

(9.7)

where $\ast$ is 1D cyclic convolution and $G_{MST}$ is a normalized 1D Gaussian kernel. We set the radius $r_{MST}$ to $12^\circ$, $\sigma_{MST} = 2^\circ$, and mean of the MST kernel $G_{MST} \mu_{MST} = \bar{0}$. These parameters conservatively mimic neurophysiological studies reporting greater areal summation in primate area MSTd compared to MT (Duffy and Wurtz, 1995; Nelissen, 2006).

Finally, we introduce a dynamical competitive network in MSTd. Grossberg analyzes the following network equation, termed a recurrent competitive field (Grossberg, 1973)

$$\dot{x}_i = -Ax_i + (B - x_i)(f(x_i) + I_i) - x_i \sum_{k \neq i} f(x_k).$$

(9.8)

In Eq. 9.8, $A$ specifies the passive decay rate, $B$ defines the saturation upper bound,
The signal function $f(w)$ describes the signal function, and $I_i$ defines the external input to unit $x_i$. The signal function $f(w)$ dynamically specifies the nature of the feedback a cell receives relative to its current activity. We solve Eq. 9.8 at equilibrium (i.e. $\dot{x}_i = 0$), and use a faster-than-linear signal function $f(w) = w^2$ to form a choice, or winner-take-all, network (Grossberg, 1973). We obtain recurrent MSTd units, $M_i^t$, after substituting in the smoothed pattern match distribution, $P_i^t$, and setting $A = 1$, $B = 1$:

$$M_i^t = \frac{g(P_i^t)^2}{1 + \sum_k g(P_k^t)^2}. \quad (9.9)$$

Because the minus sign before the summation term $-x_i \sum_{k \neq i} f(x_k)$ in Eq. 9.8 occurs before the factor $x_i$, the inhibitory effect cell $k$ has on cell $i$ is, at equilibrium, divisive rather than subtractive, as shown in Eq. 9.9. The function $g(x_i)$ is defined as a linear accumulation of network activity between stimulus frames:

$$g(x_i) = cg(x_{i-1}) + (1 - c)x_i. \quad (9.10)$$

Hence with Eqs. 9.9 and 9.10 we accumulate the smoothed pattern match over time. We choose $c = 0.3$ to temporally weight network activity due to new visual information higher than that of recent history. Following MSTd competition, we determine the judged heading direction by selecting the template which has the most activation in the final frame in the motion sequence. The judged heading $i^*$ is found by the following equation at the last frame $n$:

$$i^* = \arg \max_i M_i^n. \quad (9.11)$$

Finally, we determine the heading bias of the model by subtracting the judged heading $i^*$ from that generated by the network in the absence of an IMO. Because Royden & Hildreth (1996) define the sign of the bias differently than that of Warren & Saunders
(1996), we employ the conventions used in each respective study when reporting the model results. Hence, in the non-approaching object condition, we define a positive bias as a heading estimation too far to the right of the screen. In the approaching object condition, we define a positive bias as a heading estimation too far toward the center of the screen. Figure 9·9 displays some example MSTd responses along with their biases. Due to the random selection of dots in each experimental condition, we ran each configuration 10 times and averaged to obtain the reported headings.

**Figure 9·9:** MSTd responses at different times during the presentation of random-dot display sequences. Red depicts activity in the absence of an IMO, whereas green curves show the time-averaged responses with the IMO present. The top row shows the 5th, 10th, and 20th frame from left to right of the non-approaching object sequence simulating the Royden L4 condition. The observer heading is 5° to the right and a bias of -0.75° is generated in this trial. The second row shows sample MSTd responses in the approaching IMO condition, showing frames 10, 20, and 45. The observer heading is 5.5°, while the object FoE is 11.5°. The bias is 2.7° to the right, away from the center of the screen.
Figure 9·10: Simulated biases averaged across different heading conditions for the non-approaching IMO. Panels (a) and (c) show the simulation results (green) and human psychophysical data (blue) for leftward and rightward object motion, respectively. Model results for condition R6 ($\bar{x} = -2.46^\circ$) is shown in the inset of (c). Using Pearson’s correlation, we found $r = 0.94$ and $r = 0.71$ for the left and right conditions, respectively. Human subject biases were approximated from Figure 4 of Royden & Hildreth (1996). Error bars indicate 1 SEM. Panels b and d show the performance of the model of Royden (2002) (white) using asymmetric differential operators compared against that yielded by humans (black).
9.3 Results

Panels \(a\) and \(c\) of Figure 9-10 depict the average model performance when viewing the non-approaching IMO compared against the data from Royden & Hildreth (1996), adapted from Royden (2002). Panels \(b\) and \(d\) of Figure 9-10 are adopted from Figure 6 of Royden (2002), showing the model performance of Royden (2002) compared against the same human data in panels \(a\) and \(c\). Royden & Hildreth (1996) averaged across subjects and observer headings, we average over observer headings and perform 10 trials per heading. Although our model returns deterministic results given repeated presentations of the same dot motion sequence, the displays consisted of random dot patterns, which introduced inter-trial variation. Our model yields similar bias curves to those found by Royden & Hildreth (1996). Across leftward and rightward IMO conditions, we consistently obtained biases in the direction of the object motion: left for a leftward IMO, right for a rightward IMO. We obtained \(r = 0.94\) and \(r = 0.71\) using Pearson’s correlation when comparing our performance against the human data for the leftward and rightward IMO conditions, respectively. Similarly to the psychophysics results, the model produces the largest error when the IMO occludes the observer heading. When the IMO does not, or hardly, covers the observer heading during the trial, the bias is small.

Figure 9-11 displays the model performance in the approaching IMO condition averaged across observer and object headings for a given path angle \(\delta\). Also drawn are biases averaged across subjects and observer and object FoE reported by Warren & Saunders (1995); these values are approximated based on Figure 4 from Warren & Saunders (1995). The model fits Warren & Saunders’s (1995) data well, with \(r = 0.99\), \(r = 0.98\), and \(r = 0.86\) in the opaque, transparent, and black object conditions, respectively. These results demonstrate that approaching IMOs result in biases in the direction of the object FoE (Warren and Saunders, 1995).
Figure 9.11: A comparison between average subjects psychophysical data in Warren & Saunders (1995) (blue) and the average present model results (green) as a function of path angle (denoted $\delta$) for the approaching IMO condition. Subject biases were approximated from Figure 4 of Warren & Saunders (1995). Error bars indicate 1 SEM. (a) Model findings closely reflect those reported by Warren & Saunders, indicating in the presence of an IMO occluding the observer’s FoE, heading biases occur in the direction of the object FoE. Using Pearson’s correlation, we find $r = 0.99$. (b) Model results also closely match the human data when the approaching moving object is transparent ($r = 0.98$). (c) Simulation results compared to the human heading data in the black object condition ($r = 0.86$).
9.4 Discussion

We have presented a motion-pooling model of MT\textsuperscript{+} and MSTd that explains the heading biases produced in the psychophysical studies of Royden & Hildreth (1996) and Warren & Saunders (1995). As found in humans by Royden & Hildreth (1996), when non-approaching objects cover the FoE of a translating observer, the model produces biases in the direction of object motion. By contrast, when the object approaches the translating observer and covers the observer FoE, the model generates biases in the direction of the object FoE. Our model unifies the results of both studies while remaining consistent with known neurophysiology. The model also indicates that the primate visual system can determine the direction of heading by pooling motion in MT\textsuperscript{+} and competition in MSTd, without needing units sensitive to differential motion. As depicted in Figure 9·9, our model explains the two sets of data using a peak shift in the MSTd unit distributions. In the non-approaching case, the peak shift occurs because when the IMO occludes the observer’s FoE, the motion around the FoE is inconsistent with a heading in that direction. The MSTd population distribution thus has a trough around the position of the IMO, which in turn causes the peak to split into a bimodal distribution with a maximum peak on one side of the IMO. In the approaching case, the peak shift occurs due to the MSTd distributions corresponding to the observer’s and object’s FoE being close enough that they merge and produce a peak in between the FoE positions. Royden (2002) notes that her model based on MT\textsuperscript{-} differential motion cells replicates the human heading results without actively removing the IMO as would other models, such as Hildreth (1992). Our model also replicates the human data without removing the IMO to compute heading, and because it does not require differential operators, we claim that it is more consistent with neurophysiological data.
9.4.1 Approaching IMO

When the approaching IMO occluded the observer’s FoE, the optic flow pattern attributed to the IMO progressively became more influential during the trial due to the increase in size of the IMO’s representation in model area V1. Peak activities in MSTd units reflected this trend by beginning closer to the observer’s FoE and over time shifting toward the IMO’s FoE. The fit we obtained in the approaching IMO simulations ($r = 0.99$ opaque IMO, $r = 0.98$ transparent IMO, $r = 0.86$ black IMO) was not surprising, since the model of Warren & Saunders (1995) pools motion and also explains also these data. Since the black object had no dot motion defined within its boundary, we obtain a relatively flat heading bias curve as a function of path angle, also mimicking the decrease in bias as path angle increased seen in the human data. The bias curve in the black object condition is not flat because of inter-trial variation and positional effects attributed to the object always beginning $\pm 6^\circ$ from the center of the display. We discovered relative speed of the object compared to that of observer translation and the amount of motion pooling in MT$^+$ altered the model performance. For a given path angle, increasing the speed of the approaching object tended to globally shift the biases produced for all tested observer and object FoE pairs. This is because pooling locally disperses motion direction contributions, consequently increasing neighboring template match scores. Adjusting the amount of MT$^+$ pooling ($r_{MT}$, $\Sigma_{MT}$; Eq. 9.5) had large effects and influenced each observer and object FoE pair based on the path angle and context. Similarly, as $\Sigma_{MT}$ adjusts the model MT$^+$ cell spatial integration extent, this parameter may largely shift the MSTd cell template match scores. By virtue of the constraint on the visual displays that the observer heading be on the same side as the approaching object, the motion pooling and template match distributions in MSTd were usually unimodal due to smoothing that merges proximal match activity. Network accumulation in MSTd ($c$; Eq. 9.10)
also impacted heading biases by influencing the temporal sensitivity of match scores. Our selection of $c$ allowed the network to integrate information over time, but not disregard the recent past. Less smoothing in MSTd ($\sigma_{MST}$) also increased the network sensitivity to peak shifts and to other changes in the match scores between frames. Furthermore, we observed an expected symmetry of trials conducted on the right and left sides of the screen. That is, if we reflected each frame of an approaching moving object sequence about the center of the screen, we obtained the same biases. This is not true of the non-approaching object due to the lack of positional symmetry in the design of the study (Royden and Hildreth, 1996). Interestingly, biases remained insensitive to a variety of dot densities, echoing the findings of Warren & Hannon (1990) that dot density did not impact percent correct performance in their 2AFC paradigm.

![Figure 9-12: Model response to a receding IMO. (a) The first frame in the sequence. (b) The last frame in the sequence. (c) The response to the receding IMO in model MSTd. Our model predicts a heading bias in the direction opposite of the FoC relative to the observer FoE position (green) compared to when no object is present (red).](image)

Although approaching IMOs have been studied in the literature, we do not know
of any thorough investigations of receding IMOs. Figure 9-12a – b show the first and last frames of a receding IMO sequence. Our model predicts heading bias in the direction opposite of the FoC relative to the observer FoE position. Our analysis indicates that differential motion models, such as that of Royden (2002), may not make the same prediction because the direction of bias is dependent on the relative speed between the IMO and background. For example, if the receding object speed is much greater than that of the background (perhaps similar to the right side of the IMO in Figure 9-12a), the object vectors dominate in the vector subtraction and the heading estimate will be biased toward the object.

9.4.2 Non-approaching IMO

MSTd template-match distributions in the non-approaching condition were often bimodal, due to ‘good’ matches immediately around the discontinuities between the object and the background (i.e. motion boundaries). The activation within the object boundaries was reduced because MSTd units obtain suboptimal pattern matches when sampling within the IMO’s extent and become suppressed in the competition. Depending on the amount of pooling, the proportion of the trial that the object occludes the observer FoE, and the amount of competition in MSTd, the heading that gives rise to the “surviving” peak in MSTd may change. For example, strong competition magnifies small differences between both peaks because the recurrent competition field with a faster-than-linear signal function must make a choice (Grossberg, 1973). Additionally, before or after the IMO passes over the observer FoE, weaker competition can either expedite the dominance of an emerging peak or stronger competition can prolong the dominance of an existing peak. We use distance-dependent weighting to help fit the human data. This is unlike the model of Royden (2002), which relies on distance-dependent weighting to prevent the network from producing biases when the IMO is positioned far away from the translational FoE. The recurrent competitive
field and motion accumulation after the MSTd pattern match preclude such biases in our model because the object would produce relatively low match scores compared to the visible translational FoE and hence lose the competition over time (Royden, 2002; Grossberg, 1973). Our model response to the R6 condition did not fit the magnitude of heading bias reported experimentally, although our model, unlike the model of Royden (2002), matched the direction of bias. We note that the model of Royden (2002) also deviated on the R6 condition, producing a bias of the opposite sign of that reported by Royden & Hildreth (1996). During our analysis we were able to obtain a better fit to the Royden & Hildreth (1996) data using a different set of parameters \( r = 0.86 \), however, with these parameters the fit to the Warren & Saunders (1995) data was reduced. Parameters in the present study were chosen to match the psychophysical and neurophysiological data with the minimal number of parameters, and to utilize a single set of values across all our simulations. This reduces the model complexity and allows for greater understanding of the computations taking place. On-going research is investigating how multiple sets of receptive fields may interact within MT and MST, and how best to parameterize them within the model. Although the results of Royden & Hildreth (1996) indicate human heading bias in the direction of object motion when the observer’s FoE is occluded, the population vector of monkey MSTd cell responses may only reflect heading error if the object motion greatly deviates from that of the surrounding optic flow produced by the observer translation (Logan, 2006; Georgopoulos et al., 1986). In other words, MSTd cells in monkey may only yield a biased representation in a subset of the non-approaching IMO cases tested in this article. Our model can account for these differences with a change in parameters.
9.4.3 Optic flow illusion

Superimposing fields of radially expanding and laterally moving dots, often on different depth planes, forms an effect known in the literature as the optic flow illusion (OFI) (Duffy and Wurtz, 1993). The perceived FoE shifts in the direction of the lateral dot motion proportional to its speed (Pack and Mingolla, 1998; Royden and Conti, 2003). Royden & Conti (2003) claim this supports the hypothesis that the visual system performs local differential motion. Numerous manipulations, such as superimposing two radial fields, separating the fields by a gap, and changing the lateral field to a rotating field, have also been investigated (Royden and Conti, 2003; Duijnhouwer et al., 2006; Duijnhouwer et al., 2008). The superposition of two radial fields may be interpreted as laterally sliding the closer plane uniformly as the observer translates straight ahead, which shifts the perceived FoE in the direction opposite of the planar movement. Our model shows the same direction of bias as the psychophysics data on the two superimposed radial field case (Royden and Conti, 2003). In the simplified form described here, our model cannot account for the original OFI because full field lateral dot movement induces zero bias for slow lateral dot speeds and a slight bias in the direction opposite of the dot motion direction for faster speeds. However, we argue that the OFI may arise due to global motion integration, since the context may reflect that of visual stability during a smooth pursuit eye movement (Duffy and Wurtz, 1993; Pack and Mingolla, 1998). This context significantly differs from that of the IMOs analyzed in the present article, wherein optic flow due to object translation surrounds that of the IMO, and the visual system may therefore use different visual circuits. We believe that adding a smooth pursuit counter-flow stage, such as that proposed by Pack and colleagues (Pack and Born, 2001), multiplicatively combining retinal and extra-retinal signals (Beintema and van den Berg, 1998), or using gain fields (Elder et al., 2009) in the model would allow the model to account for the OFI.
human data (Royden and Conti, 2003). Whether counter-flow alone is sufficient to explain the human data on the original OFI, or its modifications, will be dependent on the parameterization of the modified model.

9.4.4 Timing

We assume that heading judgements in the model are made after the final frame is presented. Analogously, this assumes subjects exclusively decide on their direction of heading after viewing the information present in the final frame of motion, but humans may decide on a heading at any point during the trial and may ignore some of the available information. In fact, Royden (2002) observes that the psychophysical data reported here fits her model better earlier in a trial. In the approaching moving object condition, it would seem plausible if humans valued early information due to the object expansion that progressively obscures the optic flow field (Royden, 2002). When the object does not approach the observer, one can imagine later frames of the display providing more reliable information about the translational FoE if the object initially obscured but later moved away to reveal it (Lappe, 1998). Royden & Hildreth (1996) and Warren & Saunders (1995) employed similar presentation protocols as summarized in the introduction, with the exception of the differing means of response (cursor clicks compared to left-right judgments) and the fact that subjects in Royden & Hildreth (1996) clicked a button to initiate the trial whereas in Warren & Saunders (1995) the trial began automatically. Our model currently samples the visual field uniformly when performing template matches, however, neurophysiological evidence exists that more cells in MST may have a more peripheral preferred azimuth of FoE, but exhibit a greater sensitivity foveally (Gu et al., 2010). In a study requiring eye fixation, such as Royden & Hildreth (1996), template sampling and weighting differences may change the pattern of results.
9.4.5 MSTd cell types

At present our model does not include the full variety, or complexity, of cells found in primate MSTd. In order to keep the model simple and directed toward the assessment of self-motion, we have focused on radially expansive cells with receptive fields that cover most, if not all, of the visual field. No additional properties of MSTd cells were required to explain the human data discussed herein. Evidence suggests that a number of MSTd cells that exhibit sensitivity to radial expansion not only respond in the context of self-motion, but also to aspects of object motion in the scene. The response of individual MSTd neurons may be a complex combination of local object and global motion (Sato et al., 2010). While the response to object motion in MSTd is related to the work we present here, it does not seem necessary to explain human heading biases in the presence of IMOs. Although MST cells of differing heading direction preferences in our model inhibit each other via the term \(-x_i \sum_{k \neq i} f(x_k)\) of Eq. 9.8, which has a minus sign, the effect is neither global nor local subtraction (Royden, 2004). Our model uses divisive rather than subtractive normalization, as can most readily be seen in Eq. 9.9 due to the shunting inhibition by \(x_i\) in Eq. 9.8, which has different effects than subtraction (Grossberg, 1973; Levine and Grossberg, 1976; Heeger, 1992). Future work will clarify the contexts within which MSTd cells respond to local object or global motion and how this may influence navigation.

The model of MT+/MSTd that we present in this article demonstrates that human heading biases estimated for approaching and non-approaching IMOs can be explained using motion pooling and template matching in a competitive network while remaining consistent with known neurophysiology. Differential motion processing is not necessary to explain these data in the presence of IMOs.
Chapter 10

Heading perception and attention

10.1 Attentional signals in MSTd

Dubin and Duffy investigated the responses of single neurons in MSTd when monkeys were presented with radial motion patterns and were primed to attend to locations of the visual field some distance from the FoE (Dubin and Duffy, 2007; Dubin and Duffy, 2009). The researchers found that neurons selective for the FoE position showed an increased response when the monkeys fixated the center of the visual display and had to later saccade to the FoE location (behaviorally relevant condition). Covert attention was assumed to travel between the fixation and FoE locations during each trial because the FoE appeared randomly in one of eight locations at 30° eccentricity about the fixation point. In behaviorally irrelevant trials, an attentional prime appeared and disappeared in one of the FoE locations prior to the optic flow display onset, and the monkeys had to saccade to the prime location after the optic flow presentation. The prime did not always coincide with the forthcoming optic flow FoE location. The firing rate of neurons tuned to the FoE position was enhanced, when the prime was close to the FoE compared to when it was far away. The timing of the peak response of MSTd neurons was related, almost linearly, to the distance in visual angle between the prime and the FoE ($r^2 = 0.89$) (Dubin and Duffy, 2007; Dubin and Duffy, 2009).

We present a neural model to address the effects of spatial attention on the neural dynamics, peak population response latencies, and tuning curves of MSTd cells
reported by Dubin and Duffy. The model contains stages that correspond to primate visual areas V1, MT, MSTd, and frontal eye fields (FEF). In model V1, units detect local motion (Livingstone and Conway, 2003). Model V1 projects to model MT, wherein long-range motion pooling over the model V1 responses over time occurs. Model MT projects to model MSTd, which has units that respond to large patterns of visual motion (Duffy and Wurtz, 1991b). FEF, which projects to MSTd, is included in the model to provide top-down attentional signals related to the planning of saccades. The goal is to accurately model FEF's modulation of MSTd in the context of the Dubin and Duffy experiments, but not produce a comprehensive model of FEF.

The model simulates the linear recession of the mean MSTd population peak timing as an emergent phenomenon as distance between attention and the optic flow FoE increases. The bottom-up optic flow signal and top-down attentional signals both produce spatial distributions of activity in MSTd. When the signals are centered on similar locations in visuotopic space (attention near FoE), the two signals constructively interfere and result in an earlier peak in the mean population activity. When the signals target MSTd neurons that are far apart in visuotopic space, the spatial distributions will overlap less and the bottom-up signal takes longer to peak due to higher uncertainty and competition in the network.

The work investigates what produces the linear peak recession by analyzing the spatial dynamics in MSTd.

10.2 Contributions

The following section introduces the model in the form of a journal article published in Frontiers in Computational Neuroscience (Layton and Browning, 2012). As first author of the study, I developed the model architecture, implemented and tested it on
the set of visual displays, analyzed the results, and wrote and revised the manuscript. The coauthor of the manuscript oversaw progress and contributed in the capacity of an academic advisor.
Chapter 11

Recurrent Competition Explains Temporal Effects of Attention in MSTd

11.1 Introduction

Neurons in the dorsal medial superior temporal area (MSTd) of primate visual cortex selectively respond to radially expanding random dot patterns that span large parts of the visual field (Duffy and Wurtz, 1991b; Duffy and Wurtz, 1995). Gibson noted that animals experience radially expanding patterns of motion during navigation along a straight path, in the absence of eye movements. The center of the radial motion is known as the focus of expansion (FoE) and defines the direction of travel (heading) during locomotion (Gibson, 1979). Primate MSTd neurons exhibit selectivity to the FoE location, and researchers have proposed that these cells are important for heading perception during visually guided navigation (Duffy and Wurtz, 1991b; Born and Bradley, 2005). Neurophysiological data indicate that in primate heading sensitive MSTd cells receive feedforward projections from cells sensitive to local motion in V1 via cells that integrate motion over a large receptive field in primate medial temporal area, MT (Eifuku and Wurtz, 1998; Berezovskii and Born, 2000; Orban, 2008). Objects that move independently of the observer in the environment induce distinct patterns of motion that differ from the patterns experienced by the observer in the object’s absence. The object induces its own FoE when the independently moving object approaches the observer. Depending on whether the object approaches or re-
cedes, the object size, and other contextual information, the observer may shift his focus to different aspects of the environment (Kishore et al., 2011; Wann et al., 2011). While the effects of visual attention have been studied in ventral areas (e.g. V4) and early dorsal areas (e.g. MT), the role attention plays on MSTd neurons and visually guided navigation has only been recently examined.

**Attention Affects the Gain of Individual Neurons**

Visual attention has been characterized in the psychological literature as a spatial ‘spotlight’ with limited resources (Posner et al., 1980), which enhances a subject’s visual search and luminance detection performance (Yeshurun and Carrasco, 1998) and may reduce the subject’s response latency (Treisman, 1998). Recent neurophysiological experiments have demonstrated that attention directly modulates the activity of individual neurons throughout the occipital and parietal cortices (Corbetta and Shulman, 2002). In primate visual area V4, the response of neurons increases to both an attended “target” object and proximally located behaviorally irrelevant objects (Connor et al., 1997). When low contrast objects are presented within the receptive field, attention increases the activity of neurons to the same levels that would occur in response to objects of higher contrast (Reynolds and Chelazzi, 2004). Neural signals, which modulate the gain of visual neurons in extrastriate cortex, may originate from top-down sources further up the visual pathways, since changes in a neuron’s activity take ~ 70 msec (Martinez-Trujillo and Treue, 2004) after the visual display is presented to the subject. The apparent response gain modulation observed in visual neurons is consistent with the idea of resource limitation or normalization of activity (Reynolds and Heeger, 2009), since the increased neural activity to an attended object may accompany a diminished response to competing objects (Reynolds and Desimone, 2003).
Multiplicative or Non-multiplicative Attention

While attentional gain modulation has been well documented, exactly how an attentional signal acts on baseline neural responses remains unclear. Some researchers have proposed that spatial attention *multiplicatively* influences sensory bottom-up signals such that the size and position of the receptive field of a neuron do not change, but the preferred response and tuning curve distributions do (Martinez-Trujillo and Treue, 2004; Treue and Martinez Trujillo, 1999; McAdams and Maunsell, 2000; Williford and Maunsell, 2006). For instance, the response to the preferred motion direction of single neurons in primate visual area MT has been shown to increase when the monkey attends similar directions of motion outside the cell receptive field, and the response to anti-preferred motion decreases — suggestive of multiplicative changes in the tuning curves (Martinez-Trujillo and Treue, 2004). Others have proposed that attention acts *non-multiplicatively* on the sensory signal, which may change not only the tuning properties of neurons but also the spatial extent of their responses (Womelsdorf et al., 2006; Womelsdorf et al., 2008). For instance, the receptive fields of MT neurons have been shown to shift depending on whether the subject attends objects inside or outside the receptive field.

MSTd and attention

Dubin and Duffy investigated the responses of single neurons in MSTd when monkeys were presented with radial motion patterns and were primed to attend locations of the visual field some distance from the FoE (Dubin and Duffy, 2007; Dubin and Duffy, 2009). The researchers found that neurons selective for the FoE position showed an increased response when the monkeys fixated the center of the visual display and had to later saccade to the FoE location (*behaviorally relevant condition*). Covert attention was assumed to travel between the fixation and FoE locations during each trial.
Figure 11.1: Mean MSTd population firing rates in the behaviorally relevant (red), behaviorally irrelevant (blue), near (green), and far (cyan) conditions. The timing of the mean MSTd population responses in the near (135 msec), behaviorally relevant (216 msec), and far (312 msec) experimental conditions, respectively, after the optic flow display appeared was approximately linear. Data extracted from Figure 2c-d of Dubin and Duffy (2009) and combined into a single figure.

because the FoE appeared randomly in one of eight locations at 30° eccentricity about the fixation point. In behaviorally irrelevant trials, an attentional prime appeared and disappeared in one of the FoE locations prior to the optic flow display onset and the monkeys had to saccade to the prime location after the optic flow presentation. The prime did not always coincide with the forthcoming optic flow FoE location. The firing rate of neurons tuned to the FoE position was enhanced, when the prime was close to the FoE compared to when it was far away. Figure 11.1 shows the MSTd population firing rate results for the behaviorally relevant (red) and irrelevant conditions (blue), averaged across all neurons, showing an effect for behaviorally relevant or irrelevant trials. This plot is derived from Figure 2c-d in (Dubin and Duffy, 2009). Figure 11.1 also shows the effects of distance between the attentional prime and the FoE in the behaviorally irrelevant condition, where the prime was located at 0° (near, green) or 60° (far, cyan) eccentricity. The timing of the peak average population re-
response of MSTd neurons was related, almost linearly, to the distance in visual angle between the prime and the FoE ($r^2 = 0.89$). The green, red, and cyan curves peak at 135 msec, 216 msec, and 312 msec, respectively. These peaks correspond to the focus of ‘attention’ at 0°, 30°, and 60° eccentricity relative to the monkey fixation point. Although the mean MSTd population firing rates exhibit distinct peaks when the prime appears near or far from the FoE location, behaviorally irrelevant trials collectively yielded a flatter population response (blue). Neurons that were not strongly tuned to the FoE location in the near (gray, solid) and far (gray, dashed) prime conditions showed lower mean population activities compared to neurons tuned to the FoE. Why does the mean MSTd population peak recede in time, and by what mechanisms?

This article introduces a computational model of MSTd to mechanistically explain why the mean MSTd population response peak recedes in time as the visuotopic distance traveled by attention increases.

Our neural model addresses the effects of spatial attention on the neural dynamics, peak population response latencies, and tuning curves of MSTd cells (Dubin and Duffy, 2007; Dubin and Duffy, 2009). The model contains stages that correspond to primate visual areas V1, MT, MSTd, and frontal eye fields (FEF). In model V1, units detect local motion (Livingstone and Conway, 2003). To focus on the dynamics of MSTd, we employ an analytic motion vector representation in model V1. Model V1 projects to model MT, wherein we perform a long-range motion pooling over the model V1 responses over time. Cells in primate MT elicit aperture-resolved responses to large fields containing uniform velocity patterns. Model MT projects to model MSTd, which has units that respond to large patterns of visual motion (Duffy and Wurtz, 1991b). We include model area FEF, which projects to MSTd, to provide top-down attentional signals.

The spatial priming paradigm of Dubin & Duffy involves saccadic planning, de-
cision making, and expectation formation in monkey subjects. Area FEF in both non-human primates (Krauzlis, 2005) and humans (Corbetta et al., 1998; Corbetta and Shulman, 2002) has been strongly implicated in goal-oriented top-down selection and spatial orientation (Schall, 2004) in a tightly integrated ocular-motor and attention neural circuit that projects to MST (Colby et al., 1988). Our model predicts that the attentional signals that modulate MSTd neurons in the experiments of Dubin & Duffy originate in FEF. The effects of attention have been documented in earlier visual areas, such as V1 (Paradiso, 2002), however, there are no data from V1 (or MT) during the modeled experiments and so we do not attempt to address those attentional effects. We demonstrate that recurrent competition between units in MSTd explains the data of Dubin and Duffy, irrespective of the particular attentional signal used. Deflections in the population temporal activity are predicted to signify dynamic shifts in neural activity between units sensitive to the FoE and the prime locations, in other words as attention shifts from the prime to the FoE.

11.2 Materials and Methods

Our simulation conditions mimic the monkey behavioral paradigm of Dubin and Duffy (Dubin and Duffy, 2007; Dubin and Duffy, 2009). In their experiments, two rhesus monkeys viewed radial dot optic flow displays composed of 1000 dots moving at \( \sim 40^\circ/\text{sec} \) on a \( 90^\circ \) tangent screen. Single cells were recorded from area MSTd 50-300 msec following the onset of the optic flow display. The optic flow FoE appeared in one of eight locations regularly spaced by 45 degrees eccentricity around a centrally located fixation point. The monkeys maintained fixation for 2 sec during a trial, then the optic flow display appeared for 1 sec. Monkeys were required to fixate within a \( 2^\circ \times 2^\circ \) window of the center of the screen and were rewarded for completing a saccade task. On behaviorally relevant trials, a grid of eight targets, each corresponding to a
possible FoE location, followed the optic flow display and the monkeys had 500 msec to saccade to the target that marked the position of the FoE during the trial. Dubin and Duffy interleaved behaviorally relevant trials with those that were behaviorally irrelevant. A square randomly appeared in one of the eight possible FoE locations (the prime or cue) for 1 sec, and disappeared before the optic flow display onset. A delay of 150-300 msec was introduced between the prime and optic flow presentations. Instead of being instructed to saccade to the location of the FoE, the monkeys were trained to saccade to the location of the prime, irrespective of the FoE location. Due to the random positioning of the prime, it could have occupied the three nearest (near condition) or farthest (far condition) positions relative to the FoE during the trial. The near and far conditions therefore constituted a subset of the behaviorally irrelevant trials. Monkeys saccaded to the prime location more than a second after the prime disappeared, hence it is assumed that some attentional signal was maintained on the prime location until the trial concluded. Behaviorally relevant trials did not include a prime, but it is assumed that the monkeys attended the centrally located fixation point. The monkey had no information to predict where the FoE would appear. We simulated the two experimental conditions of Dubin and Duffy: behaviorally relevant and behaviorally irrelevant trials.

Dubin and Duffy analyzed neuronal tuning curves as a function of FoE location, and the average firing rates of neurons over time across the trials were derived for subpopulations of cells that showed statistically significant effects for the behaviorally relevant and behaviorally irrelevant trials (near and far). Significance was assessed using analysis of variance (ANOVA) with Greenhouse-Greyser correction for nonspherical variance. Dubin and Duffy recorded from 135 MSTd neurons in total, and 32 cells showed significant effects to conditions assessed in the study. With respect to experimental condition (behaviorally relevant and irrelevant), 16 exhibited significant
effects for a single condition, 6 showed significant effects to both conditions, and 6 showed a task by FoE location effect (28 cells total). With respect to the relative position between the FoE and the prime, 11 cells exhibited significant effects for near vs. far, 4 for behaviorally relevant vs. irrelevant task types, 1 for both types, and 6 for task by FoE location (22 cells total).

Figure 11·2: A sample optic flow field used in the model simulations. The base of each arrow represents the instantaneous position of one of 1000 dots, the direction indicates the dot’s instantaneous direction of travel, and the arrow length is proportional to the speed.

In our simulations, we generated 1000 dot radial optic flow displays that occupied 256 x 256 pixels, as shown in Figure 11·2. Following the protocol of Dubin and Duffy, we constrained the dot movement speed to follow a $\cos(\theta)\sin(\theta)$ function, where $\theta$ denotes the visual angle between the observer’s gaze or line of sight on the FoE and each dot position located 1 m in depth (Duffy and Wurtz, 1991b). After adjusting the speed of each dot according to this formula, we scaled the dot speeds such that the average velocity of all the dots over the entire trial was $\sim 40^\circ$/sec. Our results were the same whether we constrained dots to move at an average fixed speed ($\sim 40^\circ$/sec) as
Dubin and Duffy did or simulated random dots observed at a walking speed toward a simulated fronto-parallel plane 1 m in depth. We simulated the dynamics of 128 MSTd neurons, each having a receptive field centered on equally spaced positions along the middle horizontal axis of the display. Each MSTd unit had a large 90° x 90° receptive field, consistent with neurophysiological data (Duffy and Wurtz, 1995). To analyze the model MSTd population results, we averaged the activity of all neurons over 500 msec following the onset of the optic flow display. For simplicity, we simulated FoE locations only along the center horizontal axis. In order to derive neuronal tuning curves, we selected ten neurons tuned to similar FoE and analyzed their responses to behaviorally relevant and behaviorally irrelevant trials whereby the optic flow field FoE occupied one of nine evenly spaced locations along the 90° middle horizontal axis.

Our selection of ten neurons is comparable to the eleven used by Dubin and Duffy. For each FoE location and trial type, we averaged the neurons’ activity 50 msec-250 msec after the onset on the optic flow display.

In the behaviorally relevant condition the monkeys could not anticipate the FoE location and in the behaviorally irrelevant condition the prime disappeared over a second before the monkey had to saccade. We find it unlikely that the difference in results can be attributed to residual bottom-up activity. We therefore assume that the monkeys attended the prime position between the time it appeared and when the saccade was made. Some spatial (attentional) neural signal indicated whether the monkey should saccade to either the prime or FoE location. In our behaviorally relevant trials (without a prime), we simulated a spatial distribution that enhanced the activity of neural units sampling the FoE position. By contrast, in the behaviorally irrelevant condition, the activity of neural units sampling the prime location were enhanced. We modeled the experimental conditions by adjusting the spatial position of a Gaussian-distributed neural signal originating from FEF. Microstimulation of
FEF has been shown to locally modulate the contrast sensitivity of cells in areas such as V4 (Reynolds and Chelazzi, 2004).

\[ G(x; c, \mu, \sigma^2) := \frac{c}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]  \hspace{1cm} (11.1)

Eq. 11.1 specifies the Gaussian distribution that was sampled to derive the spatial profiles of the neural signals from model areas FEF to MSTd. In the behaviorally relevant condition, we set \( \mu \) such that \( G \) was centered on the fixation point, since the monkey was trained to saccade to the radial center of the optic flow located 30° away. In the near and far conditions, we set \( \mu \) such that the Gaussian was centered 43% and 86% closer to the edge of the screen away from the FoE position. This simulates the 30° and 60° radial distance between FoE and prime locations used by Dubin and Duffy in their near and far conditions. We configured the Gaussian distribution to wrap around the population boundaries such that the area under the curve for all conditions remained constant. Across these three conditions, we fixed \( c = 6.5 \) and \( \sigma = 18.5° \).

In the experiments of Dubin and Duffy, the behaviorally irrelevant trials encompassed both near and far conditions. To simulate the behaviorally irrelevant condition, we averaged across all trials with different prime locations (near and far). To investigate whether or not the type of MSTd attentional modulation is important, we applied the spatial signal from FEF in three different ways: by multiplying or adding the signal to the input of MSTd, and by modulating the gain of MSTd sensory inputs. Adding the attentional signal to the sensory input that MSTd units receive can shift their receptive fields, which approximates the behavior of non-multiplicative attention (Womelsdorf et al., 2006). Multiplying the FEF signal with the sensory input MSTd units receive enhances or suppresses existing activity, which approximates the behavior of multiplicative attention (Martinez-Trujillo and Treue, 2004). To ensure our model tested a variety of attention types, we multiplied the model FEF signal to
the MSTd sensory input such to only augment the gain of existing neural activity. In other words, units influenced by attention in this case exhibit higher gain in their activations than those that are not.

11.2.1 The Model

As noted in the Introduction, the model contains stages that correspond to primate visual areas V1, MT, MSTd, and frontal eye fields (FEF). Model V1 projects to MT, which provides bottom-up inputs to MSTd. FEF provides top-down attentional signals to MSTd.

All simulations were run on an 8-core 2.66 Ghz Mac Pro with 64 GB of memory using Mathematica 8. Parameter values listed in the text specify those that remained constant throughout all simulations. Table 15.1 contains parameters values that varied for different experimental conditions.

Equations in our model describe the temporal dynamics of individual neurons or populations of neurons that densely sample the visual field. Model neurons obey ordinary differential equations that feature shunting competitive dynamics (Grossberg, 1968). These equations perform a leaky integration of their inputs and simulate many known properties of neurons, such as divisive normalization (Heeger, 1992; Carandini and Heeger, 2011) and automatic gain control (Grossberg, 1983). Model equations for area MSTd resemble the following membrane equation, termed a recurrent com-

### Table 11.1: Parameter values used in simulations.

<table>
<thead>
<tr>
<th>Attention Case</th>
<th>$\gamma_{MST}$</th>
<th>$\delta$</th>
<th>$n$</th>
<th>$\zeta$</th>
<th>$w_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-multiplicative</td>
<td>0.03</td>
<td>1</td>
<td>3</td>
<td>0.0001</td>
<td>0.15</td>
</tr>
<tr>
<td>Multiplicative</td>
<td>0.014</td>
<td>1</td>
<td>3</td>
<td>$4 \times 10^{-14}$</td>
<td>0</td>
</tr>
<tr>
<td>Multiplicative Gain</td>
<td>0.03</td>
<td>1</td>
<td>3</td>
<td>0.0005</td>
<td>0</td>
</tr>
<tr>
<td>Modified Sigmoid</td>
<td>0.03</td>
<td>1.2</td>
<td>6</td>
<td>$3 \times 10^{-9}$</td>
<td>0.08</td>
</tr>
<tr>
<td>Modified Sigmoid (Windowed)</td>
<td>0.03</td>
<td>1.1</td>
<td>6</td>
<td>$2 \times 10^{-9}$</td>
<td>0.065</td>
</tr>
<tr>
<td>Modified Sigmoid (Distance-dependent)</td>
<td>0.03</td>
<td>1</td>
<td>6</td>
<td>$4.5 \times 10^{-9}$</td>
<td>0.057</td>
</tr>
</tbody>
</table>
petitive field (Grossberg, 1973):

\[
\frac{dx_i}{dt} = -\alpha x_i + (\beta - x_i)(f(x_i) + I_i) - (x_i + \gamma) \sum_{k \neq i} f(x_k). \tag{11.2}
\]

Eq. 11.2 is a shunting equation that describes the activity, \( x \), of the \( i \)th cell in a neural network layer. The parameters \( \alpha, \beta, \gamma \) define the passive decay rate \( (sec^{-1}) \), saturation upper bound, and hyperpolarizing lower bound of the cell, respectively. 

The terms \( (\beta - x_i)(f(x_i) + I_i) \) and \( -(x_i + \gamma) \sum_{k \neq i} f(x_k) \) of Eq. 11.2 specify the shunting excitation by input \( I \) and surround inhibition, respectively. In Eq. 11.2, \( f(x) \) is a signal function (Grossberg, 1973) that specifies the nature of the feedback from cells in the same network layer. A sigmoidal signal function, Eq. 15.16, induces winner-take-all, pattern-preserving, and uniformizing behavior when the activation of units in the model falls in the faster-than-linear, linear, and slower-than-linear regions of the signal function, respectively. For a more comprehensive analysis of recurrent competitive fields, such as those defined in Eqs. 11.2 & 15.15 (Grossberg, 1973). The parameters \( \delta, w_0, \zeta, \) and \( n \) adjust the gain, threshold, slope and position of the linear portion, and slope of the sigmoid, respectively. In Eq. 15.16, \( [\cdot]^+ \) denotes the half-wave rectification, \( \max(\cdot, 0) \).

\[
f(w; \delta, w_0, \zeta, n) = \frac{\delta ([w - w_0]^+)^n}{\zeta + ([w - w_0]^+)^n} \tag{11.3}
\]

V1

In model V1, we analytically compute first-order optic flow field representations according to Eq. 11.4 (Longuet-Higgins and Prazdny, 1980), to create motion representations similar to those shown in Figure 11.2.

\[
I^1(x, y) := \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} x l_z - l_x \\ y l_z \end{pmatrix}. \tag{11.4}
\]
In Eq 11.4, \((\dot{x}, \dot{y})\) represent the horizontal and vertical flow components at the position \((x, y)\) at time \(t\), \(l_z\) signifies the depth component of the translational velocity of the observer \((m/sec)\), \(l_x\) indicates the horizontal component of the observer translation \((m/sec)\), \(Z\) is the distance \((m)\) from the observer to the point in space represented by the dot. We use the notation \(I^t(x, y)\) to represent the vector-valued optic flow field \((\dot{x}, \dot{y})\) with spatial location \((x, y)\) at time \(t\).

**Model MT**

Model V1 projects directly to Model MT, where cells have receptive fields that integrate over particular velocities (speed and direction). We define the pooled MT motion \(\tilde{M}(x, y)\) according to

\[
\frac{d\tilde{M}(x, y)}{dt} = -\alpha_{MT}\tilde{M}(x, y) + (I^t(x, y) \ast G_{MT})(x, y; \mu_{MT}, \Sigma_{MT}, r_{MT}) \quad (11.5)
\]

where \(\ast\) denotes the 2D convolution operator, \(\alpha_{MT}\) represents the passive decay rate \((sec^{-1})\) of each MT component unit, \(G_{MT}\) is a 2D discrete multivariate Gaussian kernel with mean \(\mu_{MT}\) and covariance matrix \(\Sigma_{MT}\) normalized such that all points in the kernel’s support sum to unity, and \(r_{MT}\) defines the kernel radius \((^{\circ})\). The points \((x, y)\) refer to positions in 2D retinotopic coordinates. For all simulations, we set \(\alpha_{MT} = 3\). We model MT cells with circular receptive fields, hence we set \(\mu = \tilde{0}\) and \(\Sigma_{MT}\) such that \(\sigma_x = \sigma_y = \sigma_{MT}\) and the covariance \(\rho\) between \(x\) and \(y\) is zero. We used \(\sigma_{MT} = 0.01^{\circ}\) and \(r_{MT} = 3^{\circ}\) to conservatively simulate MT cell receptive field properties found in neurophysiological studies (Born and Bradley, 2005; Churchland et al., 2005). Rather than integrating each component of Eq. 11.5, we used the analytical solution shown in Eq. 11.6 to evaluate each model MT cell at time \(t\).

\[
\tilde{M}^t(x, y) = \frac{I^t(x, y)}{\alpha} \ast G_{MT}(x, y)(1 - e^{-\alpha t}) \quad (11.6)
\]
Model MSTd

In model MSTd, we perform a template match between MT units $\tilde{M}^t(x, y)$ and all templates $\tilde{T}_i(x, y)$. That is, for a given $\tilde{M}^t(x, y)$, we match against $\tilde{T}_i(x, y)$ for all horizontal headings $i$. Templates are defined as normalized radial optic flow fields computed according to Eq. 11.4 (Layton et al., 2012a). We used 128 templates, each with a different FoE position uniformly sampled along the 90° middle horizontal axis. We weight the template match using inverse Euclidean distance. In Eq. 11.7, we obtain a scalar value $p^t_i$ for time $t$ at the horizontal heading $i$, representing the cosine similarity (i.e. inner product) between distance-weighted vectors at each time and those in the template.

$$p^t_i = \lambda \sum_{\{x,y\}} W_i(x, y) \left( \sum_{\{\hat{x}, \hat{y}\}} \frac{\tilde{T}_i(x, y) \odot \tilde{M}^t(x, y)}{||\tilde{M}^t(x, y)||} \right)$$  \hspace{1cm} (11.7)

In Eq 11.7, $W_i(x, y)$ represents a distance-dependent weighting from the horizontal heading indexed $i$. We use inverse 2D Euclidean distance, scaled by a parameter $\lambda$ to adjust the effective spatial extent of the templates. We selected $\lambda = 200$, which broadly scales the distance-dependent weights across the visual field. The inner summation performs component-wise multiplication (denoted by the $\odot$ operator) between vectors in the template $\tilde{T}_i(x, y)$ and on MT $\tilde{M}^t(x, y)$ for every spatial location. The resulting vector is normalized by the $L^2$ (Euclidean) norm (denoted $||\tilde{M}^t(x, y)||$) and then the vector components $\{\hat{x}, \hat{y}\}$ are summed.

We subsequently smoothed then sharpened the 1D pattern match distribution in MSTd according to

$$P^t_i := (p^t \ast G_{MST})^{nMST}(i; \mu_{MST}, \sigma_{MST}, \tau_{MST})$$  \hspace{1cm} (11.8)

where $\ast$ is 1D cyclic convolution and $G_{MST}$ is a normalized 1D Gaussian kernel. We
Figure 11.3: (a) The model mean MSTd population activation yields a linear ($R^2 = 1$) separation between the peak timings in the near (165 msec), behaviorally relevant (192 msec), and far (219 msec) conditions, respectively. In these respective conditions, the model simulates the increasing distance attention traveled between the fixation and FoE locations. (b) Summary of the population peak timing in the near, behaviorally relevant, and far conditions as a function of the type of attention modeled. Additive, multiplicative, and multiplicative gain attention types all produced a linear peak timing ($R^2 > 0.98$). Modifying the sigmoidal signal function in model MSTd, as described in the text, yielded the best timing correspondence to the Dubin and Duffy (2009) data (gray line).
set the radius $r_{MST}$ to 40°, $\sigma_{MST} = 10^\circ$, and mean of the MST kernel $G_{MST} \mu_{MST} = \bar{\sigma}$. We sharpened the resulting distribution with $n_{MST} = 30$.

Finally, we introduce a dynamical competitive network to describe each MSTd unit $B$ at location $i$ according to Eq. 15.15.

$$\frac{dB_i}{dt} = -\alpha_{MST} B_i + (\beta_{MST} - B_i)(f(B_i) + P'_i) - (\gamma_{MST} + B_i) \sum_{k \neq i} f(B_k). \quad (11.9)$$

In Eq. 15.15, $f(B_i)$ and $\sum_{k \neq i} f(B_k)$ represent excitatory and inhibitory recurrent inputs, respectively. The self-excitation term $f(B_i)$ can be considered as a form of post-synaptic input if one makes an assumption that the cell does not have a physical synaptic connection with itself. We set $\alpha_{MST} = 0.01 (sec^{-1})$ and $\beta_{MST} = 1$.

**Model FEF**

In the studies performed by Dubin and Duffy attention was maintained prior to the optic flow presentation. We modeled the FEF signal as decaying (Eq. 11.10) after the initial conditions were set according to Eq. 11.1 at $t = 0$. The parameters varied according to the experimental conditions defined above. We set $\alpha_{FEF} = 0.01 (sec^{-1})$.

$$\frac{dA_i}{dt} = -\alpha_{FEF} A_i. \quad (11.10)$$

Modulation of neuronal activity due to attention has been documented in visual areas fewer synapses away from the retina than MSTd, such as V1 (Paradiso, 2002). This article is focused on the response properties of MSTd as described by Dubin and Duffy, we therefore implemented attentional effects only as far as MSTd and investigated whether the effects of attentional signals on the MSTd population may be sufficient to explain the data.
Attention Cases

Exactly how attention acts in cortex is not clear, but researchers have proposed that it could either multiplicatively or non-multiplicatively affect neural signals. We tested several possible ways an attentional signal could interact with MSTd cells. First, we considered *additive attention* on MSTd cells by changing the excitatory (2nd) term in Eq. 15.15 to \((\beta_{MST} - B_i)(f(B_i) + P_i^t + A_i)\), where \(A_i\) is defined according to model FEF (Eq. 11.10). This modification provides an additive (non-multiplicative) influence on MSTd unit inputs (Womelsdorf et al., 2008). Second, we considered the effects of *multiplicative attention* on MSTd dynamics by modifying the excitatory term in Eq. 15.15 to \((\beta_{MST} - B_i)(f(B_i) + P_i^t \times A_i)\). In this case, MSTd activity is enhanced or suppressed but not induced (Martinez-Trujillo and Treue, 2004). We also examined the effects of *multiplicative gain* by scaling the sensory input \(P_i^t\) in Eq. 15.15 by the spatial pattern from FEF, which modulates the gain of model cell responses: \((\beta_{MST} - B_i)(f(B_i) + P_i^t \times (A_i + 1))\).

11.3 Results

MSTd Population Timing

In order to assess the model’s ability to fit the basic neuronal tuning curve and linear trend in the peak population temporal activity of Dubin and Duffy, we employ the sigmoidal signal function defined by Eq. 15.16 with \(n = 3\). Figure 11·3b summarizes the peak timing results as a function of attention. As noted, we simulated non-multiplicative (additive), multiplicative, and multiplicative gain attention conditions. All attention simulation results exhibit a linear peak timing trend similar to that found in the data of Dubin and Duffy (Figure 11·1). Figure 11·3a shows the mean MSTd population response in the additive attention case. The timing of the MSTd population activity peaks are at 165 msec, 192 msec, and 219 msec \((R^2 = 1)\).
corresponding to when the prime is at 0° eccentricity (near condition), the FoE is at 30° eccentricity (behaviorally relevant condition), and the prime is at 60° eccentricity (far condition). In this case, the peak latencies exhibit an exact linear trend. In the multiplicative attention case, we obtained MSTd population peak latencies of 63 msec, 72 msec, and 81 msec for the near, behaviorally relevant, and far conditions, respectively. In the multiplicative gain case, we obtained MSTd population peak latencies of 147 msec, 156 msec, and 165 msec for the near, behaviorally relevant, and far conditions, respectively. We obtained linear peak timings ($R^2 > 0.98$ in all conditions) and similar qualitative appearances irrespective of the attention condition (Figure 11·3b). For the remainder of this article we focus on the additive case.

When we modified the MSTd network signal function (‘Modified Sigmoid’) as specified in Table 15.1, we were able to better model the temporal dynamics of the MSTd neurons and obtain closer correspondence to the Dubin and Duffy data (Figure 11·3b). Changing the sigmoid exponent $n$ from 3 to 6 steepened the slope of the sigmoid. The implications of this change are that the activity of network units is less likely to enter the linear range of the sigmoid signal function because it is narrower, and units whose activity do enter the linear range are more likely to be ‘pushed’ out to the faster-than-linear or slower-than-linear segments. Figure 11·4a depicts the average population behavior over time with the new signal function. The timing of the MSTd population activity peaks are at 123 msec, 225 msec, and 321 msec corresponding to when the prime is at 0° eccentricity (near condition), the FoE is at 30° eccentricity (behaviorally relevant condition), and the prime is at 60° eccentricity (far condition). Similar to the data of Dubin and Duffy, the behaviorally irrelevant population response (blue) did not yield a dominant peak. The modified signal function better fits the data of Dubin and Duffy. We verified the network was making use of the full dynamic range of the sigmoidal signal function by comparing the network behavior to
Figure 11.4: (a) Model mean MSTd population activation using a steeper sigmoidal signal function compared to Figure 11.3a (see Table 15.1). The model produces linear peak timing in near (green), behaviorally relevant (red), and far (cyan) conditions while qualitatively simulating the low frequency deflections present in the Dubin and Duffy (2009) data (Figure 11.1). The behaviorally irrelevant condition (blue) does not yield a distinct peak, thereby qualitatively matching the data of Dubin and Duffy (2009). (b) Model mean population MSTd population activation of cells selective to the FoE position (windowed). The solid and dashed gray curves show the activation of units not selective to the FoE in the near and far conditions, respectively. (c) Model mean MSTd population activation in a network that features distance-dependent competition and windowing, which better reflect neurophysiological and experimental conditions. Consistent with the data of Dubin and Duffy (2009) (Figure 11.1), the curves for all conditions drop off more steeply after peaking, the peak in the behaviorally relevant case has higher contrast, and the behaviorally irrelevant condition response is flatter.
when we made $f(w)$ a step function with threshold $\Gamma = 0.25$ to emulate the shape of the sigmoid without the linear portion. The step signal function preserved the linear peak timings, but did not yield the multiple deflections seen in Figure 11-4a, which indicates that model MSTd units use the linear region. The modified signal function gives rise to a number of low frequency deflections in the population activity that resemble those present in the Dubin & Duffy data (Figure 11-1). Our simulations did not contain noise, so we investigated the source of these deflections.

Figure 11-5 presents snapshots of the spatial activity in the template matching layer for the near, behaviorally relevant, and far conditions. Each subplot shows 120 msec of network activity within model area MSTd. We used intervals of 120 msec to capture the dynamics around the low frequency deflections present in Figure 11-4. In each set of plots, we identified a group of MSTd units that respond primarily to the attentional signal (the *attentional subpopulation*) and another group primarily driven by the sensory bottom-up input (the *sensory subpopulation*). Note, the subpopulations are in the same model MSTd layer and are only classified as such based on their response to either the sensory optic flow or attentional signal. The results in Figure 11-5 indicate that the deflections arise due to competitive interactions between the attentional (e.g. Figure 11-5c, right peak) and sensory (e.g. Figure 11-5c, left peak) subpopulations. Both attentional and sensory inputs influence model neurons within each subpopulation, so we use the terms attentional and sensory subpopulations to refer to the primary distributions of activity within the network induced by the respective signals. In the near condition (Figure 11-5a), the first deflection, which happens to be the overall population peak activity, occurs due to the superposition of the MSTd response driven by the attentional signal (built up prior to the arrival of the sensory input) and the sensory input due to the optic flow (activity still ramping up at $\sim$120 msec) (Figure 11-5a, left panel). Due to the competition, the average
activity across the network subsequently reduces for the next ~50 msec. Finally, the MSTd response to the sensory input reaches its peak at ~225 msec and then reduces due to competition at later times (Figure 11·5a, center and right panels). In the behaviorally-relevant condition, covert attention must travel 30° relative to fixation. Unlike the near condition, the sensory and attentional signal superposition is weaker when the first deflection occurs at ~120 msec (Figure 11·5b, left panel). However, the superposition is stronger later, resulting in the second deflection at 225 msec when attention reinforces the bottom up response to the optic flow (Figure 11·5b, center panel). In the far condition, the prime is located at 60° eccentricity relative to the FoE. The first deflection corresponds to the high MSTd activity due to the attentional subpopulation (Figure 11·5c, left panel). Since the sensory response is still developing in MSTd and the distance is far from the prime, less superposition occurs and the network response is more evenly distributed across MSTd, resulting in a lower average population activity than the other experimental conditions. The second deflection in the far condition arises due to fierce competition between the sensory and attentional subpopulations (Figure 11·5c, center panel). The attentional subpopulation loses the competition due to the emerging sensory response, which results in a sudden dip in the average network activity at ~300 msec. Finally, the third deflection at 310 msec corresponds to the sensory subpopulation reaching its maximal activity and the rapid decay of the attentional subpopulation (Figure 11·5c, right panel). The sensory subpopulation response takes longer to develop, because of the strong competition.

**Windowing**

Dubin and Duffy included only neurons that showed significant responses to the experimental conditions in their analyses. We initially included all model cells. In order to better approximate electrophysiological conditions, we introduced windowing whereby we only included cells moderately or highly selective to the FoE of the optic
Figure 11-5: The spatial dynamics of the model MSTd heading template match layer. (a), (b), and (c) show the temporal evolution of MSTd cells in the near, behaviorally-relevant, and far conditions, respectively. Each subplot contains snapshots of the network activity at uniformly sampled times within 120 msec intervals. From left to right, the plots show snapshots from contiguous 120 msec intervals. The FoE is located in the center of the visual display (0°). We call the cells primarily influenced by the sensory optic flow (left peak) and attentional (right peak) signals the sensory subpopulation and attentional subpopulation, respectively.
flow display. This way, we do not include cells that would not elicit a response to the sensory optic flow in our averaging. This was the criterion employed by Dubin and Duffy to select neurons for analyses. It should be noted that this affords a better data fit, but does not qualitatively affect our results. The simulation results are shown in Figure 11·4b. Figure 11·4b shows that peak latencies were 105 msec, 219 msec, and 339 msec for the near, behaviorally relevant, and far conditions, respectively. The behaviorally irrelevant condition (blue) did not yield a clear peak. We plotted the activations of units excluded from the colored curve averages in the near and far conditions in solid and dashed gray, respectively. Similar to the Dubin and Duffy data, the gray curves resided below the colored curve conditions, and the solid gray curve showed higher average activation than that of the dashed gray curve.

**Distance-dependent competition**

Our model assumes all cells globally compete with one another with equal weight, despite the fact that they may have very different visuotopic preferred FoE locations. Since MST has a rough topography (Born and Bradley, 2005), we introduce a distance-dependent weighting in Eq. 15.15 such that units compete locally. Each MSTd unit now receives bottom-up input from a single visuotopic location but receives only local inhibitory input from neighboring units. The extent of the local competition in MSTd is determined by the inhibitory kernel $G_{MST}$. As shown in Figure 11·4c, the distance-dependent network results combined with windowing further improves the qualitative fit with the Dubin and Duffy data. The smaller peak in the near condition appears after the second deflection in the far condition, the contrast between the first two behaviorally relevant peaks is higher, and the mean MSTd population activity drops off faster after the peaks occur in all conditions. Peak latencies were 99 msec, 195 msec,
and 291 msec for the near, behaviorally relevant, and far conditions, respectively.

\[ \frac{dB_i}{dt} = -\alpha_{MST} B_i + (\beta_{MST} - B_i) (f(B_i) + P_i) \]
\[ - (\gamma_{MST} + B_i) \sum_{k \neq i} \sum_l G_{MST}(k - l, \sigma^2) f(B_l). \quad (11.11) \]

Eq. 11.11 shows the modified MSTd equation to implement for the distance-dependent interactions via \( G_{MST}(\mu_{MST}, \sigma^2_{MST}) \), with \( c_{MST} = 4, \sigma = 60^\circ \). Simulation results are shown in Figure 11.4c. Compared to the results shown in Figure 11.4a – b, the attentional and sensory subpopulations are composed of fewer units due to the local inhibition, which results in larger drops in average network activity when the sensory representation wins over the attentional representation in the competition.

**MSTd Neuronal Tuning Curves**

Figure 11.6 shows the tuning curves of model neurons selective to similar optic flow FoE in the near vs. far (a) and behaviorally relevant vs. behaviorally irrelevant (b) conditions. When presenting an optic flow display at MSTd units’ preferred FoE location, units exhibited a higher gain in the near and behaviorally relevant conditions compared to far and behaviorally irrelevant conditions, respectively. When the units were not tuned to the FoE position, units in these respective conditions showed suppression. The tuning curves derived in Figure 11.6 reflect the experimental findings of Dubin and Duffy (Dubin and Duffy, 2007).

**11.4 Discussion**

We have presented a dynamical model of primate MSTd that simulates the data of Dubin and Duffy. Our model produced peaks in the average population activity of MSTd units with timings spaced linearly as the distance increased between the center of the spatial attention signal and the FoE. The linear trend in the peak timings
Figure 11.6: (a) Behaviorally Relevant (red) vs. Behaviorally Irrelevant (blue) attention tuning curves. Units selective to radial center of motion exhibited higher grain in the behaviorally relevant condition compared to the behaviorally irrelevant condition. Optic flow displays with FoE far from the preferred location elicited a suppressed response in units. (b) Near (green) vs. far (cyan) attention tuning curves. Units selective to radial center of motion exhibited higher grain in the near condition compared to the far condition. Optic flow displays with FoE far from the preferred location elicited a suppressed response in units. The tuning curves were derived from ten model cells.
was robust to a range of parameters, including the steepness of the sigmoidal signal function slope and signal function type (e.g. faster-than-linear and step functions). Our quantitative fits of the data of Dubin and Duffy improved when we excluded units from the population average that were not strongly selective for the optic flow FoE and introduced distance-dependent competition into the network. Both modifications more accurately simulate the electrophysiological and neurophysiological conditions of Dubin and Duffy. Our neuronal tuning curves exhibited increased gain about the neuron’s preferred FoE when comparing the curves generated from the aggregate behaviorally relevant vs. behaviorally irrelevant conditions and the near vs. far conditions. These findings also match the data of Dubin and Duffy. Finally, we simulated multiplicative and non-multiplicative types of attention acting on model area MSTd, and our model produced qualitatively similar results irrespective of the particular type of attention. Hence, our model is agnostic with respect to, and compatible with all of these forms of attention. Our MSTd network equations feature built-in normalization (Eq. 15.15), and are also compatible with proposals of attention that incorporate a normalization property (Reynolds and Heeger, 2009; Carandini and Heeger, 2011).

The primary focus in selecting our model structure and parameters (Table 15.1) was to use the minimal possible mechanisms to fit the linear peak shift latencies. We tested three different attentional mechanisms by adding as few parameters the model as possible, while still capturing the expected behavior of each attention type. Relative curve separations in the data (Figure 11·1) and our model (Figure 11·4) agree well quantitatively at different temporal milestones. Early in the ‘Windowed’ and ‘Distance-dependent’ simulations at ~ 100 msec, the ratio between the green curve peak height and the heights of the cyan, purple, red curves underneath are close to those in the data. Late in all the simulations, the ratio between heights
of the cyan curve peak and the green curve is also close to the data. However, in some cases the ordinal heights of the curves in simulations differed from that of the data. For example, under the cyan curve peak in the data, the green curve was the next highest, followed by the red and purple curves. In the ‘Distant-dependent simulations’, the secondary green curve peak that occurs after 200 msec decreases faster than it does in the data and therefore it is the least active at the time of the cyan curve peak. Conversely, the behaviorally irrelevant condition curve decreases more slowly than it does in the data and therefore is higher than the red curve. There are several reasons why this may occur. First, an exhaustive search through the model parameter space has not been performed, so parameters that improve fits to the relative separations between experimental curves may exist. Second, the model fits the average neural data from Dubin & Duffy and does not consider the variance. Third, many known properties of MSTd cells are not modeled, such as differential receptive field sizes (Duffy and Wurtz, 1991b), dynamic ranges, and speed tunings (Duffy and Wurtz, 1997). Our goal was identify the simplest core neural mechanisms that are required to give rise to the data of Dubin and Duffy.

The exact implementation of attention (additive, multiplicative, multiplicative gain) in the model did not affect the linear separation in the peak latencies, but the fact that the model attentional signal was concentrated in particular spatial locations and assumed a Gaussian-like form is crucial for the model’s ability to fit the data. In particular, the attentional signal modulated the recurrent inputs in Eq. 15.15, which resulted in higher firing rate amplification closer the FoE, despite the attentional signal’s exact form, and therefore decreased the mean MSTd population activity peak latency when the attentional signal acted more proximally to the FoE location. Note that if the recurrent competition is not included in the model, the population peak activity will always occur at the same time, defined by the temporal, not the spatial,
separation of the attentional and sensory signals. Without the recurrent competition components, Eq.9 becomes a leaky integrator and the absolute magnitude of the input signals is the same in each experimental condition. With windowing, the effect of additional stimulation from the attentional signal on the FoE sensitive population varies depending on the spatial position of the attentional prime, however, unless a subset of MSTd cells saturate, these differences will have only minimal effects on the timing of the peak population response. Recurrent competition prolongs the time that the attentional signal produces a high activation response across the cell population after the attentional prime is removed. This latent activation remains until there is a competing sensory signal. When the latent activation peak is close to the sensory activation peak the two populations merge to produce an overall population peak soon after the presentation of the sensory signal. When the latent activation peak is far from the sensory activation peak, the two populations compete. Initially the latent activation is able to suppress the sensory activation, but as the competition continues the sensory activation overtakes and suppresses the latent activation. In this case, the overall population peak occurs when uncertainty as to the winner is at its highest, in other words when the latent activation and the sensory activation are roughly equal. We predict that attentional signals from FEF act on target MSTd neurons that possess a FoE preference near the spatially primed location of the visual field and receive the most modulation. Neurons with distal FoE preferences are predicted to successively receive less modulation as a function of visuotopic distance. Our model would require revision if a future experiment demonstrated that saccade-related attentional signals from FEF do not mostly target MSTd neurons with FoE preferences spatially coincident with the primed location or if the signal targets MSTd neurons in an erratic rather than a visuotopic, Gaussian-like fashion. In our model simulations, we assumed that attention modulates neurons in MSTd prior to sensory
signals. This assumption is based on a recent study of saccadic planning, which reports that saccade-related neural activity arrives 30 msec earlier in MSTd than would otherwise occur without saccades during fixation (Crowder et al., 2009). Future studies that probe the difference in latencies between MSTd afferents from MT and FEF could quantitatively test the prediction made by the model that attentional signals arrives prior to sensory signals, perhaps by simultaneously recording from neurons in MSTd and FEF. Experiments could also investigate the duration of FEF modulation on MSTd neurons when monkeys are presented with optic flow to test whether the signal influence decreases over time as predicted by the model.

As noted, the recurrent inputs to MSTd units in Eq. 15.15 played an important role in fitting the data. Recurrent excitation and inhibition in model MSTd is modulated by a sigmoidal signal function \(f(w)\) in Eq. 15.15). While we were able to obtain evenly spaced peak separations, akin to those shown in Figure 11-3, with a faster-than-linear signal function, such as \(f(x) = x^2\), we obtained the best performance using a sigmoid function. Depending on a unit’s activity relative to the spatial pattern of activity in the network, sigmoid signal functions afford analog winner-take-all behavior in the network. Units whose activity falls in the slower-than-linear portion of the sigmoidal signal function apply pressure on the rest of the population to suppress lower activity maxima, which in turn further enhances the global activity maximum. Unlike prior analysis of recurrent competitive fields with sigmoid signal functions (Grossberg, 1973), our MSTd simulations have dynamic, continuously varying inputs, which directly impact the recurrent feedback to MSTd units. Due to the continuously changing inputs and the use of a sigmoid signal function, strong bottom-up inputs can override stable network patterns, such as in the case when the attention signal is the main MSTd input, and apply pressure on the MSTd population to shift the location of the globally most active unit. This interplay between
bottom-up, top-down, and recurrent inputs gives rise to the activity patterns shown in Figure 11-5.

Recurrent competition, FoE selective cells in MSTd, and temporal competitive dynamics represent essential characteristics of the model. We know of no other model that can fit the data of Dubin and Duffy. Balancing rising MT unit activation and decaying attentional signals from FEF in the model allows competition within MSTd to produce the distinct activity peaks observed in Figure 11-5, which correspond to temporal landmarks that indicate that a particular signal is winning the competition. The dynamical properties of the MT and FEF signals represent important aspects of the model that allows the mean population activity in MSTd to peak at the appropriate times. A prior version of the model cannot simulate this because it employs difference equations (Layton et al., 2012a). Other models that lack temporal integration (Royden, 1997; Raudies et al., 2011), that are filter-based (Perrone, 1992) or that lack continuous-time dynamics (Browning, 2012), lack the necessary mechanisms to balance bottom-up and top-down signals and are therefore unlikely to be able to simulate the timing of peaks in the Dubin and Duffy data. Because our model does not include spiking, synaptic and conduction delays, temporal dynamics at the neuronal level may differ compared to those implemented in the model. Further experiments that show finer grain timing details about how attentional signals interact with MSTd neurons may require the model to be modified.

Our quantitative fits to each condition in the data of Dubin and Duffy were achieved by the increased distance between the attentional and sensory signals defined in the experimental paradigm. All model parameters were fixed, except where explicitly noted in the text. The differences in peak timings across the behaviorally relevant, near, and far experimental conditions arise in the simulation due to the competitive interactions between attentional and sensory representations in MSTd.
Although the low frequency deflections in the data of Dubin and Duffy may appear to reflect random fluctuations (Dubin and Duffy, 2009), we simulated these properties of the MSTd population response without any noise (Figure 11·4). We therefore predict that the low frequency deflections present in the neurophysiological data may represent dynamic shifts between the sensory and attentional subpopulations in MSTd. Figure 11·5 shows how the increased distance between the attentional and sensory signals result in fiercer competitive interactions and oscillations between the MSTd subpopulations. The temporal integration of the recurrent competition and sigmoidal signal function were critical to achieving these results.

If the hypothesis tested by the model that the low frequency deflections in the neurophysiological data actually represent important events in the processing of optic flow and attention is correct, then the number and latency of subpeaks, should remain the same for each experimental condition. The fact that each data curve (Figure 11·1) is the result of averaging a large population of neurons over many trials suggests this may be the case. However, if the undulation properties vary over more trials, as would be the case if the undulations are just noise, our hypothesis would be incorrect. In our ‘Windowed’ simulation, we discovered that increasing the window size had the effect of compressing the temporal spacing of and at times merging the average population peaks together. We therefore predict that characteristics of the low frequency deflections, such as number of subpeaks, change when the neuron population size over which one averages changes, but the deflection properties should remain fixed for a constant sample size.

Although our model successfully reproduces many results of Dubin and Duffy, the model cannot simulate one of the employed paradigms. In Experiment 2 of Dubin and Duffy, monkeys observed behaviorally relevant and irrelevant trials much like those described the present paper (Dubin and Duffy, 2007). However, in behaviorally
irrelevant trials, monkeys were presented with one of four shapes rather than a strictly spatial prime prior to the onset of the optic flow. The monkeys were trained to saccade to its location when a grid of all four shapes appeared. The shape was flashed at the fixation point prior to optic flow onset and would reliably appear in a fixed location different from the fixation point in the grid of four shapes after the optic flow was presented. Therefore, the monkey could learn a specific shape-to-space mapping. In Experiment 3, Dubin & Duffy randomized the location of the shape in the grid to preclude the monkeys from determining the saccade location prior to the shape grid appearance and thereby preventing a space-to-space mapping from forming. In both cases, the neuronal tuning curves qualitatively match those shown in Figure 11-6: behaviorally relevant and near trials elicited greater average firing in MSTd neurons than in behaviorally irrelevant and far trials, respectively, when the optic flow FoE was nearby the preferred FoE location. Our model cannot simulate the shape paradigm because it does not implement the learning factors required by the monkeys to perform the task. Due to the qualitative similarity between the neuronal tuning curves in the strictly spatial and shape priming paradigms, however, the underlying attentional signal and MSTd dynamics may be similar. If the model was updated to include a learning mechanism to associate a shape with a particular spatial location, we believe our model would simulate the results from Experiments 2 and 3.

As stated previously, we obtained qualitatively similar results for all types of attention when acting within a realistic parameterization of the physiological timing. However, with multiplicative attention, when we employed a different parameterization and slowed down the growth of each unit’s integration of its inputs, the relative timing between peaks increased exponentially. This result is because multiplicative modulation of the bottom up signal gives rise to exponential amplification over time. If input amplitude were logarithmically transformed, the MSTd population peak tim-
ings would again be linear. A logarithmic scaling of MSTd inputs is consistent with cortical magnification factor (Elder et al., 2009). If attention acts multiplicatively in cortex and the cortical magnification factor logarithmically transforms the sensory signal, then our analysis would be compatible.

Our results suggest that when attention is engaged during visually-guided navigation, recurrent competition in primate area MSTd modulates the time the population will take to reach its highest activity, and individual neurons tuned to the optic flow FoE, when attended, exhibit higher firing rates. The model predicts that saccade planning and attention modulate the temporal behavior of MSTd neurons, which may affect decision making when primates navigate in the environment. More work needs to be done to understand how primates engage attention in more ecologically relevant scenes, and with independently moving objects. Evidence exists that heading is useful for steering and as such, our model predicts that under divided attention the competition in MSTd will take longer and the time required to make confident steering decisions will also be longer. Recent neurophysiological evidence exists that steering according to an independently moving object or by surrounding optic flow alters the responses of MSTd neurons (Kishore et al., 2011). These findings are consistent with the model results showing that attention to different aspects of the environment changes the response properties of MSTd neurons.
Chapter 12

A neural theory of path perception

Self-motion, steering, and obstacle avoidance during navigation in the real world require humans to travel along curved paths (Rushton et al., 1998; Wann, 2000; Wann and Swapp, 2000; Fajen and Warren, 2003). The neural representation of self-motion along curved trajectories may be more fundamental than for those that are straight. Li and Cheng investigated human path perception along circular paths using random dot and textured ground plane visual displays (Li and Cheng, 2011b). Subjects fixated the center of the screen and were presented with displays that simulated movement along circular paths. At the end of each 1 sec trial, subjects were asked to place a marker a fixed distance away on their future path that would be reached if the trial were extended. Experiments varied the simulated direction of observer gaze along the circular path. Although observers did not move their eyes from the center of the screen during each trial, Li and Cheng constructed the displays to show observers the optic flow they would observe while fixating different targets. Gaze conditions included tracking a point 15° inside, on, or 15° outside the future path, and the natural case of gazing along the instantaneous heading, tangent to the path. The optic flow the observers viewed contained a combination of translational and rotational optic flow due to the intrinsic path curvature. Li and Cheng found that when subjects gazed inside (outside) the circle, path errors indicated that subjects overestimated (underestimated) the path curvature. Subjects also underestimated the path curvature when looking at a point on the future path. Subjects did not
produce significant errors when gazing along instantaneous heading. Another gaze condition was considered wherein the simulated head orientation counter-rotated as the subject traversed the circle such that gaze remained on the horizon (gaze-along-z-axis), and subjects saw a radially expanding optic flow field with a FoE that drifted throughout the trial. Subjects yielded errors consistent with the percept of a straight path. The visual displays composition (dots or textured surfaces) did not impact subject performance (Li and Cheng, 2011b).

Humans travel along curvilinear paths when steering toward goals. We define curvilinear path here as one that has zero (straight) or nonzero curvature, which are assumed to be circular. An analysis performed by Wann and Swapp shows that when humans fixate a goal that consists of two posts through which the observer will pass, whether the observer under- or over-steers can be determined by the retinal optic flow field over time (Wann, 2000). If the subject travels on an interception course with the goal, there is a sequence of vertical flow lines that lead to the goal. If the observer under- (over-) steers such that the trajectory will pass behind (in front of) the goal, the streamlines curve toward (away from) the target. The visual displays considered by Li and Cheng and Wann and Swapp both produce spiral patterns of optic flow.

Many perceptual models have been proposed that focus on heading (van den Berg, 1992; Warren, 1998; Warren et al., 2001; Britten, 2008), which specifies the direction of travel along straight paths, but not on path curvature, which humans accurately perceive and is critical to everyday locomotion (Saunders, 2010; Li and Cheng, 2011b; Cheng and Li, 2012). In primates, including humans, dorsal medial superior temporal area (MSTd) has been implicated in heading perception (Duffy and Wurtz, 1991a; Duffy and Wurtz, 1991b; Duffy and Wurtz, 1995; Bradley et al., 1996). However, the majority of MSTd neurons respond optimally to spiral patterns, rather than the radial expansion patterns associated with heading perception (Graziano et al., 1994). No
existing theory of curved path perception explains the neural mechanisms by which humans accurately assess path and no functional role for spiral-tuned cells has yet been proposed. Here we present a computational model that demonstrates how the continuum of observed cells (radial to circular) in MSTd can simultaneously code curvature and heading across the neural population.

I hypothesize that neurons sensitive to spiral motion in MSTd are important for curvilinear path perception. The circular paths in the simulated gaze conditions of Li and Cheng yield characteristic spiral optic flow patterns. We present a model in the next section in which spiral flow selective units in MSTd afford sensitivity to the path curvature. Competition between cells that are sensitive to a continuum of spiral, translational, and rotational optic flow fields produce peak activations biased in the same directions as human path errors (Li and Cheng, 2011b). Detection of path curvature by spiral selective MSTd units could indicate whether an observer is on course, understeering, or oversteering toward a goal. The center of motion is a generalization of the FoE when optic flow contains rotation, such as in spiral fields. Many neurons in MSTd have been shown to exhibit a property known as position invariance (Graziano et al., 1994), whereby the neurons respond to their preferred spiral pattern of motion irrespective of the center of motion (Duffy and Wurtz, 1991b; Duffy and Wurtz, 1995). This property may suggest that the global spiral pattern and curvature are important properties for primate navigation along curvilinear paths.

Curvature is encoded in the model through the spirality of the most active cell, and heading is encoded through the visuotopic location of the center of the most active cell’s receptive field. Curvature and heading errors made by the model fit those made by humans (Li and Cheng, 2011b). Our model challenges the view that the function of MSTd is heading estimation, based on our analysis we claim that it is primarily
concerned with trajectory estimation and the simultaneous representation of both curvature and heading. We offer the testable prediction that the optimal stimulus for MSTd cells reflects motion on a curved path rather than the pure spirals that have been hitherto studied (Graziano et al., 1994). In our model, temporal dynamics afford time-history in the neural representation of optic flow, which may modulate its structure. This has far-reaching implications for the interpretation of studies that assume that optic flow is, and should be, represented as an instantaneous vector field (Froehler and Duffy, 2002; Li and Cheng, 2011b). Our results suggest that spiral motion patterns that emerge in spatio-temporal optic flow are essential for guiding self-motion along complex trajectories, and that cells in MSTd are specifically tuned to extract complex trajectory estimation from flow.
Chapter 13

A unified model of heading and path perception in primate MSTd

13.1 Introduction

Gibson noted that animals can navigate about their environment using the changing pattern of light distributions falling on the retina, which is now known as optic flow (Gibson, 1979). Travel parallel to a ground surface, along a straight path, without eye movements or body rotations produces characteristic patterns of optic flow, which Gibson called a “melon-shaped family of curves”. These flow patterns contain a singularity known as the focus of expansion (FoE). When the path of travel is straight, the optic flow field radially expands and the FoE specifies the direction in which the animal is going (heading). Gibson observed that animals could navigate using optic flow by aligning the FoE with the direction in which the animal wishes to travel.

Since Gibson proposed this strategy for navigation, much psychophysical research has focused on understanding human perception of heading (Warren, 1998). For the remainder of this article, we define heading to refer to the instantaneous direction of travel of an observer, and we define curvilinear path as the trajectory of travel, which may be curved. Psychophysical studies of human heading judgments have largely been based on static environments, consisting of dot or textured ground planes or 3D dot clouds. The observer typically travels along a straight path. In such environments,
humans accurately judge their heading, within 1° (Warren et al., 1988; van den Berg, 1992; van den Berg, 1996). Biologically inspired models of human heading perception often make use of depth variations in the visual scene (motion parallax) (Rieger and Lawton, 1985; Hildreth, 1992) to estimate the observer’s heading given a two dimensional (2D) retinal velocity field. Template matching (Warren and Saunders, 1995; Royden, 2002), whereby the retinal optic flow is compared to a number of radial templates, or a combination of the approaches is also used. In more complex environments, human heading judgments have been shown to be less accurate. When the environment contains independently moving objects, humans demonstrate systematic bias in their heading judgments depending on the motion of the object. An independently moving object traveling perpendicularly to the observer’s path gets closer to the observer over time and human heading judgements are biased opposite the direction of the object motion (Warren and Saunders, 1995). By contrast, a moving object that maintains its distance from the observer over time results in biases in human heading judgements in the direction of object motion (Royden and Hildreth, 1996). In general, models of the heading perception in primate cortex explain the bias in human judgments by either relying on object segmentation (Royden, 2002) or pooling motion over large parts of the visual field (Layton et al., 2012a).

Navigation under natural conditions is more complex than traveling on a straight path without any rotation. When the observer travels along a straight path, factors, such as eye movements and gaze, introduce rotation, which may result in optic flow that is not radially expanding or contracting. Sources of rotation are either considered retinal or extra-retinal (Warren and Hannon, 1990). Rotations that occur through the actions of the observer, such as eye, head, or body movements, which also result in vestibular signals, are considered extra-retinal, whereas rotations due to path curvature are considered retinal. Research on heading perception during
smooth-pursuit eye movements shows that human bias in heading judgments remains constant (< ±5°) and independent of angular rotation due to eye movement velocities (Royden et al., 1994; Royden et al., 1992). However, when human subjects fixate (no extra-retinal rotation) and are shown optic flow displays that simulate what would be seen by an observer traveling along a straight path with a constant amount of rotation, humans make large errors in heading judgments that is proportional to the rate of rotation (Royden et al., 1994). This is often referred to as the simulated rotation condition, in which the retinal rotation experienced by human observers is due to the simulated eye movements. Subjects typically note the experience of traveling along a curved path and not a straight path with eye rotation, which is the assumption of the experimenters. Mathematical analyses indicate that the optic flow experienced by subjects when eye movements are simulated is similar to that experienced traveling on a curved path over the time period of a typical experimental trial (Royden, 1994). For longer viewing times, the optic flow in the two scenarios diverges and could potentially allow the subjects to disambiguate curved paths from simulated eye rotations.

Electrophysiological evidence suggests that radial expansion sensitive neurons in MSTd, which are thought to encode heading, demonstrate modulation due to eye vestibular signals (Bradley et al., 1996; Shenoy et al., 1999; Shenoy, 2002). The modest human bias demonstrated by humans during eye-movements is less than would be expected given the magnitude of the rotations. Assuming MSTd is involved in heading perception, this could be explained by the extra-retinal signals to MSTd imperfectly ‘canceling out’ the rotations. Computational models have employed gain fields in MSTd as the mechanism by which this ‘canceling out’ of rotation may occur (Churchland et al., 2005; Elder et al., 2009).

Animals navigate complex terrain and their paths are rarely straight (Rushton
et al., 1998; Fajen and Warren, 2007), but few studies have examined human navigation along curved paths. Those that tend to examine human path perception in the context of circular paths (Fajen and Kim, 2002; Saunders, 2010; Cheng and Li, 2012; Li and Cheng, 2011a; Warren et al., 1991; Saunders and Ma, 2011), in the present article, if the curvilinear path is not straight, we also make the assumption of a circular path. When traveling along a curved path without eye or body movements, all rotation in the retinal optic flow is due to the path curvature. Research indicates that in environments composed of random dots, humans can accurately judge the curvature of their path in static environments (Warren et al., 1991; Kim et al., 2000; Kim, 2008). Judgments remain accurate in the presence of independently moving objects (Fajen and Kim, 2002), when the observer gaze or instantaneous heading direction and body orientation are always tangent to the path of travel. This naturally occurs during locomotion. Human judgements of path curvature are not affected by whether the environment is composed of sparse dots, limited lifetime dots, or dense textures (Li and Cheng, 2011b). However, many studies that investigate curvilinear navigation are confounded by whether subjects report heading or future path (Li and Warren Jr., 2004).

### 13.1.1 Theories of Path Perception

Existing theories of path perception are heuristics that do not specify the mechanisms by which the future path is perceived. Some theories depend on the active tracking of ‘features’ in the visual scene (Lee and Lishman, 1977; Warren et al., 1991), while others implicate an extensive cognitive component, such as updating path estimates with respect to external reference objects (Li and Warren, 2000; Li and Cheng, 2011b). We first summarize theories of path perception that assume active track of certain visual ‘features’ in the optic flow field. The passing flow line hypothesis observes that optic flow integrated over an extended period of time yields
Figure 13-1: Exemplar first-order optic flow fields. (a) Radially expanding optic flow experienced by an observer traveling along a straight path on a ground plane. The optic flow contains the focus of expansion (FoE) singularity on the horizon, which indicates the heading direction. (b) Movement of an observer along a straight path, as in (a), but with a constant amount of rotation added to the first-order optic flow (simulated rotation condition). Human subjects that view displays with simulated rotation report traveling along a circular path. (c) First-order optic flow experienced by an observer traveling on a circular path whose gaze is along heading or tangent to the circular path.

a streamline that passes underneath the observer that coincides with the path of travel (Lee and Lishman, 1977). This hypothesis assumes the observer gaze is in the direction of heading and requires the environment to have texture that passes directly underneath the observer. A related hypothesis proposed by (Wann and Swapp, 2000), which we call the vertical vector hypothesis, notes that if the observer maintains gaze on the destination of travel, the path can be recovered from retinal flow by integrating first-order flow vectors that are vertically aligned. This strategy does not require knowledge of heading. The vertical flow line hypothesis posits that the visual system tracks the constellation of vertical optic flow streamlines that exist when the observer fixates a point on the future path. This strategy assumes that humans fixate on their destination while traveling along curvilinear paths. The reversal boundary
hypothesis notes that the future path of travel coincides with direction reversals or “zero-crossings” in the horizontal motion component once the optic flow has been projected onto the retina (Warren et al., 1991); the horizontal motion component of texture inside (outside) the path will be rightward (leftward), or vice versa depending on whether the circle is traversed clockwise (CW) or counterclockwise (CCW). This hypothesis requires gaze to always be in the heading direction. While psychophysical evidence suggests that humans are most accurate in judging path curvature when the gaze direction is aligned with the instantaneous heading, it is not clear how the strategy may be used with momentary fluctuations in gaze. Warren and colleagues have proposed a vector normal hypothesis whereby the center of the circular path can be determined by computing the intersection of the vector normals of two points in the environment (Warren et al., 1991). Using the vector normals, knowledge of the circular path center, and the observer’s current position, the radius and therefore the curvature of the path of travel can be recovered. This strategy also assumes that the observer gaze is in the heading direction.

The hypotheses reviewed above suffer from rigid constraints about the environment or observer gaze, and are unlikely to represent general theories of human path perception. The strategies proposed by the passing flow line, vector normal, and vertical vector hypotheses only hold when observers look where they are going—i.e. gaze is along the heading direction. Judgments indicate that path curvature is perceived by humans when gaze is not along the heading direction (Saunders, 2010), which these hypotheses cannot explain. Humans can also perceive their path of travel in sparse environments composed of small quantities of dots. The boundary reversal hypothesis, however, requires dense optic flow to ascertain the horizontal motion zero-crossing. From a neural computation point of view, it is unclear how the brain could track the context-specific local features proposed by the above hypotheses over
time.

The following path estimation theories rely on external references in the environment. The *reference object hypothesis* posits that observers either update their self-position or integrate the change in heading over time with respect to an external object reference (Li and Warren, 2000). Subjects in the experiments of (Li and Cheng, 2011b) were able to judge their future path of travel in the absence of persistent objects in the environment, rendering the reference object hypothesis an incomplete strategy (Li and Cheng, 2011b). (Li and Cheng, 2011b) tested whether humans can integrate the change in heading without a reference object by tracking the “drift” in the FoE over time with gaze remained fixed along a particular axis when the observer travels along the circle without any rotation (Z-axis condition, Figure 15.6). Subject responses were consistent with the percept of moving along a straight path, not a circle, making the FoE drift hypothesis unlikely (Li and Cheng, 2011b). Finally, (Li and Cheng, 2011b) proposed that observers first estimate heading as a reference to then estimate the path curvature, which is mathematically defined in the circular path case as the ratio of the optic flow rotation to translation. It is not clear if or how mechanisms in the brain do or could perform these operations.

In summary, theories of path perception either treat path perception as independent of heading or depend on the prior determination of heading. In the present article, we propose a neural model of the primate visual system in which representations of heading and path are determined simultaneously and dynamically interact in the same population of neurons.

13.1.2 Neurophysiology of Path Perception

Neurons in the primate medial superior temporal area (MST) in the superior temporal sulcus (STS) exhibit tuning to radially expanding optic flow patterns, similar to those experienced by an observer moving on a straight path, and therefore have been
the focus of neurophysiological investigations of the mechanisms underlying visually-guided navigation. MST is the earliest visual area, the fewest synapses away from the retina in the primate dorsal stream, that responds to large field pattern motion. Evidence suggests that MST in monkey is composed of functionally distinct dorsal (MSTd) and ventral (MSTv) regions. Whereas neurons in MSTd exhibit sensitivity to optic flow patterns that occupy areas of the visual field as large as 100°, MSTv neurons have smaller receptive field sizes and are suspected to be involved in the perception of object motion (Orban, 2008). When stimulated in the laboratory, neurons in MSTd exhibit selectivity to large constellations of dots that form radially expanding and contracting patterns (Duffy and Wurtz, 1991a), resembling those experienced by an observer traveling along a straight path. MSTd neurons demonstrate sensitivity to dot speed (Duffy and Wurtz, 1997) and spatial shifts in FoE position (Duffy and Wurtz, 1995), and therefore are thought to be involved in heading perception (Britten, 2008).
Figure 13-2: Experimental paradigm and sample optic flow fields from (Froehler and Duffy, 2002), who report the existence of path selective neurons in MSTd. A monkey seated in a sled traveled CCW (a) or CW (c) along a circular track while maintaining gaze on the distal wall of luminous dots. The body, head, and eye did not rotate so that the monkey always directly faced the distal wall. The monkey therefore experienced radially expanding or contracting optic flow without sources of rotation. (b,d) Instantaneous optic flow experienced by the monkey at different locations along the circular track. In (b) at $t_0$, the monkey views a radially expanding optic flow while moving CCW when the heading direction is straight ahead, which is the same as the optic flow viewed CW 180° on the other side of the circle. Between $t_0$ and $t_1$, the FoE drifts rightward until at $t_1$ it is out of view. At $t_2$, the monkey experiences radial contraction.
(Froehler and Duffy, 2002) have conducted the only neurophysiological study to date that reports the existence of “path selective” neurons in cortex (Froehler and Duffy, 2002). Monkeys were placed on a sled in a dark room that contained bright dots on the three walls that were within view. The sled moved CW or CCW along a circular path (Figure 15.2). The sled was configured not to rotate the body as it traversed the circular path. The monkeys maintained gaze, throughout the trial, on a target that was projected from the sled onto the distal wall. Because the projector was attached to the sled and the monkey was trained to maintain gaze on the target, the fixation point occupied the same position within the monkeys’ visual field over time. The optic flow experienced by the monkeys contained no sources of rotation and appeared to radially expand or contract at each instant, with a FoE or focus of contraction (FoC) that ‘drifted’ horizontally during the trial. A monkey traveling once around the circle on the sled therefore viewed a sequence of instantaneous headings and each had an equivalent at antipodal positions in both the CW and CCW trials. (Froehler and Duffy, 2002) recorded from single neurons in MSTd and 73% elicited differential activity at antipodal positions on the track, where expansion/contraction optic flow patterns were identical. The neurons’ response depended on whether the circle was traversed CW or CCW, and as a result the authors claimed these cells demonstrated path selectivity. The authors also found heading selective cells, which fired when the optic flow contained their preferred heading irrespective of the CW or CCW traversal direction, and place selective cells, which responded when the monkey moved to a particular location of the room irrespective of the visual motion pattern. The selectivity of neurons in the sample was distributed along a continuum, ranging between demonstrating high (path selective) to low (heading selective) CW v.s. CCW differential activity. The mechanisms that underlie how these cells in MSTd respond to along a continuum to heading and path were not evaluated by the study.
In summary, neurons in MSTd have been shown to selectively respond to optic patterns that would be viewed by an observer traveling on a straight path, and may exhibit sensitivity to path in the absence of rotation. Primate locomotion along curved paths typically involves rotation, so, if the neurons discovered by (Froehler and Duffy, 2002) are in fact path-selective, it remains unclear how their response patterns would generalize to more natural movement conditions. Our model proposes mechanisms by which the MSTd neurons identified by (Froehler and Duffy, 2002) elicit differential firing rates when the instantaneous visual motion appears the same, yet the monkey moves CW or CCW around the circle. Our analysis integrates the findings with other known properties of MSTd neurons.

13.1.3 Spiral-selective MSTd cells dynamically encode path and heading direction

Figure 13.3: Spiral space continuum of motion patterns employed in electrophysiological studies to probe cell selectivity to spiral motion. Sensitivity to radial expansion and center motion is tested by the left and right ends of the continuum, respectively. Spiral patterns exist in between as an interpolation between the radial and center patterns. The spiral space also contains contracting spirals and those with CCW orientations (not shown).
If the primary role of MSTd were to determine heading, most MSTd neurons would be expected to preferentially respond to radial expansion and contraction. While many neurons in MSTd are tuned to such patterns, many others exhibit preferential responses to patterns in a spiral space spanned by radial and center templates (Figure 15·3). Moreover, neurons in MSTd would be expected to discount retinal rotation, as many appear to do with extra-retinal rotation (Bradley et al., 1996), to recover the FoE location if heading detection was of paramount importance. However, (Orban et al., 1992) demonstrated that MSTd neurons tuned to radial expansion did not respond to expansion displays with added rotational components, introducing simulated retinal rotation. Therefore, retinal rotation does not appear to be discounted in MSTd neurons and may be integral to MSTd response properties. (Graziano et al., 1994) found that more neurons preferentially responded to CW and CCW spirals than to rotation or contraction, and the tuning curve width and selectivity did not differ across the MSTd population for radial, spiral, and center patterns. That is, neurons tuned to radial expansion did not exhibit sharper tuning curves than those tuned to spirals. Spiral tuning also appears in neurons in the ventral parietal area (VIP) (Schaafsma and Duysens, 1996) and area 7a (Read and Siegel, 1997), two of the brain regions to which MSTd projects (Born and Bradley, 2005). Despite the response tuning diversity in MSTd, no functional role has been proposed for MSTd neuron tuning in spiral space.

We claim that selectivity to optic flow across a spiral space continuum simultaneously affords MSTd with sensitivity to the curvature of the path and to the heading direction. When an observer travels along a curvilinear path on a ground plane with a fixed direction of gaze, a spiral-like pattern is experienced and optic flow contains rotation that specifies the path curvature (Li and Cheng, 2011b). Theoretically, spiral selective neurons should be sensitive to the curvature of their preferred spiral pat-
tern and would therefore be capable of extracting information about the future path. Although the spiral space tuning is not likely to be cleaned defined mathematically \textit{in vivo} (Mineault et al., 2012), we assume MSTd spiral space selectivity spans the continuum between radial and center patterns that has been electrophysiologically tested (Orban et al., 1992; Graziano et al., 1994; Schaafsma and Duysens, 1996; Read and Siegel, 1997).
Figure 13-4: Schematic depiction of the model coding of path curvature and heading in MSTd. Neurons in a model MSTd hypercolumn possess selectivities across a spiral space spanning CW, CCW, radial expansion, radial contraction, and center motion patterns. The length and width dimensions of the schematic MSTd selectivity volume correspond to neurons with 2D visuotopic tuning. Therefore, at every position in the visual field, there is a model MSTd hypercolumn with a full set units tuned to radial, spiral, and center optic flow patterns. The hypercolumn expanded on the top left corresponds to MSTd units with receptive fields centered on the top left portion of the visual field, which have focus of expansion or center of motion tuning in that location. Travel along a circular path elicits a distribution of activity within the MSTd volume (overlaid heat map). The position of the activity peak across the volume in the spiral space (depth) dimension corresponds to the model path curvature estimate, and the 2D position of the peak in the spatial dimensions (length and width) indicates the estimated heading direction.
Figure 15-4 shows a visualization of the proposed functional organization of MSTd with respect to spiral space tuning. Each cylindrical volume represents a functional MSTd hypercolumn with respect to spiral selectivity. A hypercolumn contains a subpopulation of MSTd neurons that are sensitive to a spectrum of optic flow patterns in spiral space that have receptive field centered at the same location of visuotopic space. The horizontal and vertical axes specify the spatial dimensions of the shown MSTd visuotopic map. Each point in this two-dimensional space corresponds to the subpopulation of MSTd neurons tuned to expansion, contraction, spiral, or center optic flow patterns that have the FoE, FoC, or more generally the center of motion (CoM) centered at that visuotopic location. For example, the subpopulation of MSTd neurons shown in the top-right portion of Figure 15-4 that are tuned to radial expansion preferentially respond when retinal motion patterns contain an FoE located on the top-right region of the visual field. The axis than spans the depth of the cylinder represents the degree of spiral tuning for the subpopulation of neurons that have receptive fields centered at a particular location of the visual field. The depth of each hypercolumn contains MSTd neurons tuned to CW or CCW spiral patterns that either expand or contract. Spiral patterns smoothly vary in ‘spirality’ along the space between patterns that are radial with no curvature (top and bottom), and those that are centers (left and right). We propose that the ‘spirality’ of the most active subpopulation of neurons in MSTd encodes the curvature of the path, and the two-dimensional visuotopic position of that maximally active subpopulation represents the heading.

In the simple case of traveling along a straight path, we expect neurons on the radial expansion portion of the spiral space continuum to be most active, indicating no path curvature, and we anticipate the peak to be spatially coincident with the FoE, indicating the heading. Therefore, the population MSTd response in this example is
the same as if there were only neurons selective to radial patterns. In the case of a circular path, we expect the spiral-selective neurons with spiral arms that best match the path curvature to be most active. As reported in (Royden et al., 1994; Li and Cheng, 2011b), different gaze patterns modulate the rotation present in the optic flow. In the present paper, we test whether the maximal activity of neurons tuned in spiral space map onto human judgments of path curvature as gaze varies.

We present a dynamical model of primate MSTd that builds on electrophysiological findings and explains a range of human psychophysical data on path and heading perception with and without eye movements. The main goal is to present a mechanistic hypothesis of path perception that provides a unified framework to interpret psychophysical and neurophysiological data on heading and path perception. Our model goes beyond existing heuristics by providing a mathematical description and biologically-plausible implementation that is readily testable. Our analysis and simulations show that the model yields performance similar to humans under different gaze conditions, circular path radii, and eye movement patterns. The model predicts that the neurons reported by (Froehler and Duffy, 2002) obtain their path selectivity through a spiral pattern tuning.
Figure 13-5: Diagram of model V1-MT-MSTd. First-order local motion is computed in model V1. Model MT receives projections and spatially pools motion signals from model V1. A vestibular eye velocity gain field acts on the afferent signals from model MT in MSTd, which compensates for rotation introduced by pursuit eye movements proportional to the eye movement speed in the direction opposite that of the eye movement. A template match occurs in model MSTd, whereby the similarity is assessed between the afferent motion signal and motion field templates sampled in spiral space. A distance-dependent weighting exponentially discounts vector matches by distance from the template singularity. Finally, neurons selective to different spiral patterns, expansion and contraction, CW and CCW orientations, and 2D visuotopic location compete.
13.2 Materials and Methods

Our objective was to create a biologically plausible model of the primate visual system that demonstrates the mechanisms by which perception of heading and path may arise from populations and systems of neurons that process optic flow. The model consists of systems of *shunting* differential equations, each of which models the activity neurons in cortex (Grossberg, 1973). This architecture affords realistic neural temporal and competitive dynamics, including recurrent competition and feedback, gain control, and normalization. By creating a computational model using known functional properties of neurons in the magnocellular pathway of the dorsal stream, we can simultaneously connect neurophysiological mechanisms to human data and our test our hypotheses on diverse types of psychophysical data.

13.2.1 Model Area Descriptions

The proposed neural model contains three stages corresponding to primate primary visual cortex (V1), medial temporal area (MT), and the dorsal medial superior temporal area (MSTd) (Figure 15·5). In this paper, we do not model retinal input, but rather use analytical equations to model the representation in V1. A prior version of the model demonstrates how retinal inputs are processed through neural circuits to generate those representations (Browning et al., 2009a; Browning et al., 2009b)

**V1 (Local motion detection)**

We generated videos of dots distributed on a ground or frontoparallel plane, which served as input to the model. The videos approximate the visual displays shown to human subjects in psychophysical experiments that assess human heading and path perception. The local motion of the dots was computed according to a planar pin-hole camera model (Raudies and Neumann, 2012) and the first-order optic
flow equations with translation vector $\vec{T} = (T_x, T_y, T_z)$ and rotation vector $\vec{R} = (R_x, R_y, R_z)$ (Longuet-Higgins and Prazdny, 1980):

$$
\begin{pmatrix}
  u_x \\
  u_y 
\end{pmatrix} = \begin{pmatrix}
  \hat{x} \\
  \hat{y} 
\end{pmatrix} = \frac{1}{Z} \begin{pmatrix}
  -1 & 0 & x \\
  0 & -1 & y 
\end{pmatrix} \begin{pmatrix}
  T_x \\
  T_y \\
  T_z 
\end{pmatrix} + \begin{pmatrix}
  xy \\
  (1 + y^2) \\
  -xy \\
  -x
\end{pmatrix} \begin{pmatrix}
  R_x \\
  R_y \\
  R_z
\end{pmatrix}.
$$

Eq. 15.8 computes the model V1 representation of the dot motion $(u_x, u_y)$ (i.e. first-order optic flow), which corresponds to the instantaneous velocity of each projected dot. In Eq. 15.8, $Z$ signifies the depth of the projected dot in the world and $(x, y)$ correspond to the spatial position in the 2D projection plane. Values for the parameters $\vec{T}$ and $\vec{R}$ varied according to experimental conditions and follow in Experimental Descriptions. For simplicity we use a Cartesian representation of space in V1, although prior work has demonstrated how motion can be processed with cortical magnification (Elder et al., 2009).

**MT (Motion pooling)**

Model MT units that pool the V1 response vectors $(u_x, u_y)$ component-wise with a Gaussian receptive field kernel $G_{MT}(\mu_{MT}, \sigma_{MT})$. We configured model MT neurons with $\mu_{MT} = 0$, $\sigma_{MT} = 0.05^\circ$, and radius $r = 3^\circ$, as in (Layton et al., 2012a) to mimic the larger receptive fields in MT as compared with V1. Model MT units respond to large fields of uniform motion and project to MSTd. The pooled model V1 activity in model MT is denoted $(v_x, v_y)$.

**MSTd (Gain fields, spiral template matching, recurrent competition)**

Model MSTd consists of three stages: 1) eye velocity gain fields, 2) template matching in spiral space, and 3) dynamical recurrent competition. When the eye velocity is
nonzero (e.g. during a smooth-pursuit eye movement), a vestibular signal $\vec{p}(t)$ acts presynaptically to MSTd (Churchland et al., 2005; Elder et al., 2009):

$$ (w_x, w_y) = (v_x, v_y) - \vec{p}(t). \quad (13.2) $$

In Eq. 13.2, $(v_x, v_y)$ represents the output of model MT and $(w_x, w_y)$ is the result of the vestibular compensation. We simulated the conditions of (Cheng and Li, 2012) whereby subjects made judgments about their future curvilinear path while visually tracking a horizontally moving target. Because the target moved at a constant velocity and the experimenters discarded data $150 \text{ msec}$ from the onset of the eye movement, we set $\vec{p}(t) = (\nu, 0)$, where $\nu$ is proportional to the mean pursuit eye movement speed across subjects in each respective condition. The sign depends on the eye movement direction, which varied in the experimental conditions of (Cheng and Li, 2012).

We generated spiral templates that spanned the entire visual field through the interpolation between radial and center vector field patterns (Figure 15·3) (Grossberg et al., 1999). Eq. 13.3 defines a radial field $A$ and a center field $B$:

$$ A = \Delta_A (x - x_0, y - y_0) \quad B = \Delta_B (y - y_0, -(x + x_0)) \quad \Delta_A, \Delta_B \in \{-1, 1\}. \quad (13.3) $$

Radial expansion and contraction templates are obtained by setting $\Delta_A = 1$ and $\Delta_A = -1$, respectively. Center templates with CW and CCW orientations are constructed by setting $\Delta_B = 1$ and $\Delta_B = -1$, respectively. The values of $x_0$ and $y_0$ determine the horizontal and vertical spatial offset of the FoE in the radial template and the center of motion (CoM) in the center field. Eq. 13.4 defines the spiral template, and the value of $\psi$ determines the degree of spirality, with $0 \leq \psi \leq 1$.

$$ C = (1 - \psi)A + \psi B \quad (13.4) $$
When $\psi = 0$, the template is radial, when $\psi = 1$, the template is a CW center, and $C$ is a spiral template for other values of $\psi$.

We created a neural model with 11500 MSTd neurons with motion pattern selectivities determined by the templates in spiral space. Each model neuron receives afferent signals from model MT, which is passed through a template match to assess the degree of similarity between the input signal and model neuron’s pattern tuning. The match score at time $t$, $M(s, o, x, y, t)$, for the neuron at location $(x, y)$ with preferred spirality $s$ and orientation $o$ is computed according to the following inner product:

$$M(s, o, x, y, t) = \lambda \sum_{m,n} e^{-((m-x)^2+(n-y)^2)} \times \left( \sum_{w_x,w_y} C(s, o, m, n) \circ (w_x, w_y) \right)$$  \hspace{1cm} (13.5)

Eq. 15.14 computes an inner product (i.e. cosine similarity) by performing component-wise multiplication, indicated by $\circ$, between the input optic flow $(w_x, w_y)$ and the spiral template $C$. The result is normalized by the $L^2$ norm of the optic flow vector and the vector components are summed. An exponential distance-dependent weighting is applied to give matches near the center of motion greater weight, following by the summing over all spatial locations to obtain a scalar match score. The parameter $\lambda$ is set to the reciprocal of the number of dots such that the match score is not biased by the number of vector samples.

Eq. 15.15 defines a dynamical competitive network that describes the activation of model MSTd neuron $S$ at spatial location $(x, y)$ that is selective to a spiral pattern with spirality $s$ and spiral orientation $o$ (CW v.s. CCW).

$$\frac{dS_{s,o,x,y}}{dt} = \epsilon(-\alpha S_{s,o,x,y} + (\beta - S_{s,o,x,y})(S_{s,o,x,y}^2 + f(M(s, o, x, y, t)))
- S_{s,o,x,y} \left( \sum_{i\neq s} \sum_{j\neq o} \mu \left( \sum_{k\neq x} \sum_{l\neq y} S_{i,j,k,l}^2 \right) \right) \right).$$  \hspace{1cm} (13.6)
Eq. 15.15 is a recurrent competitive field and is configured as a contrast-enhancing or winner-take-all network (Grossberg, 1973). Competition between neurons in the network occurs across location and spiral template space. The constant $\epsilon$ is defined as the reciprocal of the membrane time constant of the model neuron and scales how fast the neuron responds, $\alpha$ signifies the passive decay rate, and $\beta$ is the saturation upper bound of the model neuron. In Eq. 15.15, the inhibition model neurons receive from others in the network that have a different spiral pattern and orientation sensitivities is set to unity weight, and $\mu$ differentially weights the spatial competition. Table 15.1 summarizes parameters values that were used in configuring the MSTd dynamics. The function $f(w)$ in Eq. 15.15 is a sigmoidal transfer function defined as

$$f(w) = \frac{([w - \Gamma]^+)^2}{\zeta + ([w - \Gamma]^+)^2},$$

(13.7)

where $[\cdot]^+$ indicates the half-wave rectification $\text{max}(\cdot, 0)$, $\Gamma$ is a threshold on the input from model MT, and $\zeta$ is a sigmoid shape parameter defining the inflection point.

Path curvature $c^*$ and heading $h^*$ is computed according to Eqs. 13.8 and 13.9, respectively, by considering the spirality and spatial position that elicited the maximal MSTd subpopulation activation.

$$c^* = \arg\max_o S_{s,o,x,y}$$

(13.8)

$$h^* = \arg\max_{(x,y)} S_{s,o,x,y}$$

(13.9)

All simulations were run on a 8-core 2.66 GHz Mac Pro with 64 GB of memory using Mathematica 8. Routines involving numerical integration of network dynamics (Eq. 15.15) and template matching (Eq. 15.14) were written in C++. Parameter values listed in the text specify those that remained constant throughout all simulations.
13.2.2 Experimental Descriptions

Unless otherwise noted, all simulation parameters matched those used in the following psychophysical experiment descriptions.

Path Perception & Gaze

We simulated the five experimental conditions of (Li and Cheng, 2011a) to compare path estimates produced by the model to those produced by human subjects (Figure 15·6). In the experiment, subjects viewed computer displays in which an observer traveled along a circular path. All coordinates are given with respect to a three-dimensional world coordinate system whereby the origin corresponds to the center of the circular path, the observer begins movement at \((X, Y, Z) = (\delta r, \bar{y}, 0)\), and the observer’s position at time \(t\) is given by \((\delta r \cos(\omega t), \bar{y}, \delta r \sin(\omega t))\), where \(r\) represents the radius of the circular path, the observer either moves CW or CCW about the path, \(\delta\) is 1 for CCW path traversals and -1 for CW traversals, \(\omega\) signifies the rate of traversal around the circle, and \(\bar{y}\) corresponds to the observer eye height. The observer translation vector \(\vec{T}\) is given by:

\[
\vec{T} = (T_x, T_y, T_z) = \frac{d}{dt}(\delta r \cos(\omega t), \bar{y}, \delta r \sin(\omega t)) = (-\delta r \omega \sin(\omega t), 0, \delta r \omega \cos(\omega t)) \tag{13.10}
\]

Each trial lasted 1 sec during which the observer traveled 3 m around the circular path. Therefore, we fix \(\omega = \frac{3}{r}\). No trial resulted in a traversal greater than a quarter circle. Since the observer motion remained parallel with respect to the XZ plane throughout each trial, \(T_y = 0\). In the experiments of (Li and Cheng, 2011a), the gaze conditions were simulated in the computer display while subjects fixated a stationary fixation cross above the ground plane horizon throughout the trial. Gaze was simulated to only vary at eye height. Because the eyes of human subjects did not move throughout the trial, the sources of rotation were both simulated gaze and the
path curvature. Therefore, $R_y$ depended on the gaze condition and $R_x = R_z = 0$.

Each condition was identical except for the simulated observer gaze (i.e. no eye movements). In the Z-axis condition, an observer was simulated to travel on a circular path and gaze remained parallel to the Z-axis (Figure 15-6a). The instantaneous vector field contained no rotation, the field at any time appeared to radially expand, and over time the FoE laterally ‘drifted’. Since there was no rotation in the Z-axis condition, $R_y = 0$. In the outside path condition, the simulated gaze was on a target 15° outside the circular path (Figure 15-6b). In this case,

$$R_y = \frac{r\omega(x_0\cos(\omega t) + z_0\sin(\omega t) - r)}{r^2 + x_0^2 + z_0^2 - 2rx_0\cos(\omega t) - 2rz_0\sin(\omega t)},$$  \hspace{1cm} (13.11)

where $(x_0, \bar{y}, z_0)$ is the position of the simulated gaze target, which was $rm$ from
Figure 13.6: Observer gaze conditions during travel along a circular path tested in the model from (Li and Cheng, 2011b). The gaze in each condition is “simulated” within the computer display because human subjects in the experiments of (Li and Cheng, 2011b) maintained fixation throughout the trial. We also tested the model on analogous conditions with pursuit eye movements (see Figure 13.10).

(a) Z axis condition. The observer maintains a fixed body, head, and eye orientation, in the direction of the ‘Z axis’, during travel along the circular path. The optic flow field at every instant is radially expansive, and over time the FoE drifts horizontally. (b) Outside path condition. Observer gaze was maintained on a target positioned 15° outside the path. (c) On path condition. The observer maintained gaze on a target on the future path positioned 30° from the initial heading. (d) Inside path condition. Observer gaze was maintained on a target positioned 15° inside the path. (e) Gaze along heading condition. Observer gaze is always tangent to the circular path, which is most often the case during human locomotion. Human subjects in the experiments of (Li and Cheng, 2011b) underestimated path curvature in the Z axis, outside path, and on path conditions, overestimated path curvature in the inside path condition, and yielded low error in their judgments in the gaze along heading condition.

the observer’s initial position (see (Layton and Browning, 2013) for derivations). In the on path condition, the simulated gaze was on a target 30° away from the initial heading and \( R_y = \frac{\omega}{2} \) (Figure 15.6c). In the inside path condition, the simulated gaze was on a target located 15° inside the path and \( R_y \) is equal in magnitude but not direction to the value in the outside path condition (Figure 15.6d). The gaze along heading condition is the natural case whereby the observer’s gaze was simulated to be aligned and rotate with the body and the observer’s heading was always tangent to the path, so \( R_y = \omega \) (Figure 15.6e).

For all path conditions, the observer traveled along circular paths with radii 28 m, 38 m, and 58 m. The environment consisted of 200 dots randomly distributed along a ground plane 1.4–50 m in depth. An analysis of model performance as a function of environmental dot count is shown in Results. We clipped dots that were projected
outside the 120° field of view the observer had in the experiment. The computer projector had a 60 Hz refresh rate, so we simulated observer motion across 60 frames of video.

Path Perception & Eye Movements

The experiment of (Cheng and Li, 2012) followed the same paradigm as (Li and Cheng, 2011b), but introduced real eye rotations through two conditions in which subjects performed smooth-pursuit eye movement on a horizontally moving target on the computer display. The orientation along heading condition was the same as the gaze along heading condition, except subjects tracked a target moving toward the outside of the path, which had the effect of linearizing the optic flow (Kim and Turvey, 1999). The orientation along Z-axis condition was the same as the Z-axis condition, except subjects tracked a target moving toward the inside of the path, which had the effect of adding extra-retinal rotation. The two conditions were configured such that the first-order retinal optic flow appeared the same. In the orientation along heading condition, the mean subject pursuit eye movement speeds were 1.42°/sec, 2.05°/sec, and 2.75°/sec for path rotation rates of 3.0°/sec, 4.5°/sec, and 6.0°/sec. In the orientation along heading condition, the mean subject pursuit eye movement speeds were 1.5°/sec, 2.05°/sec, and 2.63°/sec for path rotation rates of 3.0°/sec, 4.5°/sec, and 6.0°/sec.

13.3 Results

13.3.1 Path Perception & Gaze

Figure 15-7a depicts the path error obtained in each path experiment condition, averaged across the three path curvature conditions. Random-dot displays in our model simulations and human experiments both contained 200 dots. Positive and negative
path errors correspond to an overestimation and underestimation of the path curvature, respectively. Zero path error signifies an accurate assessment of path curvature. Model mean path errors agree well with those produced by humans in the experiments of (Li and Cheng, 2011b). Both the model and human subjects on average underestimated the path curvature in the Z-axis, outside path, and on target condi-
Figure 13.7: Path errors obtained by the model in the five gaze conditions. (a) Path error averaged across circular path radius. Positive and negative path errors indicate overestimations and underestimations of path curvature, respectively, and zero path error signifies veridical performance. Both humans and the model underestimated path curvature in the Z axis, outside path, and on path conditions, overestimated path curvature in the inside path condition, and elicited near veridical performance in the gaze along heading condition. (c) Model MSTd activity across spiral pattern selectivity space during an exemplar trial with a $38 \, m$ path radius for the five gaze conditions. The location of each peak across the spiral continuum determines the model estimate of path curvature. For example, in the Z axis condition (black), the MSTd activity peak occurs in the subpopulation sensitive to radial expansion ($\psi = 0$), and therefore the model indicates zero path curvature (straight path). (c) Model path errors (solid lines) compared to human data from (Li and Cheng, 2011b) (replotted, dashed lines) in the five gaze conditions as a function of path curvature. Model path errors were in good agreement in all gaze condition with those based on human judgments ($r > 0.94$), and path error decreased linearly ($R^2 > 0.95$) with path curvature. Error bars correspond to standard error of the mean (SEM).

When optic flow experienced by an observer moving along a curvilinear path is presented to the model, a subpopulation of units in a particular model MSTd hypercolumn become most active (Figure 15-4). Path curvature is coded by the spiral tuning of these most active units in MSTd. The visuotopic tuning of this maximally active subpopulation does not impact the encoding of path curvature. Figure 15-7b plots the peak magnitude of each MSTd unit tuned to a different template in spiral space, irrespective of the unit’s tuning in visuotopic space, in the five gaze conditions when the path curvature was $38 \, m$. The $x$ axis corresponds to the pattern tuning across the spiral space continuum, and the $y$ axis shows the maximal activity elicited.
by units sensitive to a particular optic flow pattern in spiral space, irrespective of its visuotopic tuning. A spirality of 0 signifies a MSTd neuron that is preferentially tuned to radial expansion, a spirality of 1 indicates a tuning to CCW center motion patterns, and intermediate values correspond to preferential responses to CCW spiral patterns. In the Z axis and outside path conditions, the maximally active MSTd unit was the one that was sensitive to radial expansion ($\psi = 0$). The positions of MSTd activity peaks in the Z axis (black) and outside path (red) conditions were to the far left of the spiral space continuum. Radially expansive patterns contain no curvature, therefore, the model signals, similar to humans, in the Z axis and outside path conditions that the path is straight.

To compute path error from representations of path curvature in the model, we have to ground the spiral continuum into perceptual space. When humans look where they are going, judgements of path curvature are accurate. This is most often the case during normal locomotion (Li and Cheng, 2011a), so we calibrate the model around the distribution of activity in model MSTd yielded in the natural gaze along heading condition (Figure 15·7b, blue). We subtracted the spirality of the peak obtained in each condition ($c^*$) from that obtained in the gaze along heading condition to yield the model path error.

The ordinal positions of peaks shown in Figure 15·7b correspond to path errors made by humans in the experiments of (Li and Cheng, 2011b). As mentioned above, the MSTd activity peaks in the Z axis and outside path conditions are produced by units tuned to radial expansion. These peaks are positioned far to the left compared to the activity peak in the gaze along heading condition, and subtraction of their abscissae yields large magnitude negative path errors, consistent with large underestimations of path curvature by human subjects. The position of the activity peak in the on path condition (pink) is closer to that in the gaze along heading condition. This
yields a negative path error, albeit lower in magnitude than those produced in the Z axis and outside path conditions. Therefore, the model signals an underestimation of path curvature, consistent with the judgments of human subjects.

Figure 15.7c compares the average path errors produced by the model (solid lines) with those yielded by human subjects (dashed lines) in the five gaze conditions of (Li and Cheng, 2011b). Model path error is assessed on 58 m, 38 m, and 28 m radii circular paths with curvatures of 0.017 m⁻¹, 0.026 m⁻¹, and 0.035 m⁻¹, respectively. Error bars in Figure 15.7c correspond to the standard error of the mean (SEM) yielded over 100 simulations of the model. Our model is deterministic, but the random dot positions in the input introduced variance into the model results. Model path estimates produced a good fit to those yielded by human subjects in the Z-axis (r = 0.98), outside path (r = 0.99), on target (r = 0.99), inside path (r = 0.96), and gaze along heading (r = 0.95) conditions. Similar to human subjects, the model overestimated path curvature when gaze was inside the path (green) that had the least curvature (0.017 m⁻¹). As the path curvature increased, path curvature estimates in the model converged to those obtained in the gaze along heading condition. In the highest path curvature condition, the model path curvature estimates followed the tendency for humans to largely underestimate the path curvature in the on target, outside path, and Z-axis conditions, respectively. Across all conditions, the decrease in path error varied as a linear function of increasing path radius ($R^2 > 0.95$).

When the observer gaze was inside the circular path (green), a bimodal distribution emerged in model MSTd, with a peak on the CCW center side of the spectrum and a subpeak located closer to the middle of the spiral continuum (Figure 15.7b). As depicted in Figure 15.7c, when the path curvature is high, the human data and model yield similar path errors in the gaze along heading and inside path conditions. The additional rotation introduced by the high path curvature makes the subpeak that
appears in the gaze inside path condition dominant (Figure 15·7b), which decreases the distance between the peak positions in the gaze along heading and inside path conditions. The increased proximity between the two peaks explains the model performance when path curvature is high. When path curvature is low, human subjects greatly overestimate path curvature in the inside path condition and yield small path errors in the gaze along heading and on path conditions. In the model, the peak in the gaze along heading condition shifts (Figure 15·7b, blue) further to the left due to the decreased path curvature in the optic flow. Hence, the distance between the positions of the peaks in the gaze along heading (Figure 15·7b, green) and on path (Figure 15·7b, pink) conditions decreases and the distance between the gaze along heading and inside path (Figure 15·7b, green) conditions increases. The shift results in increased path errors in the model for the inside path condition and comparable errors in the gaze along heading and on path conditions.

13.3.2 Different dot densities

We tested the model stability and path curvature estimation performance as a function of the number of dots in the scene. The path curvature judgments made by human subjects in the experiments of (Li and Cheng, 2011b) and the model results shown in Figure 15·7 were derived from environments with 200 dots. Figure 13·8 shows model performance across the path curvature conditions as function of scene dot count. The y axis plots the path error deviation, which indicates the relative path error compared to that obtained with 200 dots. Independent of the path radius, the model yields reliable results, with modest mean path error deviations (< 5°) even with only 25 dots. Human path curvature judgments have been tested with varying dot densities in conditions that most closely resemble those in the gaze along heading condition, and model produces similar errors to these human data (Warren et al., 1991). Path errors in scenes with greater numbers of dots than 200 also yielded low
Figure 13·8: Robustness in model path curvature estimates for scenes containing 25–1500 dots. Panels a–c plot how much path errors deviate from those shown in Figure 15·7 when path radii were 28 m, 38 m, and 58 m, respectively. Deviations in path error were modest, with mean errors falling within ±5° of those depicted in Figure 15·7. There were only small deviations in any condition when the scene contained at least ≥ 200 dots.
magnitude path error deviations, which indicates that the model results shown in Figure 15-7 are stable and model parameters did not overfit the human data.

### 13.3.3 Heading

In Figure 13-9a, heading bias in the model for the outside path, on path, and inside path conditions is compared to that of human subjects in the experiments of (Li and Cheng, 2011b). Heading was read out in the model according to the preferred 2D visuotopic position of the maximally active MSTd neurons (see Materials and Methods). Positive and negative heading errors correspond to heading judgments that were biased in the direction of and direction opposite to the path curvature, respectively. Human heading judgements were slightly biased outside the path in the outside path and on path conditions, and more greatly biased toward the inside of the path in the inside path condition (Figure 13-9a, red) (Li and Cheng, 2011b). The model produced similar heading errors, but unlike the human data, model heading estimates were veridical in the outside path condition. This occurred because the model was not sensitive enough to detect differences between the MSTd activity peaks in the Z axis and outside path conditions (Figure 15-7b), so the model signals the veridical heading. Neither heading errors produced by the human subjects nor by the model were influenced by the path radius.

Figures 13-9b-c depict the temporal evolution of the spatial distribution in MSTd of the maximally active subpopulation with competition (Figures 13-9b) and without competition Figures 13-9c. The simulation is of the Z axis condition, wherein the instantaneous optic flow is always expanding radially without rotation, and Figures 13-9c shows the activity of model neurons tuned to radial expansion. The visuotopic positions of the activity peaks in MSTd do not change due to the competition, but the model competitive interactions sharpen the spatial distribution. Any heading bias therefore is preserved in the model through the competition in MSTd.
Figure 13-9: Heading errors produced by the model during travel along a circular path. Positive and negative heading errors indicate bias in heading judgements in the direction of and the direction opposite to the path curvature, respectively. The model and humans produced small negative heading errors in the on path condition, and more substantial positive bias in the inside path condition. The model yielded veridical heading performance in the outside path condition, which occurred because the model is not sensitive enough to differences in the optic flow in the outside path and Z axis conditions. Heading bias in the model is preserved over time without (b) and with (c) competition in MSTd.
13.3.4 Path Perception & Eye Movements

Figure 13-10: Model path error in conditions that involve smooth pursuit eye movements. The optic flow that appears on the observer retinal during smooth pursuit of a horizontally moving target is identical in the orientation along heading and orientation along Z axis conditions. The orientation along heading condition is similar to the gaze along heading condition, except the observer tracks a target that moves in the direction opposite of the path curvature. The orientation along Z axis condition is similar to the Z axis condition, except the observer tracks a target that moves in the same direction of the path curvature. Similar to human subjects, the model yields low path errors for all the path radius conditions because model gain fields compensate in the direction opposite that of the eye movements. The model increasingly underestimates path curvature in the orientation along Z axis condition, similar to humans.

Figure 13-10 plots model performance in the orientation along heading and orientation along Z axis conditions, in which human subjects in the experiments of (Cheng and Li, 2012) performed smooth pursuit eye movements to track a moving target. Model gain fields signaled pursuit compensation proportional to the mean eye track-
ing speeds of human subjects, which increased with path curvature. Human subjects yielded small path errors independently of the path radius in the orientation along heading condition, but increasingly underestimated the path curvature in tandem with increases in path curvature for the orientation along Z axis condition. Because in the orientation along Z axis condition, the tracking occurred in the same direction as the FoE drift, and compensation occurs in the direction opposite of the eye movement, model path estimates increasingly underestimated the path curvature as the curvature increased. In the orientation along heading condition, whereby the pursuit eye movement was performed in the direction opposite to the path curvature, the vestibular signal in the model gain fields almost fully compensates for the additional source of rotation and path errors are small. Model results fit the human data well in the orientation along heading ($r = 0.97$) and orientation along Z axis ($r = 0.99$) conditions.

13.3.5 Is Competition in MSTd Necessary?

To determine whether competition across spiral space, spiral orientation (CW v.s. CCW), and visuotopic space in model MSTd was necessary to produce path errors comparable to humans, we selectively lesioned certain competitive interactions between model neurons. Figure 13·11 compares human and intact model mean path errors with those produced when the three types of competition in the model were lesioned. In all cases, omitting a particular type of competitive interaction in the model resulted in changes in path errors. For instance, lesioning the horizontal spatial interactions between model MSTd neurons resulted in a shift and compression in path error across all path radii: the path errors for the inside path, on path, and gaze along heading conditions converged to the same value for each path radius, and path errors in the Z axis and outside path conditions converged to a different value. Introducing lesions into model MSTd connections garnered results that did not follow similar
patterns to human judgements. Human behavioral performance is compatible with neural computation results from competitive interactions between subpopulations of cells in MSTd.
**Figure 13.11:** The impact lesions to model MSTd have on path error. The mean path errors for human subjects and the model from Figure 15.7 are plotted on the two leftmost data columns. Lesions were introduced in the model MSTd connectivity by zeroing out competitive interactions in spiral space, across spiral orientation, and across 2D space between neurons in MSTd (see Eq. 15.15). Lesions had a detrimental impact on model performance, and path curvature estimates no longer mapped onto human judgments. The three competitive interactions in model MSTd were necessary to obtain our results.

**Figure 13.12:** Heading bias yielded by the model in the simulated rotation condition with rotation rates between $\pm 6^\circ$. When human subjects fixate on optic flow displays wherein an observer moves along a straight path with rotation, humans make large heading errors in the direction of the simulated rotation and report the perception of travel along a curved path. The model (blue) produced the same sigmoidal pattern of heading bias as human subjects (red). Both sets of data points were fit well with a hyperbolic tangent function. The similarity between model and human heading bias, suggests the model mechanisms can explain the curved path percept reported by human subjects.

### 13.3.6 Simulated Rotation

In human psychophysical studies that employ a simulated eye rotation condition, the observer moves on a straight path with an added amount of rotation (Royden et al.,
However, human subjects report the perception of moving along a curved path (Royden, 1994). We tested whether our model produces similar heading bias to human subjects in the simulated rotation condition, which would offer a mechanistic explanation of the curved path percepts. To compute heading bias, we compared the heading garnered by the model in the gaze along heading condition with that obtained when simulating observer travel along a straight path with added rotation rates between ±6°. Figure 13·12 depicts model heading bias (blue) for different amounts of simulated rotation fitted by a hyperbolic tangent function \( a \times \tanh(bx) \), where \( a = 18.45 \) and \( b = 0.34 \), \( R^2 = 0.98 \). The red curve in Figure 13·12 shows the hyperbolic tangent function fit \( 30.88 \times \tanh(0.12x) \), \( R^2 = 0.99 \) to mean human data from (Royden et al., 1994). The sigmoidal functions fit the human data and model well, and the two were well correlated with one another \( r = 0.98 \). Figure 13·12 shows that heading was biased in the direction of the simulated rotation, which is the same sign of error observed in Figure 13·9a. Therefore, both the model and humans data exhibit heading bias in the simulated rotation condition, which may explain the curved path percepts in humans.

**13.3.7 Path Selective Neurons**

Figure 13·13a shows a model simulation of first-order optic flow experienced by the monkey in the experiments of (Froehler and Duffy, 2002). The gaze of the monkey traveling along the circular track was tantamount to that of the Z axis condition. Therefore, the radial subpopulation of cells in MSTd are expected to be maximally active due to the lack of rotation in the optic flow. However, in our simulations the maximally active model MSTd subpopulation was tuned to spiral patterns rather than those that are radial (dark orange). When the angular rotation rate \( \omega \) exceeded that used by (Froehler and Duffy, 2002) \( (\omega > \omega_b) \), MSTd neurons in the model tuned to spiral patterns remained the most active. When the angular rotation rate was
comparable to that used in the $Z$ axis condition of (Li and Cheng, 2011b) ($\omega < \omega_0$), the model neurons selective to radial patterns were most active. Our analysis indicates that temporal accumulation and the distance-dependent weighting ($e^{-(m-x)^2+(n-y)^2}$, see Eq. 15.14) used in the model induced a peak shift in spiral space, from neurons sensitive to radial pattern to those sensitive to spirals. As shown in Figure 13-13b, when the model views the visual display, the temporal accumulation and spatial weightings transform the sequence of radial patterns with a shifting FoE into a spiral pattern with a fixed FoE. When the speed around the circle is slower than that of monkey in the experiments of (Froehler and Duffy, 2002), the activity in MSTd spiral
Figure 13.13: Model simulation of the experiment of (Froehler and Duffy, 2002). (a) Responses of the maximally-active model MSTd subpopulations in spiral space as a function of the angular rotation rate (i.e. how fast the circular path is traversed per unit of time). When the angular rotation rate matched that used by (Froehler and Duffy, 2002) ($\omega = \omega_0$, dark orange), model MSTd neurons most sensitive to spiral patterns were most active. This also occurred when for larger angular rotation rates ($\omega > \omega_0$). When the angular rotation was set to a comparable rate to that used in the $Z$ axis condition (Li and Cheng, 2011b) ($\omega < \omega_0$), model MSTd neurons most sensitive to radial expansion elicited the maximal activation. (b) Simplified model mechanisms that explain why neurons that are sensitive to spirals produced the peak activity in spiral space in the simulation of (Froehler and Duffy, 2002), but did not in the simulation of the $Z$ axis condition. Consider the first-order optic flow ($A$ and $B$) at two times ($t_0$ and $t_1$) during the circular path traversal (top row). Template matching in the model is inversely weighted by distance to the FoE or CoM (second row). The third row shows the optimal templates inversely weighted by distance ($A^*$ and $B^*$). Because model MSTd dynamically integrates afferent signals from model MT, activation due to the input at $t_0$ influences the activation due to the input at $t_1$. Temporal accumulation in the model can be approximated by considering $(1 - \alpha)A + \alpha B$, which temporally blends the two weighted fields. This yields a spiral field (bottom row), and explains why model MSTd neurons sensitive to spiral patterns are most active when the angular rotation rate about the circular path is sufficiently large.

space is distributed so that the subpopulation of units tuned to radial expansion is most active. At higher speeds around the circle, the position of the MSTd peak shifts so that units sensitive to spiral patterns are most active (Figure 13.13). Our analysis suggests that the path selective neurons identified by (Froehler and Duffy, 2002) in MSTd are in fact preferentially tuned to spiral patterns, and the spiral space competition employed in our model can explain the mechanism underlying their path selective properties.
13.4 Discussion

In this article, we present experiments using a computational model of the primate dorsal stream to test the hypothesis that area MSTd can simultaneously code heading and path curvature. We posit that the underlying mechanism involves competition between neurons in MSTd that are sensitive to large field spiral motion patterns. Our model results are supported by electrophysiological data that demonstrates that MSTd neurons exhibit selectivity across a spiral continuum, ranging from radially expansion or contraction to CW and CCW center motion patterns (Figure 15·3). We tested this through model simulations of observers moving along curvilinear paths and comparing results to those garnered by studies of human path perception. We first tested the hypothesis by simulating the experiments of (Li and Cheng, 2011b), whereby observers viewed displays of traveling along circular paths with different radii and loci of gaze. The model produced similar errors to humans when estimating path curvature across the five gaze conditions (Figure 15·7). This indicates that, as for human subjects, perception of path curvature is underestimated when gaze is along a fixed direction in the world (along the ‘Z axis’), outside the path, and on a location down the future path, overestimated with gaze is inside the path, and relatively accurate when gaze changes such that it is always in the instantaneous heading direction (i.e. tangent to the circle). Figure 15·7b shows that the model explains the human path errors through the rank ordering of activity peaks distributed along the spiral space sensitivity continuum of MSTd neurons. Like humans, the model overestimates or underestimates path curvature when the gaze results in the location of peaks in MSTd spiral space to be displaced, left or right corresponding to a lower or higher $\psi$ value, respectively, compared to when gaze is along the heading direction.

In the experiments of (Li and Cheng, 2011b), human path errors were not modu-
lated by the structure of the visual scene. Our simulations of the five gaze conditions of (Li and Cheng, 2011b) with different dot densities demonstrated that the model yielded minimal deviations in path errors when dot densities exceeded that tested with human subjects, and deviations at lower dot densities were modest (Figure 13·8). This is consistent with the findings of (Li and Cheng, 2011b) that denser textured environments did not modulate human path judgments. The robustness of the model results to dot density is consistent with findings that indicate that path perception does not depend on local features in the environment (Li and Cheng, 2011b). The stability of path errors across different types of scenes in humans and the model suggests that mechanisms underlying path perception depend on an area such as MSTd that prefers large field pattern motion. MSTd receives input from MT, which is known to selectively respond to aperture-resolved coherent motion directions (Pack and Born, 2001). MT neurons may serve a crucial role in path perception by mitigating motion error estimates before the signal arrives to MSTd neurons and afford MSTd with stable path-related responses despite scene layout variations.

The simulated rotation condition of (Royden et al., 1994) presents an interesting test for the model. Although the path of travel is straight with simulated eye movements, subjects perceive traversing a curved path. Due to the added rotation in the display (Figure 15·1b), spiral cells in model MSTd elicit the most activity, and the model heading bias provides a good fit to that of human subjects (Figure 13·12). Our hypothesis predicts that humans perceive that they are traveling along a curved path in the simulated rotation condition due to the activation of spiral-selective neurons in MSTd. Conversely, in the Z-axis condition of (Li and Cheng, 2011b), human subjects responded as if they were traversing a straight path despite actually traveling along a curved path. In this case, radially expansive neurons in the model elicited the maximal activity, which signals a lack of curvature in the path and is consistent
with human path errors. In the simulated rotation and Z axis conditions, the spiral space mechanisms in the model correctly predict the perceived path curvature. This suggests that humans rely on retinal rotation (i.e. rotation not due to extra-retinal sources) to perceive the curvilinear path and that MSTd neuronal tuning to spirals extracts information about path curvature. The large heading bias produced by humans when simulated retinal rotation is added (Royden et al., 1994) is consistent with the finding of (Orban et al., 1992) that MSTd neurons tuned to expansion do not appear to compensate for rotational components in the optic flow field except when accompanied by an extra-retinal signal. The model heading bias in the outside path, on path, and inside path gaze conditions were also comparable to those found in human subjects (Figure 13·9).

In order to relate model MSTd activity to human path error judgements, we compared the location of the MSTd activity peak in spiral space (Figure 15·3 and 15·7c) to that yielded when the observer gaze was aligned with the heading direction. The reasons for this are twofold. First, directing gaze in the direction of the instantaneous heading is the natural condition that most often accompanies activities, such as locomotion and driving. Gaze along heading appears to be important for human perception of path because only in this gaze condition did humans accurately assess the path curvature (Li and Cheng, 2011b). The results of (Li and Cheng, 2011b) are supported by a number of similar studies (Warren et al., 1991; Fajen and Kim, 2002; Saunders, 2010; Saunders and Ma, 2011). When human mothers carry their infants, statistics during locomotion indicate that gaze is most often maintained within 20°of the heading direction (Raudies et al., 2012). Second, human perception of metric space has been demonstrated to be inaccurate and it therefore would seem more likely that humans perceive path relative to conditions afforded during normal locomotion (i.e. when gaze naturally changes with heading direction) rather than
perceiving path in absolute terms. For example, humans exhibit distorted judgments of distance and slant (Norman et al., 1996; Witt et al., 2004). The rank order of the model MSTd activity peak positions in spiral space followed that of path errors made by human subjects (Figure 15·7) across different gaze and path curvature conditions. This supports the idea that humans perceive their path of travel by using the pattern of MSTd activity yielded during natural location as a reference for when gaze changes.

The MSTd maximal activity curves in Figure 15·7b exhibit different widths and sharpnesses in spiral space. Because model MSTd was configured as a winner-take-all network (Eq. 15.15), given sufficient time, the network will select a single MSTd subpopulation to be active and all other model neurons will be suppressed through the competition. At the end of the 1 sec trial, some activity distributions across MSTd, such as those produced in the on path and inside path conditions, appear sharp, while others appear broader. As noted in (Browning et al., 2009a; Browning et al., 2009b; Layton et al., 2012a; Browning, 2012), broad activation distributions across the network could implicate a greater degree of uncertainty about the path curvature and the dynamic competitive interactions require longer to resolve a clear winner. We configured model MSTd with a single set of parameters, but it is possible in vivo that different subpopulations exhibit differential response latencies (Layton and Browning, 2012).

We selected spiral templates in the model to resemble the optic flow patterns used in a number of electrophysiological studies (Duffy and Wurtz, 1991a; Graziano et al., 1994; Read and Siegel, 1997) to investigate large motion pattern selectivity in neurons located in MSTd and other areas of the STS. Although electrophysiological studies report tuning in the spiral space that spans radial expansion, contraction, and center fields, actual MSTd neuron receptive fields may exhibit far greater complexity.
(Mineault et al., 2012) modeled the feedforward subunit structure of MSTd neurons based on single-cell recordings and discovered complicated subunit configurations that deviated from characteristic radial, spiral, and center motion patterns. Feedback and other types of horizontal connectivity was not modeled, and only ~50% of the MSTd response variance was accounted for, so the actual receptive fields of MSTd units are likely even more complex. MSTd receptive fields may follow the motion statistics experienced by primates during ecological locomotion conditions along a ground surface. For instance, model templates spanned the entire visual field, but ‘ecological templates’ may be biased toward the lower portion of the visual field. The statistics of videos collected from head-mounted cameras on human mothers carrying infants show that the optic flow during locomotion is fairly evenly distributed across expanding, contracting, upward, downward, CW, and CCW motion patterns, with a bias for expansion (Raudies et al., 2012). The selectivity of MSTd neurons in the sample of (Graziano et al., 1994) also are biased toward expansive motion patterns. Humans accurately judge heading in environments with many different structures, even with dynamic occlusion, unless the textures become unstructured (Kim, 2008). Therefore, ecological statistics may be important for guiding the development of MSTd receptive fields.

In simulating monkey movement along a circular path, we found different model MSTd activity peak locations depending on the speed by which the circular path is traversed. At speeds slower around the circular track than that used by (Froehler and Duffy, 2002), the optic flow more closely mimicked the Z axis condition of (Li and Cheng, 2011b) and the subpopulation of MSTd neurons tuned to radially expanding motion patterns was most active—thereby signaling travel along a straight path. However, when the path traversal speed equaled or exceeded that used in (Froehler and Duffy, 2002), the activity peak shifted rightward, signaling navigation along a
Table 13.1: Parameter values used in simulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon$</td>
<td>1.0</td>
<td>Inverse cell time constant</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>3.25</td>
<td>Passive decay rate</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.0</td>
<td>Activation upper bound</td>
</tr>
<tr>
<td>$\mu$</td>
<td>2.5</td>
<td>Strength of inhibition from spatial competition</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>0.01</td>
<td>MSTd presynaptic threshold</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>0.07</td>
<td>Sigmoid shape parameter</td>
</tr>
</tbody>
</table>

curved path. Our analysis indicates that at a sufficiently fast speed around the track, the motion pattern MSTd neurons receive in the experiment of (Froehler and Duffy, 2002) is temporally ‘blurred’ and actually resembles a spiral pattern (Figure 13.13). (Froehler and Duffy, 2002) did not report testing selectivity to spirals in their sample. Our analysis and simulation results predict that the path selective neurons discovered by (Froehler and Duffy, 2002) were tuned to be selective to spiral rather than expansion patterns in spiral space. We predict that human subjects would produce path curvature judgments consistent with the percept of traveling along a curved path in a psychophysical experiment with the Z axis gaze condition when the rotation rate along the circle is increased. In this proposed experiment, the model makes the prediction that humans would produce different path errors in the Z axis condition, depending on how much of and the speed at which the circular path is traversed.

Our model results suggest information about future path may be processed in areas as early as MSTd. Path estimation may more fundamentally indicate the functional role of area MSTd in primates.
Chapter 14

Analysis of gaze along curvilinear paths

In the following chapter, we present mathematical derivations of the curvilinear trajectories produced by humans in the experiments of Li and Cheng (Li and Cheng, 2011b). In particular, we derive the rotation present in the optic flow field based on the geometry of the observer gaze toward targets in the Z-axis, outside path, target on path, gaze along heading, and inside path conditions. We assess the impact that recurrent competition within model area MSTd has on heading bias, and its behavior over time.

14.1 Contributions

The following section introduces the model in the form of a journal article accepted in IJCNN 2013 (Layton and Browning, 2013). As first author of the study, I developed the model architecture, implemented and tested it on the set of visual displays, analyzed the results, and wrote and revised the manuscript. The coauthor of the manuscript oversaw progress and contributed in the capacity of an academic advisor.
Chapter 15

The Simultaneous Coding of Heading and Path in Primate MSTd

15.1 Introduction

Optic flow is defined as the spatial displacement of luminance over time on a 2D image. Present neural networks for visually-guided navigation and optimization methods can competently estimate the parameters in first-order optic flow (Longuet-Higgins and Prazdny, 1980) that are necessary for computing the instantaneous direction of travel (heading) from visual motion fields (Raudies and Neumann, 2012). However, only estimating instantaneous heading when processing optic flow for the guidance of mobile robotic systems has its limitations. First, when a mobile agent travels along a circular path, the representation of path remains constant, whereas the instantaneous heading needs to be constantly recomputed. Thus, incorporating path estimation in a neural network for autonomous navigation has a computational advantage over heading. In the present paper, we use heading to refer to the instantaneous, tangential, straight course of travel, and path refers to the general future trajectory, which may be straight or curved. Second, path information is useful for future route planning and detecting potential interceptions with stationary and moving targets. Human drivers have been shown to safely negotiate steering around curvilinear road bends, even at high speeds (Wilkie and Wann, 2003b; Li and Cheng, 2011a). Therefore, anticipating and/or determining the future path of travel can prevent an autonomous
agent from veering off course.

The ability of humans to navigate on varied terrain with ease has ignited much research interest. Research demonstrates that humans can judge their heading in rigid environments with little error \((<1^\circ)\) (Warren et al., 1988), and when moving their eyes with error that depends on the eye rotation rate (Royden et al., 1992). In the presence of independently moving objects, human heading bias depends on whether the object approaches or maintains a fixed distance with respect to the observer (Warren and Saunders, 1995; Royden and Hildreth, 1996). Because humans estimate their heading with reasonable accuracy in a short amount of time, the computational mechanisms in the brain that afford this performance are of interest for neural network models that process visual motion for autonomous navigation. When an observer travels along a straight path in a rigid environment toward a frontoparallel plane without eye movements, the observer experiences a radially expanding optic flow field that contains a singularity known as the focus of expansion (FoE) that specifies the heading direction (Gibson, 1979). Neurophysiological experiments have shown that the dorsal medial superior temporal area (MSTd) in primates contains neurons that selectively respond to radially expanding optic flow fields similar to those viewed when traveling along a straight path (Duffy and Wurtz, 1991a; Orban, 2008). Neural network models have demonstrated that the activity of model MSTd neurons tuned to radial motion patterns yields the same patterns of heading errors as humans in rigid environments and in the presence of independently moving objects (Warren and Saunders, 1995; Royden, 2002; Layton et al., 2012a), indicating that MSTd may represent a crucial brain area for heading perception.

In the natural case of travel along a curvilinear path, instantaneous heading may be less important for guiding navigation than the perception of the future, potentially curved, path. Unlike the straight path case wherein the instantaneous heading spec-
ifies the future path, curvilinear navigation produces optic flow with translation, and rotation that is intrinsic to the pattern of motion on the human retina (i.e. not due to eye, heading, or body movements). In the present paper, we assume navigation is along a circular path and all rotation in the optic flow is intrinsic. Therefore, the optic flow contains information about the path curvature. When traveling along a circular path and the observer gaze is in the same direction as heading (i.e. looking where one is going), humans can accurately judge the path curvature (Warren et al., 1991), even in the presence of independently moving objects (Fajen and Kim, 2002). Li & Cheng showed that human judgments underestimate the curvature of the circular path when gaze is outside the circle and overestimate the curvature when gaze is on or inside the path (Li and Cheng, 2011b).

No existing theory of path perception explains the underlying neural mechanisms or clarifies whether or not heading is involved. We present a neural network model of path perception motivated by primate brain area MSTd that encodes path through the maximally active subpopulation of units selective along a continuum of radially expansive, spiral, and center global motion patterns, and the two dimensional (2D) visuotopic location of the maximally active unit corresponds to the heading direction. Our neural network predicts that the representation of heading and path are intimately related and dynamically interact over time.

The rest of this paper is organized as follows: Section II presents a mechanism inspired by primate brain area MSTd that may simultaneously code heading and path. Section III introduces a neural network inspired by MSTd and shows derivations of the path conditions simulated in the network. Simulations results of conditions that resemble those in human psychophysical experiments of heading and path perception are shown in Section IV. Finally, Section V concludes the paper.
Figure 15·1: Model representation of path curvature and path MSTd visualized as a 3D volume. Each position in 2D cross-sections corresponds to a neuron subpopulation that is selective to a global optic flow pattern with a vertically and horizontally displaced center of motion. The depth of the volume at a fixed spatial position corresponds to selectivity along a continuum to radial (top), spiral (middle), and center (bottom) patterns. The blue arrows show a vector representation optic flow experienced by an observer traveling along a circular path, and the orange ‘heat maps’ show sample model neuron activation in each cross-section. Model neurons compete across space, spiral orientation (not shown) and over expansion vs. contraction (not shown). Path curvature and heading are coded in the model by the location of the peak in volume depth (spiral space) and spatial location of the peak, respectively.
15.2 Heading and Path in MSTd

![Diagram of MSTd and its components]

**Figure 15.2:** Model diagram. Local motion in the retinal optic flow is analytically computed (Eq. 15.8) and it is spatially pooled in model MT. Model MSTd performs a template match (Eq. 15.14) on the model MT distribution and model MSTd neurons compete in a recurrent network (Eq. 15.15) across 2D position (top), spirality (middle), and orientation (bottom).

Due to the tendency of neurons in brain area MSTd to respond to large radial expansion or contraction, researchers have suspected that the primary functional role of MSTd in navigation is heading perception (Duffy and Wurtz, 1991a; Warren, 1998). Although most neurophysiological investigations of MSTd have focused on radial optic flow sensitivity, some studies demonstrate that many MSTd neurons exhibit highly selective tuning to spiral (Graziano et al., 1994) and center (Lappe et al., 1996; Duffy and Wurtz, 1991b) optic flow patterns. No functional role has been ascribed to spiral optic flow selectivity.

We propose that spiral optic flow selectivity affords MSTd with sensitivity to the
Figure 15-3: Experimental conditions of Li & Cheng (2011) that are simulated in the model. In each condition, an observer travels for 1 sec along a circular path defined in the $XZ$ plane. The observer heading direction is depicted by the black arrow and the observer gaze is indicated by the dashed line. Conditions varied by simulated observer gaze about the y-axis. In the Z-axis condition, observer gaze remained fixed along the z-axis (a, top panel) and the optic flow was radially expansive without rotation (a, bottom panel). In the outside path condition, simulated gaze was on a target 15° outside the future path (b). Simulated gaze was on a target on the future path 30° from the initial heading in the target on path condition (c). The gaze along heading condition is the most ecologically relevant condition whereby the observer looks where he is going (d). In the inside path condition, the simulated gaze of the observer is on a target 15° inside the future path. The bottom panels depict sample vector representations of the optic flow experienced by the observer in each respective condition.
future path of travel. When humans travel along a curvilinear path in a flat environment with a fixed direction of gaze, spiral optic flow is experienced that contains information about the path curvature. Neurons that are selective to different spirals that also vary in the curvature of their preferred spiral pattern would therefore be capable of extracting information about the future path. Figure 15·1 depicts our proposed spiral space functional organization of MSTd. Each rectangular cross-section in the volume spans the selectivity across 2D visuotopic space of neurons with same spiral tuning. For a fixed 2D location, the depth of the 3D volume corresponds to selectivity to different patterns with counter-clockwise (CCW) orientations in spiral space spanning radial fields (top), spirals (middle), and centers (bottom). There are also complementary selectivity volumes for clockwise (CW) patterns and contraction (not shown). The collection of blue arrows correspond to a vector field representation of the optic flow experienced by an observer moving along a circular path. The ‘heat map’ superimposed on each cross-section shows the spatial activity pattern of hypothetical neurons. We hypothesize that the spiral selectivity of the most active MSTd subpopulation across the volume depth codes path and its cross-sectional spatial location indicates the heading. In the simple case of traveling along a straight path, we expect neurons on the radial side of the spiral selectivity continuum to be most active, indicating no path curvature, and we anticipate the peak to be located in the 2D cross-section at the FoE, indicating the heading. Therefore, the population MSTd response in this example is the same as if there were only neurons selective to radial patterns. In the case of a circular path, we expect the spiral-selective neurons with spiral ‘arms’ that best match the path curvature to be most active.

The following neural network implements the mechanisms by which MSTd may exploit tuning in spiral space to simultaneously code path curvature and heading.
15.3 Model Description

In this section, we present a recurrent neural network model that simultaneously codes heading and path. The model consists of stages that correspond to primate primary visual cortex (V1) for local motion detection, medial temporal area (MT) for short range motion grouping, and MSTd for the sensitivity to large patterns of motion (Figure 15-2). In our neural network, units process optic flow and compete across spiral space, spiral orientation, and visuotopic location. Before we introduce the model stages, we summarize the psychophysical experiments of Li & Cheng (2011) that we simulate to compare our neural network performance with human judgments of path and heading.

15.3.1 Experimental Conditions

The displays were composed of 1000 dots distributed along a ground plane 1.41–50 m away in depth from the observer’s initial position. Subjects had a 110° horizontal field of view and fixated the center of the screen, slightly above the ground plane horizon, throughout the 1 sec simulated movement along the circular path. There were five experimental conditions that varied according to the simulated gaze direction of the observer: gaze along Z-axis (Figure 15-3a), gaze outside path (Figure 15-3b), gaze on target on path (Figure 15-3c), gaze along heading (Figure 15-3d), and gaze inside path (Figure 15-3e). In the Z-axis condition, no observer rotation occurred throughout the trial and the observer maintained gaze in the direction of the initial heading, which was defined as the z-axis of the world coordinate system. The subject observed radially expansive optic flow that contained a FoE that horizontally drifted over time. In the outside path condition, the subject was shown optic flow with simulated gaze on a target 15° outside the path with respect to the initial heading and rm away, where r represents the circular path radius. In the target on path condition, the simulated
observer gaze was on a target located 30° with respect to the initial heading on the future path. The gaze along heading condition is the natural case wherein the observer always looks in the instantaneous direction in which he is going. Finally, the observer gaze in the inside path condition was on a target located 15° interior to the path from the initial heading and at a distance of \( rm \) away. Three different circle radii were tested: 28.5 \( m \), 38.5 \( m \), and 59 \( m \). Subjects were asked to manipulate the horizontal position of a marker at the end of the trial such that it intersected the future path of travel. In separate experiments, Li & Cheng asked subjects to respond with their perceived heading at the end of the trial.

### 15.3.2 Model V1 & Optic Flow in Each Condition

Our neural network takes analytical optic flow as input corresponding to that viewed by a moving observer in each experimental condition. ViSTARS, a related neural network of primate navigation, shows how optic flow can be processed in a neural dynamical system (Browning et al., 2009a). This section shows the derivation of parameters that are necessary to specify the input to the model in each experimental condition.

Consider a three-dimensional (3D) Cartesian world coordinate system in which the observer travels CCW along a circular path whereby the origin corresponds to the center of the circle and the human observer’s eye at time \( t \) in world coordinates is

\[
p_h(t) = (x_h(t), y_h(t), z_h(t)) = (r \cos(\omega t), \bar{y}, r \sin(\omega t)), \tag{15.1}
\]

where \( r \) signifies the radius of the circular path, \( \omega \) specifies the rotation rate about the path (positive and negative for CCW and CW, respectively), and \( \bar{y} \) indicates the observer’s eye height (Figure 15·4). We fix \( \bar{y} \) to the mean eye height of Li & Cheng’s
Figure 15.4: The geometry of observer gaze conditions. (a) In the outside path condition, an observer with initial position position $p_0(0)$ travels CCW around a circular path of radius $r$ defined in the $XZ$ plane with initial gaze $\theta(0)$ toward a target outside the path, which is located at position $p_o$. The gaze direction $\theta(t)$ at time $t$ is defined in terms of the spatial displacement between the observer and target ($\Delta x$ and $\Delta z$). The inside path condition is defined similarly. (b) In the target on path condition, the target is located at an angle $\theta(0)$ on the future path with respect to the initial heading. As shown in the text, $\theta(t)$ and the target position can be parameterized with respect to the initial gaze direction $\theta(0)$. 
subjects of 1.51 m and $\omega = \frac{3}{r}$ such that the observer will traverse 3 m around the circle when the trial concludes at $t = 1$, which was the case in the experiments of Li & Cheng. Hence, the initial position of the observer’s eye is given by

$$p_h(0) = (r, \bar{y}, 0)$$

and at time $t$ the observer will have traversed an arc length of $r\omega t$ about the circle (Figure 15.3a). In the outside path and inside path conditions, when gaze is on a simulated target, the target object has a position

$$p_o = (x_o, \bar{y}, z_o).$$

In the target on path condition, the target position can be parameterized with respect to the initial angle between the observer heading and the target $\theta(0)$ and the circle radius $r$ (Figure 15.4b):

$$p_o = (x_o, \bar{y}, z_o) = (r\cos(2\theta(0)), \bar{y}, r\sin(2\theta(0)))$$

Let $\vec{T} = (T_x, T_y, T_z)$ and $\vec{R} = (R_x, R_y, R_z)$ represent the translational and rotational velocities of the observer, respectively. Note that $\vec{R}$ incorporates all sources of rotation from the eye, head, and body. Consider a point $P$ in the environment that has world coordinates $(X, Y, Z)$. We use a planar camera model with unit focal length to project $P$ onto the retinal surface. The projection of $P$ has a position $p$ in retinal coordinates $(x, y)$ given by

$$p = \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} X \\ Y \end{pmatrix}.$$
conditions hold:

\[ Z > 0 \]  \hspace{1cm} (15.6)

\[ Z|\tan(\Phi)| > X, \]  \hspace{1cm} (15.7)

for \( \Phi = \pm \frac{\pi}{2} \). The instantaneous velocity of the projected point \((\dot{x}, \dot{y})\) (i.e. first-order optic flow) that is the V1 representation in the model is given by (Longuet-Higgins and Prazdny, 1980):

\[
\begin{pmatrix}
 u_x \\
 u_y 
\end{pmatrix} = \frac{1}{Z} \begin{pmatrix}
 -1 & 0 & x \\
 0 & -1 & y
\end{pmatrix} \begin{pmatrix}
 T_x \\
 T_y \\
 T_z 
\end{pmatrix} \\
+ \begin{pmatrix}
 xy \\
 (1 + y^2) \\
 -xy \\
 -x
\end{pmatrix} \begin{pmatrix}
 R_x \\
 R_y \\
 R_z
\end{pmatrix}
\]  \hspace{1cm} (15.8)

Because the observer height does not change over time, \( T_y = 0 \) in Eq. 15.8. Across all simulated conditions, gaze is directed at the same eye height as the observer. Therefore, there are no source of rotation in the \( x \) or \( z \) directions, so \( R_x = R_z = 0 \) in all conditions.

The observer translation vector \( \vec{T} \) is given by:

\[
\vec{T} = (T_x, T_y, T_z) = \frac{d}{dt} ph(t) = (-rw\sin(\omega t), 0, rw\cos(\omega t))
\]  \hspace{1cm} (15.9)

The rotational velocity about the \( y \) axis \( R_y \) differs across conditions due to variations in observer gaze. In the \textit{Z-axis condition}, gaze remained parallel to the \( Z \) axis and therefore \( R_y = 0 \). In the \textit{outside path condition}, the observer gaze angle \( \theta(t) \) (Figure 15.4a) at time \( t \) with respect to the initial heading is

\[
\theta(t) = \arctan\left(\frac{\Delta x}{\Delta z}\right) = \arctan\left(\frac{x_0-x_h}{z_0-x_h}\right) = \arctan\left(\frac{x_0-rw\cos(\omega t)}{z_0-rw\sin(\omega t)}\right)
\]  \hspace{1cm} (15.10)

\( R_y \) in the \textit{outside path condition} is then defined as the rate of change of the rotation
due to gaze:

\[ R_y = \frac{d}{dt} \theta(t) = \frac{rv(x_o \cos(\omega t) + z_o \sin(\omega t) - r)}{r^2 + x_o^2 + z_o^2 - 2rx_o \cos(\omega t) - 2rz_o \sin(\omega t)} \] (15.11)

In the target on path condition, the parameterization of the target position with respect to the initial gaze angle (Eq. 15.4) affords the following definition for \( \theta(t) \):

\[ \theta(t) = \arctan \left( \frac{\Delta x}{\Delta z} \right) = \arctan \left( \frac{x_h - x_o}{z_o - z_h} \right) = \arctan \left( \frac{\cos(\omega t) - \cos(2\theta(0))}{\sin(2\theta(0)) - \sin(\omega t)} \right) \] (15.12)

\( R_y \) in the target on path condition, is then given by:

\[ R_y = \frac{d}{dt} \theta(t) = \frac{\omega}{2} \] (15.13)

Because the observer rotates in the gaze along heading condition such that his gaze always remains tangent to the path, \( \theta(t) = \omega t \) and \( R_y = \omega \). The parameter values for \( \theta(t) \) and \( R_y \) in the inside path condition are found by Eqs. 15.10-15.11, and if the gaze angle differs in sign from that in the outside path condition, replace \( \Delta x \) with \(-\Delta x\).

Model MT smooths the optic flow vector field \( (u_x, u_y) \) with a Gaussian filter, resulting in the vectors \( (v_x, v_y) \) (Layton et al., 2012a; Layton and Browning, 2012).

### 15.3.3 Model MSTd

**Template Match**

Model units perform a template match on the pooled optic flow representation and dynamically interact in a competitive recurrent neural network. Templates vary across four dimensions: 1) spirality, 2) orientation, 3) horizontal offset, and 4) vertical offset.

We generate spiral templates through a linear interpolation between radial and center templates (Grossberg et al., 1999). The degree of interpolation is determined by the
Table 15.1: Parameter values used in simulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon$</td>
<td>6.0</td>
<td>Inverse cell time constant</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.01</td>
<td>Passive decay rate</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.0</td>
<td>Activation upper bound</td>
</tr>
<tr>
<td>$\mu_{x,y}$</td>
<td>4.0</td>
<td>Strength of inhibition from spatial competition</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>0.4</td>
<td>MSTd presynaptic threshold</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>0.07</td>
<td>Sigmoid shape parameter</td>
</tr>
</tbody>
</table>

proportion $\psi$, where $0 \leq \psi \leq 1$, $\psi = 0$ corresponds to a center template, and $\psi = 1$ corresponds to a radial template. The spiral template space is a continuum that ranges from radial and center templates (Figure 15.1).

We created a neural network with 11500 MSTd neurons with motion pattern selectivities determined by the spiral templates. Each network unit compares its input to its preferred template, yielding a template match score that assesses the degree of similarity. The match score at time $t$, $M(s, a, x, y, t)$, for the network unit at location $(x, y)$ with preferred spirality $a$ and orientation $s$ is computed according to the following inner product:

$$M(s, a, x, y, t) = \lambda \sum_{\{x,y\}} e^{-(x^2+y^2)} \times \left( \sum_{\{v_x,v_y\}} \frac{C(s, a, x, y) \odot (v_x, v_y)}{\sqrt{v_x^2 + v_y^2}} \right)$$  \hspace{1cm} (15.14)

Eq. 15.14 computes an inner product (i.e. cosine similarity) by performing component-wise multiplication, indicated by $\odot$, between the input optic flow $(v_x, v_y)$ and the spiral template $C$. The result is normalized by the $L^2$ norm of the optic flow vector and the vector components are summed. An exponential distance-dependent weighting is applied to give matches near the center of motion greater weight, following by the summing over all spatial locations to obtain a scalar match score. The parameter $\lambda$ is set to $\frac{1}{N}$, where $N$ is the number of dots in the scene.
Network Dynamics

The ordinary differential equations that describe the temporal dynamics of neural network of MSTd densely sample the visual field and feature shunting competitive dynamics (Grossberg, 1973). These equations perform a leaky integration of their inputs and simulate many known properties of neurons, such as divisive normalization (Heeger, 1992; Carandini and Heeger, 2011) and automatic gain control. Eq. 15.15 defines a dynamical competitive network that describes the activation of network unit $S$ at spatial location $(x, y)$ that is selective to a spiral pattern with spirality $a$ and orientation $s$.

$$
\frac{dS_{s,a,x,y}}{dt} = \epsilon(-\alpha S_{s,a,x,y} + (\beta - S_{s,a,x,y})(S_{s,a,x,y}^2 + f(M(s, a, x, y, t)))
- S_{s,a,x,y} \left( \sum_{i \neq s} \sum_{j \neq a} \mu_{x,y} \sum_{k \neq x} \sum_{l \neq y} S_{i,j,k,l}^2 \right) \right). \quad (15.15)
$$

Eq. 15.15 is a recurrent competitive field and is configured as a contrast-enhancing or winner-take-all network (Grossberg, 1973). Competition between neurons in the network occurs across location and spiral template space. The constant $\epsilon$ is defined as the inverse of the membrane time constant of the model neuron and scales how fast the neuron responds, $\alpha$ signifies the passive decay rate, and $\beta$ is the saturation upper bound of the model neuron. In Eq. 15.15, the inhibition model neurons receive from others in the network that have a different spiral pattern and orientation sensitivities is set to unity weight, and $\mu_{x,y}$ differentially weights the spatial competition. The function $f(w)$ in Eq. 15.15 is a sigmoidal transfer function defined as

$$
f(w) = \frac{([w - \Gamma]^+)^2}{\zeta + ([w - \Gamma]^+)^2}, \quad (15.16)
$$
where \([\cdot]^+\) indicates the half-wave rectification \(\max(\cdot, 0)\), \(\Gamma\) is a threshold on the input from model MT, and \(\zeta\) is a sigmoid shape parameter.

All simulations were run on a 8-core 2.66 Ghz Mac Pro with 64 GB of memory using Mathematica 8. Parameter values listed in the text specify those that remained constant throughout all simulations. Table 15.1 summarizes parameters values that were used in configuring the MSTd dynamics.

15.4 Simulation Results

We compute a template match on the optic flow vector representation, which serves as the input to the neural network inspired by primate MSTd in which units dynamically compete across space and spiral space. The template match is applied on a linear transformation of the veridical optic flow and tracks the model input. Because the temporal dynamics and competition in the network may result in systematic bias in the representation of heading (discussed below), we first wish to address cases wherein there should be no heading bias, whether the heading representation in the network also demonstrates no bias. Figure 15·5 shows a comparison between the template match scores (Figure 15·5a) and network activation across space of units sensitive to radial patterns (Figure 15·5b) in the Z-axis condition over the 1 sec trial when the heading direction changed 6°. Neurons in the network do not exhibit heading bias, and exhibit the same peak locations over time as the template match, which indicates heading direction in the model. Due to the spatial competition, model neurons yield a sharper distribution than the raw template match scores, which indicates confidence in the network of the heading direction at time \(t\). Therefore, our neural network can give an unbiased estimate of instantaneous heading direction, with a higher degree of confidence compared to the template match results.

Figure 15·6 summarizes model path estimation performance compared to humans
Figure 15-5: A comparison between the raw template match (a) and the MSTd network subpopulation that is sensitive to radial patterns (b) over a trial of the Z-axis condition. The template match scores track the input optic flow as the instantaneous heading changes, and show a leftward peak shift over time. The peak position in the network corresponds to heading direction and despite the nonlinear competition and temporal accumulation, the model also tracks the instantaneous heading change in the input.
Figure 15-6: (a) Maximally active subpopulations across the spiral continuum of MSTd spiral selectivity. The left, middle, and right of the x-axis correspond to subpopulations selective to radial, spiral, and center patterns, respectively. In the Z-axis and outside path conditions, the subpopulation sensitive to radial patterns is most active. In the target on path and along heading conditions, spiral subpopulations were most active. The peak in the inside path condition was closer to the center selective neurons. The distribution of peak locations in spiral space correspond to the path curvature estimates in the model in the different experimental conditions. (b) A comparison between path errors made by the model and by human subjects in the experiments of Li & Cheng. Negative and positive path error correspond to underestimations and overestimations of path curvature, respectively. The sign and magnitude of the path errors made by the model fits the human data.
in the experimental conditions of Li & Cheng. Figure 15·6a plots the maximum model activation (y-axis) yielded for each the subpopulations of network neurons tuned to different spiral templates (x-axis). Figure 15·6a shows the network activity at the end of the 1 sec trial, after which subjects would respond. The left (right) hand side of the x-axis corresponds to sensitivity to radial (center) patterns, and in between signifies spiral pattern selectivity. In the Z-axis condition and outside path condition, the MSTd neurons with radial pattern sensitivities \((\psi = 0)\) elicited the highest activity. In the Z-axis condition, this result is expected since the optic flow at every frame of the input video is radially expansive without a rotational component. The similarity in the outside path condition and the Z-axis condition model activation explains why Li & Cheng report that humans responded as if they were moving along a straight path in the outside path condition.

Spiral subpopulations in the model yielded the maximal activation in the on path condition and along heading condition. The inside path condition produced a peak in the spiral continuum closer to model neurons tuned to center templates. The ordinal distribution of the peaks in Figure 15·6a are consistent with a response to the increasing curvature in the retinal optic flow patterns throughout the respective experimental conditions. Neurons sensitive to spiral patterns are capable of extracting this information in the optic flow.

Figure 15·6b compares the position of the peak subpopulation response in the network along the spiral continuum to human path error judgments. We calibrated the model path error readout such that the network response distribution in the along heading condition constituted zero path error. We make this assumption because human subjects in the experiments of Li & Cheng only on average yielded zero path error irrespective of the path curvature or structure of the environment in the along heading condition, the condition is the most ecologically-relevant case tested wherein
gaze changes in tandem with heading direction, the network activation due to the optic flow in this condition (Figure 15.6a; blue) is therefore more likely to approximate that during actual curvilinear navigation. To produce path error readouts in the model, we subtracted the position of the peak in each experimental condition from the location in the along heading condition in the spiral space continuum. Negative and positive path errors are defined as an underestimation and overestimation of path curvature, respectively. The ordinal distribution of path errors yielded by the model (green) exactly match that of human subjects (blue). The model also produced comparable path error signs and magnitudes to human judgments: the model underestimated path curvature in all condition except along heading and inside path, exhibited no bias in the along heading condition, and overestimated the path curvature in the inside path condition. Our neural network predicts that humans responded with large underestimations of path curvature in the Z-axis condition and outside path condition because the distribution of activity in the network is similar and yields the same peak in the subpopulation of radial pattern neurons. Curvature is overestimated in the model because the location of the network activity peak in the inside path condition falls closer to the center side of the spiral continuum than that in the along heading condition.

Vestibular input about velocity, acceleration, and jerk may play an important role to reduce path estimation error. Li & Cheng (2011) measured human path errors when subjects passively viewed optic flow. Were subjects to walk or drive, perhaps in a virtual reality environment, along a circular path and judge the path curvature, we anticipate that path error would be greatly reduced. Heading sensitive neurons in primate area MSTd are not only tuned to visual optic flow, but also vestibular inputs (Orban, 2008). It may be the case that vestibular signals are combined with those derived from visual optic flow to mitigate heading and path errors.
Figure 15.7: (a) A comparison between the heading errors made by human subjects in the experiments of Li & Cheng and the model. Positive and negative heading errors correspond to heading judgments biased in the direct of and the direction opposite of the path curvature, respectively. The model provides a good fit to the heading errors made by human subjects. (b) The position of the maximally active network subpopulation peak (x-axis) over time (y-axis) in the along heading condition. The spatial location of path curvature coding remains stable throughout the trial. (c) The position of the radial subpopulation, which is maximally active, over time in the Z-axis condition. The radial subpopulation remains maximally active across time, indicating the path is perceived to be straight, but the spatial position of the peak is displaced to the left over time, indicating a change in the perceived heading.
Human subjects in the experiments of Li & Cheng also demonstrated small heading bias in the *outside path condition, on path condition, and inside path condition*. The human pattern of bias follows that of the path errors: the *on path* and *on path conditions* yielded slight bias away from the direction of path curvature and the *inside path condition* resulted in larger bias in the direction of path curvature. Figure 15.7a compares model heading errors with those produced by humans. Positive and negative heading errors correspond to judgments of heading in the direction of and opposite that of the path curvature, respectively. Model readouts of heading are taken as the *spatial* displacement of the peak network subpopulation activity compared to the peak location in the *along heading condition*. Because the experimental conditions varied due to gaze about the y-axis, we considered peak displacements along the x-axis (parallel to the horizon). The magnitude and sign of model heading errors agreed well with those of human subjects.

Our simulation results indicate that heading and path curvature can be encoded in a single neural network that performs a template match on optic flow input. This confers a number of advantages for autonomous navigation applications. First, the simultaneous coding of path and heading can reduce the computation time needed for guiding mobile robotic agents. Second, our approach estimates path and heading in parallel based on the input optic flow. Path planning need not rely on heading estimations, which may contain bias. Third, the simultaneous coding of path and heading results in a smaller network size and less memory overhead compared to computing heading and path in separate network layers.

Finally, we study the stability of the path and heading representation over time. Figure 15.7b plots the spatial position (x-axis) of the network activity peak (white; indicated by orange line) over time (y-axis) in the *along heading condition*. Over the course of the 1 sec trial, the spatial position of the maximally active network
subpopulation does not change despite large local fluctuations in the optic flow. This invariance is consistent with the fact that the curvature of the circular path did not change during the trial. Thus, our network can produce a stable response that parsimoniously anticipates the future path of travel. Figure 15.7c shows the peak position of the maximally active subpopulation in the network over time in the Z-axis condition. In this case, the radial subpopulation is the most active, and the peak drifts from right to left over time, which is consistent with the change in heading over time. The radial subpopulation is the most active throughout the entire trial, indicating that the path is straight. Unlike in the along heading condition, the spatial position of the peak displaces over time.

15.5 Conclusions

We introduced a biologically-inspired neural network that codes path curvature by the maximally active neuron subpopulation along the spiral tuning continuum, and the visuotopic position of this model neuron indicates the heading direction. The pattern of activity in Figure 15.6a supports our hypothesis that a single network can extract path curvature and heading information in the retinal optic flow. The network performs similarly to human subjects when judging the curvature of the future path of travel (Figure 15.6b and Figure 15.7a). We showed that the heading estimates dynamically update over time. Finally, we showed that responses to path curvature along constant-curvature curvilinear paths are stable and peak activation in the network spatially displaces as instantaneous heading is expected to change.
Chapter 16

Conclusion

The central thrust of the models of the ventral and dorsal streams presented here is to better understand how the brain controls behavior by simulating the dynamical interactions that occur across neuronal populations distributed across multiple visual areas in cortex. The RGB model shows how visual areas V1, V2, and V4 are recruited to dynamically give rise to border-ownership selectivity. G cells directly implement in cortex Gestalt properties, such as closure and convexity. Gestalt principles have garnered success in parsimoniously describing perceptual grouping at a psychological level. Given that border-ownership signals exist in cortex (Zhou et al., 2000; Qiu et al., 2007) and the nonlinear grouping of smaller receptive field units by larger receptive field units along the ventral stream (van Essen and Maunsell, 1983), perhaps single neurons or neural populations implement other Gestalt properties, as predicted by the G cell. Feedback projections from large receptive field G cells is crucial to the model’s operation, which indicates that intra-areal horizontal connections are not necessary to give rise to border-ownership. This is consistent with recent computational investigations of border-ownership (Supêr et al., 2010; Mihalas et al., 2011). The model demonstrates that competition between units that have on-surround receptive field (R cells) signal occlusion at T-junctions due to the junction geometry and shunting inter-scale competition. Many existing models rely extensively on explicit rules to handle human percepts around junctions (Finkel, 1992; Zhaoping, 2005; Craft et al., 2007), and R cells provide a dynamical solution in the case of T- and L-junctions.
This suggests that the visual system may not rely on specialized junction circuits to perform figure-ground segregation. Occlusion information conferred by T-junctions may reflect a more general process of inter-scale competition of grouping cells sensitive to local convex regions. To my knowledge, on-surround competition has not been studied elsewhere, and may solve other important problems in the visual system.

We presented an extension of the RGB model that yields border-ownership signals consistent with human percepts in kinetic random dot visual displays. The model predicts the existence of magnocellular and parvocellular border-ownership cells, which have distinct connectivity with MT and V4, respectively. Magnocellular and parvocellular border-ownership cells are predicted to preferentially respond to high velocity and low velocity correlation signals, respectively. The differences in selectivity are predicted to arise from differential projections to MT and V4, which group high velocity and low velocity correlation signals, respectively. The model presented here proposes that V4 neurons that respond to kinetic shapes signal the presence of a figural region surrounded by texture accretion/deletion to V4 border-ownership cells. V2 is the likely location of magnocellular border-ownership cells, because MT neurons do not respond to kinetically-defined edges (Marcar et al., 1995). Studies have identified that cells in V2 respond to kinetic edges (Marcar et al., 2000) and cells in V4 respond to kinetically-defined figures (Mysore et al., 2006), yet their functional role in a larger perceptual context has not been proposed. We propose that a subpopulation of V2 neurons sensitive to kinetic edges also possess border-ownership sensitivity. Electrophysiological studies have shown that single neurons simultaneously demonstrate border-ownership due to contrast (Zhou et al., 2000), color (Friedman et al., 2003), transparency (Qiu et al., 2007), and disparity (Qiu and von der Heydt, 2005) signals, and a subpopulation of magnocellular and parvocellular border-ownership cells may also show selectivity in a number of these other domains.
We also presented a model of shape detection and processing in the primate ventral stream, which clarifies mechanisms by which neurons elicit responses to interior but not exterior of shapes when they appear within the receptive field. In the model, competitive interactions generate activity peaks for scales and locations at which figures are likely to appear in the visual scene. Another layer of units groups the resultant activity from spatially-offset V4 subpopulations and elicits responses inside of figures in the visual scene. Results show that the exterior of figures tends to elicit uniform activity across units with different receptive field sizes and those units are suppressed through competition. Model results elucidate the importance of convexity and closure in figure-ground segregation. Moreover, the model results suggest that feedback from IT to early visual areas may serve to propagate multiplexed, multi-resolution information about shapes in the visual scene (multi-scale radial feedback) (Gattass et al., 1988; Ungerleider et al., 2007). The analysis and simulations presented here suggest that multi-resolution information propagated via feedback may increase the response gain V4 shape-selective neurons.

The model of heading perception in the presence of IMOs shows that the visual system may not need to segment and discount motion from IMOs to perceive the direction of self-motion. Electrophysiological studies show that area MST is concerned with the trajectory of IMOs, but these cells are located in the anatomically distinct subarea MSTv (Born and Tootell, 1992; Duffy, 1998), which likely receives input from MT\(^-\), which is where cells that have receptive field suited for the detection of motion boundaries are located. Our model simulates the dynamics of cells in MSTd, which have large-field receptive fields, respond to radial expansion, the FoE position, and therefore may plausibly detect heading. MSTd receives input from MT\(^+\), which is known to contain cells that pool over motion-sensitive V1 cells tuned to a single motion direction (Snowden et al., 1991). Our model computes heading using
only the V1–MT+–MSTd pathway, and generates heading bias yielded in human psychophysical experiments (Warren and Saunders, 1995; Royden and Hildreth, 1996). Our model is compatible with the flow parsing hypothesis, but our model is concerned with a different aspect of navigation. We assert that MSTd does not factor out object motion when perceiving self-motion, but this does not exclude the possibility of using radial templates to recover the trajectory of IMOs. MSTv neurons respond to object trajectory, but it is not clear to what extent the direction of motion reflects that of veridical or retinal trajectory. A future extension of the model could investigate how object motion is recovered in the visual system, and how this relates to self-motion perception.

In an extension of the model, we showed how recurrent competition in a temporally-evolving network of MSTd neurons gives rise to spatial tuning curves and temporal dynamics reported by Dubin and Duffy (Dubin and Duffy, 2007; Dubin and Duffy, 2009). After model MT+ dynamically pools short-range motion, model MSTd incorporates recurrent competition between units tuned to different radial optic flow templates, and integrates attentional signals from model area FEF. In the model, population activity peaks indicate when the recurrent competition is most active and uncertainty is greatest about the FoE location. The nature of attention, multiplicative (Martinez-Trujillo and Treue, 2004) or non-multiplicative (Womelsdorff et al., 2008) is largely irrelevant, so long as attention has a Gaussian-like profile. Using particular signal functions to modulate the recurrent feedback affords qualitative fits of deflections in the population activity that otherwise appear to be low-frequency noise. We predict that these deflections mark changes in the balance of attention between the priming and FoE locations. In activities that recruit spatial attention, such as driving or steering, the delay in the population peak in MSTd may have important behavioral consequences. For example, attending the periphery while driving may
increase reaction times and the likelihood that a collision may occur at high speeds.

Finally, this thesis develops a model that shows how the trajectory of self-motion (path) and heading can be represented across neural populations in MSTd. The model is the first to propose that MSTd neurons tuned to spiral motion patterns are crucial to the perception of the curvilinear path of self-motion through the detection the path curvature. We showed that the model results are consistent with a number of psychophysical studies that assess bias in heading and path perception (Li and Cheng, 2011b; Cheng and Li, 2012; Saunders, 2010). Spiral patterns in optic flow occur in other contexts, such as steering toward a fixated target (Wann, 2000). While the model in its present model focuses on heading and path perception, extensions to the model could investigate how actions, such as steering, may be guided by cells sensitive to spiral motion patterns and path curvature.
References


Curriculum Vitae

Oliver W. Layton

owl@bu.edu
203-550-2044
http://olayton.com

Boston University
Cognitive and Neural Systems

Education

Boston University - 3.96 GPA
Ph.D. Candidate, Cognitive and Neural Systems, 2009 - 2013
Thesis: Neural models of inter-cortical networks in the primate visual system for navigation, attention, path perception, and static and kinetic figure-ground perception
Advised by Ennio Mingolla, N. Andrew Browning, & Arash Yazdanbakhsh

Skidmore College - Magna Cum Laude (3.79 GPA) - Phi Beta Kappa
B.A. Mathematics, 2009
B.A. Computational Neuroscience (Self-Determined), 2009
Advised by Flip Phillips & Thomas O’Connell

Publications

The Border-Ownership of Kinetic Edges: A Neural Model by OW Layton & A Yazdanbakhsh. Cerebral Cortex (In Prep).

A Unified Model of Heading and Path Perception in Primate MSTd by OW Layton & NA Browning. PLOS Computational Biology (Submitted).

Dynamics of Multi-Scale Shape Integration and Detection in the Primate Ventral Stream by OW Layton & A Yazdanbakhsh. PLOS ONE (Submitted).

The Simultaneous Coding of Heading and Path in Primate MSTd, by OW Layton, International Joint Conference on Neural Networks (IJCNN) 2013 (Accepted).


Grants & Awards

Boston University
AFOSR Visual Depth Perception From Motion through Texture Accretion and Deletion: A Neural Model of Figure–Ground Segregation and Occlusion (FA9550-12-1-0436).
Coauthored and contributed to funded proposal. 2012 - present
Outstanding Teaching Fellow Award, 2011
Research Assistantship, 2010 - present
Dean’s Fellowship, 2009 - 2010

Skidmore College
Phi Beta Kappa Honors Society, 2009
Pi Mu Epilson Mathematics Honors Society, 2008
Mathematical Association of America awarded by Mathematics Department, 2008
American Mathematics Society awarded by Mathematics Department, 2007
Periclean Honors Society awarded for academic achievement, 2007
Dean of Studies Conference Travel Grant, 2009
Student Opportunity Fund Research Grant - Summer Research, 2008
Highest Honors, 2006 & 2008 & 2009

Invited Talks

The VENLab (Brown University) - November 2012
Neural Dynamics of Human Visually-Guided Navigation.
OW Layton

2nd Annual Optic Flow Workshop (Boston University) - August 2012
Time Course of MSTd Cells.
OW Layton & NA Browning

Talks

International Conference in Cognitive and Neural Systems (Boston, MA) - May 2013
A Neural Model of Illusory Contours and Shapes: A Multi-scale Investigation.
OW Layton & A Yazdanbakhsh

MODVIS (Naples, 2013) - May 2013
A multi-scale model of figure-ground segregation in the primate visual system.
OW Layton & A Yazdanbakhsh

Vision Sciences Society (Naples, FL) - May 2012
Recurrent Competition Explains Temporal Effects of Attention in MSTd.
OW Layton & NA Browning

Memberships & Affiliations

BU Optic Flow Group (http://opticflow.bu.edu)


Phi Beta Kappa
Vision Sciences Society

Posters

Vision Sciences Society - May 2013
Path is Encoded by Spiral-Selective Cells in MSTd.
OW Layton, & A Yazdanbakhsh

Vision Sciences Society - May 2013
Multi-scale Selectivity to Figures in Primate V4.
A Yazdanbakhsh & OW Layton

Society for Neuroscience (New Orleans) - October 2012
A Neural Model of Motion-Induced Border-Ownership in Random Texture Displays.
OW Layton, E Mingolla, & A Yazdanbakhsh

International Conference on Cognitive and Neural Systems - May 2012
Multiplicative Motion Correlation Detection Produces Stable Border-Ownership Signals.
OW Layton, E Mingolla, & A Yazdanbakhsh

NSF inter-Science of Learning Centers Conference (San Diego) - April 2012
A Neural Model of Border-Ownership and Motion in Early Vision
OW Layton, E Mingolla, & A Yazdanbakhsh

Vision Sciences Society - May 2012
A Neural Model of Border-Ownership and Motion in Early Vision.
A Yazdanbakhsh, OW Layton, & E Mingolla

International Conference on Cognitive and Neural Systems - May 2011
Processing of Differential Motion Signals Is Not Necessary to Explain Human Heading Bias in the Presence of Independently Moving Objects.
OW Layton, E Mingolla, & NA Browning

Vision Sciences Society - May 2011
A Model of MT Motion Pooling Explains Human Heading Bias.
OW Layton, E Mingolla, & NA Browning

Vision Sciences Society - May 2011
A Neural Model of Figure-Ground Segregation Explains Occlusion Without Junction Detectors.
A Yazdanbakhsh, OW Layton, & E Mingolla

NSF inter-Science of Learning Centers Conference - March 2011
A Model of MT Motion Pooling Explains Human Heading Bias
OW Layton, E Mingolla, & NA Browning

Vision Sciences Society - May 2009
The Traveling Salesman in the Natural Environment
Conducted modeling and psychophysical study to assess human performance and optimization strategy in flat and variable outdoor terrains.
OW Layton, T O’Connell, & F Phillips
Teaching Experience

Boston University - Department of Cognitive & Neural Systems
Taught two lectures, designed one homework simulation assignment, graded homework and projects, led weekly discussion sections.

Skidmore College - Department of Psychology & Neuroscience Program
Teaching Assistant - First-Year Experience Seminar *Designing a Mind*, 2007
Held weekly robotics labs, held class study sessions, helped students revise major papers.

Skidmore College - Department of Mathematics & Computer Science
Department Tutor - Mathematics and Computer Science courses, 2006 - 2009
Courses include: Calculus I - III, Linear Algebra, Discrete Mathematics, and Introductory Computer Science I & II.

Skidmore College - Early Childhood Center
Assistant Teacher - Early childhood education, 2005 - 2008
Worked in a classroom with 3-4 year-old preschool students, worked with small groups of children on motor, art, puzzle, and elementary mathematical activities.

Professional Experience

Boston University - Department of Cognitive & Neural Systems
Graduate Student - Professor Ennio Mingolla & Arash Yazdanbakhsh, 2010 - present
Developing neurocomputational models of human navigation, figure-ground segregation, and motion.

Skidmore College - Department of Psychology & Neuroscience Program
Research Assistant - Eye Brain Vision (EBV) Laboratory, 2006 - 2009
Give lab group talks, prepare and participate in psycho-physics experiments.

Platinum Grove Asset Management - Systems Group
Hedge Fund of Economics Nobel Prize Winner Myron Scholes Intern, 2006 & 2007
Introduced new network discovery tools, automated software distribution, installed desktops and servers, desktop support.

Skidmore College - Network & Technical Services
Intern - Networking Group, 2008 & 2009
Analyzed SQL database and created reports in Crystal Reports, created a Skidmore campus map of fiber connectivity.

Programming

Mathematica, MATLAB
Implemented neural models of brightness, perceptual grouping, spatial working memory, and motion in both environments. Regularly use optimization, image processing, and ODE solver packages.
Six years of experience.

Python, Java
Implemented computer chess agent using artificial intelligence techniques of
reinforcement learning, neural networks, back-propagation, and exponentiated gradient
descent. Player considered higher-order neighborhood interactions.
Proficient in object-oriented constructs, such as inheritance, polymorphism, and interfaces.
Proficient in regular expressions and functional programming in Python. Experience with
NumPy and SciPy.
Five years of experience.

C++
Implemented optic flow detection, GPS-video syncing, Hough transform, Canny edge
detector using OpenCV.
Two years of experience.

LaTeX
Wrote all reports in graduate school and undergraduate thesis using LaTeX and BibTeX.
Five years of experience.

CSS, XHTML
Created numerous websites since high school, including olayton.com.

Microsoft Excel, Microsoft Access
Experience with pivot tables. Computed bandwidth and MAC address to user identity
associations on a college campus network by querying several databases and writing
functions in Excel.

Competitions

Skidmore College
  William Lowell Putnam Mathematical Competition, 2008
  Earned Points on challenging national exam

Western New England College
  Skidmore College Team Participant

Brown University
  ACM International Collegiate Programming Contest, Providence, RI, 2007
  Skidmore College Team Participant