

**A THEORY FOR THE VISUAL PERCEPTION
OF OBJECT MOTION**

by

Joseph W. Norman

A Dissertation Submitted to the Faculty of
The Charles E. Schmidt College of Science
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

Florida Atlantic University

Boca Raton, FL

August 2014

UMI Number: 3647571

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3647571

Published by ProQuest LLC (2014). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Copyright by Joseph W. Norman 2014

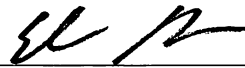
A THEORY FOR THE VISUAL PERCEPTION
OF OBJECT MOTION

by

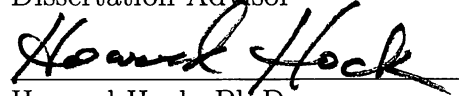
Joseph W. Norman

This dissertation was prepared under the direction of the candidate's dissertation advisor, Dr. Elan Barenholtz, Center for Complex Systems and Brain Sciences, and has been approved by the members of his supervisory committee. It was submitted to the faculty of the Charles E. Schmidt College of Science and was accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

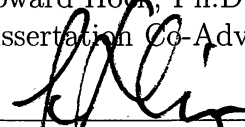
SUPERVISORY COMMITTEE:



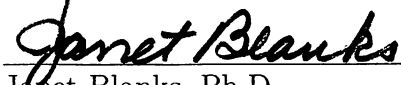
Elan Barenholtz, Ph.D.
Dissertation Advisor



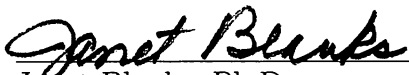
Howard Hock, Ph.D.
Dissertation Co-Advisor



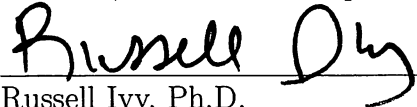
Gregor Schöner, Ph.D.




Janet Blanks, Ph.D.



Janet Blanks, Ph.D.
Director, Center for Complex Systems and Brain Sciences



Russell Ivy, Ph.D.
Interim Dean, The Charles E. Schmidt College of Science



Deborah L. Floyd, Ed.D.
Interim Dean, Graduate College

May 20, 2014
Date

ACKNOWLEDGEMENTS

Thank you: to my advisor, Dr. Howard Hock, for many fruitful discussions and for being a model of what it means to love one's work; to my co-advisor, Dr. Elan Barenholtz, for delving into the murky depths of philosophy with me; to Dr. Gregor Schöner for serving as a model of rigor and precision; and to Dr. Janet Blanks for consistent support in exploring a world of opportunities in science. I am extremely grateful to be able to call you my colleagues and friends.

ABSTRACT

Author: Joseph W. Norman
Title: A Theory for The Visual Perception of Object Motion
Institution: Florida Atlantic University
Dissertation Advisor: Dr. Elan Barenholtz
Degree: Doctor of Philosophy
Year: 2014

The perception of visual motion is an integral aspect of many organisms' engagement with the world. In this dissertation, a theory for the perception of visual object-motion is developed. Object-motion perception is distinguished from objectless-motion perception both experimentally and theoretically. A continuous-time dynamical neural model is developed in order to generalize the findings and provide a theoretical framework for continued refinement of a theory for object-motion perception. Theoretical implications as well as testable predictions of the model are discussed.

DEDICATION

To my family for their unwavering love and support: Mom, Dad, Mike, Ally, Bodhi, Trinity, Monkey, Nibbler, Chewy, and Yogi.

**A THEORY FOR THE VISUAL PERCEPTION
OF OBJECT MOTION**

List of Figures	xi
1 Introduction	1
1.1 The Perception of Object- and Objectless-Motion	1
1.2 The Counterchange Motion Detection Principle	3
1.2.1 Change detection in sensory systems	4
1.2.2 Local, Global, and Bi-local Patterns	4
1.3 The Dual Role of Perceptual Systems	5
1.4 Philosophy of Psychophysical Experiments	7
1.5 On The Modeling	9
1.5.1 Aims	9
1.5.2 Scope	11
1.5.3 Principles	13
1.6 Structure of the dissertation	15
2 A Brief History of the Study of Motion Perception	17
2.1 Gestaltists and Apparent Motion	17
2.2 Models of motion detection	19
2.2.1 Spatiotemporal Correlators	19
2.2.2 Gradient Detectors	19
2.2.3 Barlow-Levick Detectors	20

2.2.4	Counterchange Detectors	20
2.3	Optic Flow and Motion Energy	23
2.4	Conclusion	24
3	Contrasting Accounts of Short-range Motion: Direction and Shape Perception in a Random-dot Cinematogram	25
3.1	Introduction	25
3.2	Experiment	29
3.2.1	Method	30
3.2.2	Results	32
3.3	Computational Simulations	35
3.3.1	Simulations Based on the Elaborated Reichardt Detector	42
3.3.2	Simulations Based on the Counterchange Motion Detector	48
3.4	General Discussion	54
3.4.1	Source of asymmetry and reverse-phi in counterchange model	56
3.4.2	Half-Wave Rectification in the Counterchange Model	57
3.4.3	Dual Motion Pathways	58
3.4.4	The Source of Shape from Coherent Motion	63
3.4.5	Theoretical Framework for the Recovery of Depth from Counterchange Motion	65
3.4.6	Conclusion	69
4	Dynamical Preliminaries	70
4.1	Stable fixed-point model of a simple neuron with input	70
4.1.1	Stable fixed-point	71
4.1.2	Resting Level	71
4.1.3	Neuron with simple input	72
4.1.4	Stochastic fluctuations	73

4.2	Change detection neurons	74
4.2.1	Increase detection	75
4.2.2	Decrease detection	76
4.2.3	Biphasic inhibition	77
4.3	Neural Interaction	77
4.3.1	Additive and Multiplicative Synapses	80
5	A Dynamical Neural Network for Solving the Correspondence Problem	89
5.1	Introduction	89
5.2	Brief Review	91
5.2.1	Ullman’s Minimal Mapping Theory	91
5.2.2	Dawson’s Hopfield Network	92
5.2.3	Token-tracking	94
5.2.4	Hard and soft constraints	95
5.3	The model	100
5.3.1	One-dimensional correspondence network	102
5.3.2	One-dimensional Cases	114
5.3.3	Two-dimensional correspondence network	122
5.3.4	Two-dimensional cases	125
5.4	General Discussion	146
5.4.1	The necessity of cooperative inhibition	147
5.4.2	The role of the biasing array	150
5.4.3	The Dynamic Application of Constraints	153
5.4.4	Relation to neural field models	155
5.4.5	Natural Constraints	157
5.4.6	Limitations	158

5.4.7	Viability as a real-time computer vision system	160
5.4.8	Hierarchical pattern formation	161
5.4.9	Neural Correlates	164
6	Closing Remarks and Future Work	166
6.1	Change and stability	167
6.2	Perception and intentionality	167
6.3	Interplay between retinal motion and eye-movements	169
6.4	Conclusion	170
A	Computational Simulations for Chapter 3	171
A.1	Stimulus	171
A.2	Edge filters	171
A.3	Implementation of the ERD (Figure 3.4, panel a)	172
A.4	Implementation of the Counterchange Detector (Figure 3.4, panel b) .	173
A.5	Direction Decisions	175
A.6	Shape Decisions	175
B	Symmetry of Elaborated Reichardt Detector to Two-frame Same- and Inverted-polarity Stimuli for Chapter 3	176
C	Implementation, parameters, and variables for Chapter 5	178
C.1	Equations	178
C.2	Parameter and variable definitions and values	179
C.2.1	Time constants	179
C.2.2	Neuron state variables	179
C.2.3	Indices	179
C.2.4	Resting levels	180
C.2.5	Noise terms	180

C.2.6	Interaction functions	180
C.2.7	Stimulus input	180
C.2.8	Synaptic weights	181
C.3	Implementation of additive inhibition network for Section 5.4.1	181
D	Comprehensive results for motion quartet simulations in Chapter 5	182
D.1	Single motion quartet	182
D.2	Two quartets	186
D.3	Four quartets	196
E	Copyright notice for Chapter 3	206
	Bibliography	210

LIST OF FIGURES

3.1	Short-range motion stimulus	27
3.2	Experimental Results	34
3.3	Standard motion detector layout	40
3.4	Diagrams of ERD and Counterchange Detector	43
3.5	Single Trial Simulations	45
3.6	Experimental Simulations	46
3.7	Control Experiments	49
3.8	Counterchange functioning in an RDC	51
3.9	Dual Motion Pathways	62
4.1	Relaxation to stable resting level from various initial conditions	72
4.2	Neuron response to simple time-varying input	73
4.3	A random-walk vs. a neuron stabilized with negative feedback	74
4.4	Increase and decrease detector responses to simple time-varying input	75
4.5	Sigmoidal Interaction Function	78
4.6	Neural Interaction	79
4.7	Additive and Multiplicative Synapses	82
4.8	Additive and multiplicative pattern-detectors responding correctly to input	83
4.9	Additive and multiplicative pattern-detectors correctly detecting no pattern when only one input is present	84
4.10	Decreased latency with increase synaptic strength	85
4.11	False positive of an additive pattern detector	86

4.12	Additive and Multiplicative Synapses	87
4.13	Correct null response from pattern detectors with additive synapses at their maximum functional strength	88
5.1	The unique split/fusion principle	97
5.2	Cooperative inhibition in a simple correspondence problem	98
5.3	Block diagram of the correspondence network	101
5.4	Feedforward correspondence network pathway	105
5.5	A motion detector's neighborhood	107
5.6	Cooperative inhibition	109
5.7	Standard apparent motion simulation	115
5.8	Correlational motion	117
5.9	The line motion illusion	119
5.10	Splitting motion	120
5.11	Simple one-dimensional group motion	121
5.12	Expanding and Contracting Motion	126
5.13	Looming and receding motion	127
5.14	Motion quartet	128
5.15	Bistability in two-frame motion quartet with no biasing array	130
5.16	Bistability in motion quartet with no biasing array	131
5.17	Motion quartet with future-shaping interactions and noise-induced perceptual switching	133
5.18	Motion quartet with future-shaping interactions without noise	134
5.19	Motion triplet simulation	134
5.20	Simulations with two motion quartets	137
5.21	Simulations with four motion quartets	140
5.22	Visual inertia	141
5.23	Group Motion	143

5.24	Siding and splitting	145
5.25	Diagram of symbols denoting various degrees of activation for a perceptual neuron	148
5.26	Simulations for network with additive inhibition	149

Chapter 1

Introduction

1.1 THE PERCEPTION OF OBJECT- AND OBJECTLESS-MOTION

Man sieht eine Bewegung: ein Gegenstand bewegte sich von einer Lage in eine andere. [One sees motion: an object has moved from one location to another.] - Max Wertheimer, 1912

Not all motion percepts are qualitatively the same, nor do they carry the same kind of information. Think of a flowing river. We can typically perceive the direction of the current, yet it is difficult to say precisely where any *piece* of the river has moved from or to over time. The motion of the water that we perceive seems to be unattached to any particular *thing*. It is rather perceived just as we imagine the river to be, a fluid flow with a general direction but without a sense of which piece of the river went where. Now imagine a fallen leaf floating on top of the water on the river. The leaf changes location over time, yet as it does we say it is the same leaf that was once upriver that has changed its location in space over time and is now downriver. There is a perceptual identity associated with the leaf; it is an object whose position is changing, but whose underlying *thingness* remains unchanged. The flow of the water and the changing position of the leaf express unique perceptual qualities from one another and imply different opportunities for interaction for the perceiving organism.

J.J. Gibson is famous, in part, for stressing the importance of what he termed *optic flow* patterns for judging one's direction of self-motion (Gibson, 1986). While

working with Army pilots, he recognized that their success in landing depended on their ability stabilize the focus-of-expansion of an optic flow pattern. Much like the flow of the river, optic flow patterns are distributed over space and time, and sharp boundaries are not easily delineated. Like Gibson's pilots, organisms can use optic flow patterns as global estimates of heading direction without the need to differentiate or represent the particular items which are causing the flow (e.g. Kim et al., 1996). This is an *objectless* motion perception; the detailed correspondences of the features of the environment that generate the flow pattern are irrelevant. The moving organism gains ecologically vital information about its movement in relation to its environment by detecting this flow.

While the detailed spatiotemporal correspondences of visual features are unnecessary to inform the behavior afforded by global optic flow patterns, other kinds of behavior necessitate the ability to keep track of one or many visual features and where they have moved from and to over time. Detecting and tracking the changing location of the deer in the forest might be the difference between dinner and none. The visual perception of object motion entails the assignment of an invariant identity to certain subsets of visual features as they change in space and time. You not only see the motion of the deer, but you perceive that it is the same deer that was a moment ago over there, and now is somewhere else. The stimulus patterns that elicit the percept of a moving object that retains some invariant perceptual identity over successive locations are necessarily different than the global stimulus patterns which give estimates of heading and gross environmental features. In this manuscript, it is argued that the detection of *counterchange* underlies the perception of object motion across the retina. Further, a dynamical neural network is proposed to account for both the generation and (competitive) selection of motion signals. This network frames the perception of object motion as a (hierarchical) pattern-forming process, out of which

emerges a global percept.

1.2 THE COUNTERCHANGE MOTION DETECTION PRINCIPLE

The principle of counterchange follows from the basic observation that when an environmental object moves, it is both *not where it was, and is now somewhere else*. This may appear self-evident¹, but this is an ecological fact that most models of motion detection don't take into account.²

The counterchange motion detection principle then suggests a mechanism that mirrors this ecological fact. To put it in concrete terms, a counterchange detector is sensitive to the motion of a pattern x from location a to location b by virtue of detecting the (approximate) temporal coincidence of a decrease in the detection of pattern x at location a and an increase in the detection of pattern x at location b . In other words, a decrease in the presence of x where it *was* combined with an increase in the presence of x where it *is*.

A motion detection mechanism premised on this principle was first proposed in Hock et al. (2002), and the first computational model of this mechanism was published in Hock et al. (2009). A dynamical neural field version of the counterchange model was also developed in Berger et al. (2012) in order to show the viability of a counterchange detector instantiated as a temporally and spatially continuous neural model, and to probe its behavior in response to continuous motion of a single edge. This dissertation elaborates on these earlier models in several ways, including specifying several constraints that make the model robust in response to dense patterns of moving random-dots, as well as proposing a multi-scale arrangement of collections

¹Although the modern understanding of quantum physics suggests this is not necessarily a physical truth at all scales of reality.

²In fact, other common motion detector models (e.g. Barlow and Levick, 1965; van Santen and Sperling, 1984, 1985; Adelson and Bergen, 1985) take either one or the other into account, but not both.

of motion detectors and specifying their cooperative and competitive interactions.

1.2.1 Change detection in sensory systems

The counterchange principle is grounded in the notion of detecting local stimulus change (decreases and increases in local activation), so it is important that the medium of perception, the nervous system, has the potential to embody the principle if it is to be viable. It has been repeatedly observed that many sensory neurons do not simply respond to the state of their impinging stimulus, but to rates of change of that stimulus. That is, they act as differentiators. This is reflected in the existence of so-called phasic neurons, which respond transiently to changes in stimulation, but return to baseline when input is static. Phasic neurons co-exist alongside so-called tonic neurons, which typically continue firing above baseline for the duration of the stimulus to which they are sensitive (see, e.g., Arutyunyan-Kozak and kimyan, 1985). However, even tonic neurons often show their greatest activity shortly after the initial presentation of a stimulus, with activity decreasing in response to a static stimulus. This process of decreasing activity in response to static input is commonly referred to as *neural adaptation*. For this reason, some authors refer to phasic and tonic neurons as fast and slow adapters, respectively. However, there is a potentially important distinction between a true differentiator which returns to baseline when input is static, and a tonic response which is somewhat attenuated over time but does not return to baseline until input is removed. Likely, there exist a range of timescales that both (adapting) tonic neurons and phasic neurons operate over.

1.2.2 Local, Global, and Bi-local Patterns

In Section 1.1, a distinction above was drawn between the functional role of (and, therefore, detection of) global optic flow patterns and object motion patterns specified

by counterchange. So, if optic flow patterns are ‘global’, does that imply counterchange patterns are ‘local’? Commonly used detectors (ERD, ME, gradient) are conceived of as local motion detectors, whose component receptive fields are largely overlapping. They measure stimuli over a well defined portion of the visual field, one that could be circumscribed by a single boundary. However, in much of the work developed here, component receptive fields are largely non-overlapping. In other words, a counterchange detector can perceive a visual item *jump* from one location to another, non-adjacent location. In this sense, a counterchange detector could be said to be, not global or local, but *bi-local*, meaningfully connecting two (potentially) non-adjacent locations. These are the location where the motion of a visual item begins and the location where it arrives.

1.3 THE DUAL ROLE OF PERCEPTUAL SYSTEMS

Perceptual systems serve to inform an organism about the ongoing set of possible engagements with the environment. As such, they must remain in correspondence with at least some actionable features of the environment in order to be functionally meaningful. It is important, then, for a perceptual system to be able to rapidly take in potentially unexpected information, enabling situational adaptability.

At the same time, the rate at which the stimulus patterns on the sensory surface change is much faster than the duration over which goal directed actions take place. Presumably, then, for a meaningful mapping to be generated between the timescale of stimulus changes and the timescale of goal-directed behavior, certain aspects of perception must be stabilized, and meaningful spatiotemporal mappings must be generated that can inform the ongoing interaction between the organism and its environment. A perceptual system must be responsive to changing input, yet be able

to stabilize action-relevant information about transient sensory events.

The human brain is an immensely complex system. Due to the dense interconnection of the billions of component neurons, it displays a wide range of not-well-understood intrinsic dynamics. In other words, many of the activities of the nervous system are behaviors which emerge out of the endogenous structure and dynamics of the nervous system itself. But the nervous system is also coupled to the environment, so it is subject to a host of external perturbations (as well as generating self-perturbations through the environment). In this view, perception can be seen as the perturbation of the ongoing emergent dynamics of the nervous system. A balance between flexibility and internal stability must be struck. If perception is dominated by the endogenous dynamics of the brain, meaningful behavioral correspondence will cease between the organism and its environment, and the organism could be said to be hallucinating.³ If the intrinsic dynamics do not play a strong enough role, meaningful spatiotemporal correspondences will fail to maintain stability over timescales long enough to inform complex behavior. In this case, the perception would probably be something like William James's description of awareness as a 'blooming, buzzing, confusion' for the human infant (James, 1890).⁴

Marrying these dual, seemingly opposed necessities of rapid response to change and internal stability is a challenge, yet it is one that must be overcome to understand the processes leading to flexible and adaptive behavior that nevertheless maintains functional significance over multiple timescales. This dissertation is an attempt at

³In fact, Bressloff et al. (2002) have hypothesized that the geometric hallucinations often reported by those under the influence of hallucinogens such as LSD or psilocybin are a direct result of generalized disinhibition of visual cortex leading to the over-expression of its intrinsic dynamics. Additionally, Charles Bonnet syndrome is a well-known condition in which complex and vivid visual hallucinations occur after total or partial vision loss (Vukicevic M and Fitzmaurice K, 2008).

⁴In fact, there have been cases of congenitally blind individuals who gained sight relatively late in life due to medical advances, but found the newly-acquired source of optical stimulation so overwhelming and confusing as to cause depression, and often reversion back to a 'blind' lifestyle (e.g. Ackroyd et al., 1974).

fleshing out a simple theory of one microcosm of the interplay of change and stability in perception.

1.4 PHILOSOPHY OF PSYCHOPHYSICAL EXPERIMENTS

The motion stimuli used in the psychophysical experiments are, by and large, discrete in nature, both spatially and temporally. Displays generally consist of a few or very many discrete visual elements of various sizes, displayed for discrete time intervals, and are typically perceived as objects quickly ‘jumping’ from place to place across frame changes.

It might be (fairly) asked if these conditions adequately recreate the conditions of ecological viewing, and if not whether they can reveal anything important about visual perception. It is often assumed that discrete ‘apparent’ motion stimuli are somewhat unrealistic, and don’t sufficiently reproduce the smooth, continuous motions of the real world. After all, things in nature (at the scales with which are familiar) don’t simply ‘teleport’ from one place to the next without passing through the intervening space. So, the logic goes, a stimulus that does not traverse the space between two locations it sequentially occupies must not be like ‘real’ motion. Despite these concerns, there are at least several reasons to continue to explore stimuli of this nature in order to gain a deeper understanding of visual perception.

For animals with perceptual systems, one of the most relevant features of the environment is other organisms. Being able to detect and perceive other creatures may be crucial for hunting, evading predators, and realizing mating opportunities, for example. Many organisms move about the environment in a manner known as intermittent locomotion, a pattern in which movements are interspersed with pauses in which the organism remains essentially still (Kramer and McLaughlin, 2001). Thus,

it is not always the case that movements generated in nature are smooth and continuous. It might even be the case that smooth, continuous-velocity motions that last for any significant duration are exceedingly rare to observe in nature. Ecologically, the organism and its environment often interact in an *event-like* manner, such that the visual motion caused by an event occurs in essentially discrete chunks.

Ecological settings are also often quite dense and noisy. There may be moments during perception where the stimulation corresponding to a percept is temporarily unavailable. Occluders or other causes of poor viewing conditions can make an object of interest difficult to detect in certain portions of the visual field as it moves about. A jumping spider on the table may move on a timescale too fast to be detected by the visual system while it is between locations. It seems reasonable that evolution would favor perceptual systems that are robust against such momentary lapses in direct stimulation, regardless of their cause.

Furthermore, when one takes the macro-view of human activities on Earth, it can be seen that the psychophysical laboratory itself is an ecological setting. The embodied participant, typically sitting in front of a computer display, makes perceptual decisions based on generated stimuli and uses them to take environmental action (e.g. responding with a keystroke). Any complete theory of perception must be able to account for the reliable perceptual decisions and responses generated by psychophysical observers, not only those percepts deemed by the scientist to be 'ecologically valid'.

Finally, and perhaps most importantly for answering the questions pertaining to the maintenance of dynamical stability in perception, ambiguous stimuli highlight the self-organizational properties of the visual system. Minimizing the structure of stimuli allows for the possibility of 'multiple interpretations', or said in the language of dynamical systems, generates regimes of multistability, where identical stimulus presentations can give rise to multiple different percepts. While these same self-organizing

principles likely underly perception under more complicated natural viewing conditions as well, they may be more difficult to detect and study due to the enormous amount of optical information reaching the retina at any given moment, continuously interplaying with these internal tendencies. Removing a significant amount of this dense source of stimulation allows the internal dynamical tendencies to come to the fore.

1.5 ON THE MODELING

1.5.1 Aims

Models are essential to the scientific process. Models aim to distill a variety of observations or phenomena in terms of some underlying form or process. That is, a model should uncover some invariance underlying multiple instances of a system of interest. As such a model can take many forms; for example, an informal conceptual scheme or diagram aimed to organize a set of related ideas, a formalized model designed to reproduce a set of input-output relations while abstracting away the mechanisms of the system, or a simulation designed to both reproduce the behavior of the system while also capturing some of the details of the actual process from which the behavior emerges.

An information-theoretic analysis of the job of the scientist developed by Bar-Yam (2013) details the way in which the number of states a system can take on explodes exponentially as the number of relevant parameters (or experimental conditions) increases. What this implies is that it is infeasible, if not impossible, to experimentally probe a complex system under all possible conditions.⁵ In this case, a model must be developed to characterize the system in states that are not directly observed. Em-

⁵This fact has implications not only for scientific inquiry, but for the engineering of large scale systems as well. See, e.g., Norman and Kuras (2006).

pirical (phenomenological) methods, Bar-Yam argues, are not sufficient on their own to capture the behavior of complex systems. The space of possibilities is simply too vast. This dissertation serves as an example of this approach. The models function to externally represent a large number of interacting elements that would be difficult to observe empirically or to reason about analytically.

Externalizing the representation of a system of interest into a formalized model also enables one to check the soundness of one's concepts. This is an important step in assessing the validity of a concept or theory in its explanatory power. Even the most gifted human minds can only consider a small number of parts of a system and their interactions at a time. This mental myopia often leads one to see only some aspects and consequences of one's conceptual models, while masking others. Complex systems composed of many interacting parts often display emergent properties that are difficult to predict from the properties of their components. These emergent properties may reveal themselves when a model of the system and its interactions are instantiated and simulated. These collective properties are often pathological in the sense that they cause the system to behave counter to the intentions of the model-builder. These failed attempts to scale a concept to a formal model provide important scientific information. A failure of this sort implies that, 1) the concepts are not sufficient in that they entail consequences not anticipated in their original formulation and run counter to empirical observations, or 2) the formalism employed is not rich enough to capture the essence of the conceptual model. Both of these scenarios present an opportunity for scientific progress. In the first case, one is given the opportunity to reflect on the consequences of the concepts, and perhaps reformulate them to develop a better model. In the second case, one can ask what it is about the formalism that is not rich enough, and what a model would need to have in order for it to sufficiently embody the conceptual model out of which it was born. When fortune strikes, some

emergent property may reveal a *real* aspect of the system of interest. In this case one can show how a system-level property emerges out of the interactions of the system's components.

A model is also a way of asking questions of a theoretical approach. For instance, one can ask, "If the visual system functions according to this theory, how do we expect it to behave under this set of conditions?". The answer to this question becomes an empirical prediction, which may either be supported or refuted by subsequent experiments. In either case, information is gained. Perhaps, as according to the philosophy of Nassim Nicholas Taleb (2013), the failed predictions provide even *more* information about the system of interest than do successful predictions. This informational asymmetry stems from the fact that a successful prediction can only tell you that a theory *might* be correct, whereas a failed prediction tells you the theory is *certainly not* correct (under the given conditions).

To summarize, the aims of the modeling in this dissertation are to 1) characterize a complex system in states that are difficult to observe empirically, 2) assess the soundness of conceptual models, 3) hypothesize about the processes leading to observed phenomena, and 4) make predictions that can be empirically tested in order to refresh the scientific cycle.

1.5.2 Scope

In this work, what is meant by motion perception is motion relative to the retina. It might be said that when we track a moving object with smooth-pursuit eye movement we are perceiving an object in motion, but in that case the sensorimotor system works to stabilize the image on the retina such that there is minimal motion relative to it. Pursuing the interrelationship between motion relative to the retina and the initiation and termination of eye movements is an extremely important step in developing a full

account of visual perception. However, it is beyond the scope of the current work, and the discussion will have to be bracketed for the time being. As such, the timescales of interest for the current work are in the range from tens of milliseconds to several seconds, the temporal range over which eye fixations typically last.

Additionally, the models developed herein propose a *structure* on which dynamic interactions can take place between various detectors and levels of processing (described in detail in the text), out of which the percept may emerge in response to a stimulus. What is not explored in this work is the genesis of such a structure. That is, nervous systems with perceptual abilities do not materialize all at once, but are a product of formative processes at multiple timescales; namely, evolution, development, and learning. To the extent that the structure remains a viable theoretical entity, future work should seek to uncover the processes of generation that allow such a system to emerge in the world in a way that provides utility and functionality to an organism. This is an area in which little is understood scientifically at present, as the study of self-organizing systems typically focuses on spontaneous processes in systems where the parts are *given* and there is no clear functional role for the system to fulfill. Biological systems are distinct in that the parts are not simply given, but also emerge out of the same process that the parts ultimately serve; i.e. they are autopoietic (Varela et al., 1974). As such, the self-organization of functional biological structures have to be able to, in some sense, be constrained and pre-specified such that they become heritable and can be operated on by natural selection.⁶ That is, the functional structure of the nervous system is undoubtedly formed through self-organizing processes, but in order for a function to (reliably) emerge out of a self-organizing process, appropriate constraints must contextualize the formation of

⁶For a thorough treatment of this issue and attempt to begin a research program to better understand this process, see (Doursat et al., 2012)

the system which are presumably heritable.

1.5.3 Principles

The modeling in what is to follow obeys several principles that are not necessarily true of all modeling approaches, and so are worth specifying explicitly. The models presented are aimed to maintain a reasonable degree of biological plausibility. It is widely presumed (with good reason) that it is the nervous system that underlies our ability to perceive. As such, the models in this manuscript attempt to respect known features of neural tissues. For instance, most neurons couple their activity through chemical synapses. Action potentials facilitate the release of neurotransmitters at synapses, which by definition can not be negative. Therefore, all of the coupling between neural elements in the models are mediated by an interaction function that only passes greater-than-zero values. This feature of the nervous system and some its implications were explored by the cybernetician Manfred Clynes in the late 1950s and early 1960s who noted that, for example, the organization of the early visual system into complementary ON and OFF pathways for carrying information about luminance increments and decrements, respectively, was a direct result of this physiological fact.⁷ Some of the implications of this inherent asymmetry in the context of motion detection are discussed in detail in Chapter 3.

While the dissertation aims to develop models that are biologically plausible and respect known features of the nervous system, it is necessary to specify a level of

⁷Clynes generalized this concept into what he termed *rein control* to capture the ubiquity of this organizational principle in which two parallel channels are formed in order to carry information about opposite polarities or directionalities in the nervous system; especially emphasized was measuring rates of change of opposite sign. In many ways, Clynes's rein control concept foreshadowed the development of the counterchange detector, which depends on simultaneous oppositely-signed derivatives. Whereas Clynes emphasized, for example, the complementary ON and OFF pathways in the visual system, it will be seen in Chapter 3 that a viable and robust counterchange detector essentially depends on rein control *within* a polarity channel. This could be thought of as *second-order* rein control.

granularity and abstract away from the details below this level. In a practical sense, including all the known details of the physical substrate of the nervous system (e.g. the atoms it is made of) is infeasible. In a theoretical sense, including only a certain subset of underlying physical details is a way of forming a hypothesis about what the *relevant* details are. As systems show qualitatively different behaviors at different levels of analysis, a functional property at a given system-level may leverage the *emergent* properties at a level below without the more micro-level details of the system coming in to play. Such a condition could be verified by replacing some component of the system with an alternative component that preserves the relevant emergent property of the system while having a differing underlying structure, enabling the system to function normally. Through careful analysis, it might be shown that some of the excluded details do in fact matter to the macro behavior of the system. However, even if this is the case, rigorously showing this would include demonstrating the insufficiency of models excluding said detail. Additionally, limiting the number of details makes the behavior exhibited by the model tractable. If a model (impossibly) contained all the details of reality, the behavior of the model would be as mysterious as the behavior of the real system, and no deeper understanding of the system could be gleaned.

In this dissertation, neural elements are treated in an abstracted fashion. A *neural element* may be conceived as a single neuron or a collection of neurons with similar receptive fields and response properties.⁸ Neural elements have an internal activation state which represents the degree of current excitation/inhibition of that element. Individual action potentials are not modeled. Instead, the internal activation state of a neuron passes through a threshold function in order to have an effect on projected-to

⁸Throughout the text, when referring to models, the terms *neural element* and *neuron* are used interchangeably. This is not meant to imply that a neuron necessarily represents a single cell in nervous tissue.

neurons. This threshold function nonetheless respects the inherently positive value of action potentials. Details of ion channels and other small-scale physiological functions are not modeled. Instead, neurons are assumed to have an intrinsic stability that can be modeled as an fixed-point attractor. These are some of the abstractions inherent in the current approach, whether they are the appropriate ones to make remains to be determined.

1.6 STRUCTURE OF THE DISSERTATION

In what follows, a theory for the visual perception of object motion is developed. This theory emphasizes the aspect of motion perception that induces the invariance of identity over changes in location. In other words, it focuses on the role that visual information plays in determining what moved where. Remarkably, such perception does not depend on the correspondence of the detailed features of two visual items (e.g. a red square might be seen to move and become a blue circle, yet the underlying identity of the thing that has moved is perceived as stable). It is argued that the *counterchange* principle underlies the perception of object motion, as opposed to other forms of motion perception.

First, a brief review and history of the study of motion perception is presented. This is intended to contextualize the current work relative to both early motivating questions in the study of perception and more recent computational and mathematical approaches to understanding the visual perception of motion. In particular, the conceptualization of motion perception as the detection of *motion energy* (energy in the spatiotemporal Fourier domain of the stimulus) is introduced, which has served as a common theoretical basis for much of the modern work on understanding low-level motion detection processes.

The first study motivates the use of counterchange detection as a basis for the detection of object motion. Minimal mathematical models and computational simulations compare the implications of counterchange detection with commonly used ‘motion energy detection’ mechanisms (the elaborated Reichardt detector serves as a representative), and compares these results with human responses. This study serves two important theoretical purposes. 1) The random-dot cinematogram used presents a challenge to any motion detection scheme and therefore successful detection shows robustness and viability of the detection scheme to non-trivial stimuli. 2) The lack of fit between the human responses and the predictions of motion energy detection cast doubt on the explanatory sufficiency of such schemes.

The second study builds on the first in two important ways. Firstly, it extends the minimal mathematical model of the first study into the framework of dynamical neural modeling. This allows the model to apply to a much larger (less constrained) class of stimuli and to ask questions about the dynamical aspects of object motion perception (e.g. perceptual stability, switching behaviors, history dependence and hysteresis). Secondly, it proposes a perceptual principle to account for the nature of the perceived correspondence between visual elements in ambiguous stimuli and embodies that principle in a neural network architecture.

Finally, summaries and concluding remarks are presented. Limitations of the current approach as well as future directions of research are discussed.

Chapter 2

A Brief History of the Study of Motion Perception

2.1 GESTALTISTS AND APPARENT MOTION

The scientific study of motion perception arose as a coevolutionary complement to the technological achievements leading up to and culminating in motion pictures.¹ As is well-known, motion pictures present a sequence of static images that create the impression of motion with no literal physical motion present. As early as 1875, Sigmund Exner had published work detailing this so-called 'apparent motion' (AM), where sequences of changes in static lights or images elicit a motion percept in human observers. It was Max Wertheimer, one of the founders of the famous Gestalt Psychology movement, who saw AM as an opportunity to develop a new kind of psychological theory. His 1912 manuscript, *Experimentelle Studien über das Sehen von Bewegung* [Experimental Studies on Seeing Motion], carefully explored many variants of the AM stimuli and discussed theoretical implications of the experiments.

In his manuscript, Wertheimer began to hint at what would become the mantra of Gestalt Psychology, later articulated by Kurt Koffka as “the whole is different than the sum of its parts”. In the context of AM, this implies that the two static images,

¹For a wonderful overview of various domains in which technological progress preceded theoretical understanding see Taleb (2013).

presented in sequence, elicit a percept that is over and above a simple combination of the two images; namely that *motion* is a property that does not belong to either of the static entities, but is an emergent percept that cannot be reduced beyond the entire spatiotemporal pattern involving both images and their relative timing. In other words, one image (or flash of light) does not elicit *half* of a motion percept, but together the two images make a *whole* motion percept. This theoretical insight opened up a world of inquiry surrounding the nature of the relationship between parts and wholes. The exploration of this relationship continues to be one of the most exciting and challenging aspects of science, both within perceptual psychology, and in the broader context of complex systems science where one might ask how parts and wholes relate in all types of systems.

The current work presented here can be seen as a clear descendent of the Gestalt movement. This manuscript seeks to uncover the relationship between certain kinds of stimulus *patterns* and the percepts associated with them, and to explore how the local interactions of many parts can give rise to global percepts that are not simply the sum of those parts. All of the stimuli used are essentially elaborations on the simple AM stimulus. Even the simple AM stimuli used in the earliest experiments highlight points of theoretical controversy and uncertainty. There is no universally agreed upon mechanism underlying the perception of apparent motion. Several models of directionally-selective detectors are briefly discussed below, including the counter-change motion detector which is elaborated throughout this manuscript.

2.2 MODELS OF MOTION DETECTION

2.2.1 Spatiotemporal Correlators

In 1958, Bernhard Hassenstein and Werner Reichardt founded the first institute to explicitly connect the studies of physics and biology, the Max-Planck Institute for Cybernetics in Tübingen, Germany. They famously developed a mathematical model of the optomotor response of a beetle subjected to various spatiotemporal patterns of light. Essentially, the model consisted of a motion detector that received input from two spatially-separated light detectors. In order to detect motion, one input was delayed and compared to the instantaneous response of the other input via multiplication. If the signs (polarity) of the inputs were the same, a positive product would be produced and the detector would signal motion in the direction from the delayed input to the non-delayed one. If the signs of the inputs were different, a negative product signified motion in the opposite direction, from the non-delayed to the delayed input. This basic delay-and-compare architecture has served as the basis of several models to follow, and in general is known as a *spatiotemporal correlator* model, where positive correlations are associated with 'forward' motion, and negative correlations as 'reverse' motion. For examples of this class of model, see (Reichardt, 1961; van Santen and Sperling, 1984, 1985; Adelson and Bergen, 1985, Watson and Ahumada, 1985). Some of the theoretically critical implications of this class of detectors are explored in detail in Chapter 3 (in which they are referred to as spatiotemporal *comparators*).

2.2.2 Gradient Detectors

Gradient detectors (e.g., Marr and Ullman, 1981) detect local motion by measuring the local temporal derivative and dividing it by the spatial derivative of luminance.

Although this model is not explored in detail in this manuscript, typical gradient-detectors that make measurements at zero-crossing of the image suffer from some of the same shortcomings of spatiotemporal correlators. Namely, they express a symmetry with respect to polarity-inverted stimuli that is not evident in human perception. Again, this is discussed at length in Chapter 3.

2.2.3 Barlow-Levick Detectors

The Barlow-Levick model of motion detection (Barlow and Levick, 1965) depends on veto-inhibition in the non-preferred motion direction. When a stimulus is moved across the detector's receptive field in the non-preferred direction, an inhibitory signal is propagated from the initially-stimulated region to the rest of the receptive field, such that any further stimulus across the receptive field does not elicit an excitatory response. When a stimulus moves in the preferred direction, inhibition is not propagated forward and continuous excitation of the detector is achieved. One of the consequences of this scheme is that stationary stimuli will result in excitation of the detector. Thus, although it is directionally selective in the sense of giving a differential response to motions in opposite directions, it is not sensitive *only* to motion, but also to static stimuli. For this reason, and a lack of elaboration with respect to a front-end (i.e. spatial filtering), the Barlow-Levick detector is not explored in depth in this dissertation.

2.2.4 Counterchange Detectors

As discussed briefly above, the concept of a counterchange mechanism was first proposed by Hock et al. (2002) in order to account for single element apparent motion, where only one visual element is seen to move. Again, a counterchange detector measures the (approximate) temporal coincidence of a decrease and increase in activation

at a pair of locations, with motion being signified from the location of decrease to the location of increase.

Hock et al. (1997) introduced the concept of *generalized apparent motion* (GAM), where two visual elements are simultaneously visible during each frame of the stimulus. In GAM, three luminance values are relevant for defining a stimulus, the luminance of the background, and the two luminances that are exchanged between the two visual elements on each frame change.² In GAM, a luminance change of a visual element toward the background luminance is presumed to elicit a decrease in local detector activation, and a move away from the background to elicit an increase in local detector activation.

Hock et al. (2002) developed a metric, the background relative luminance contrast (*BRLC*) which they found to be a strong predictor of human perception. The *BRLC* is calculated as the difference in the two element luminance levels ($L_1 - L_2$) divided by the difference between the average luminance level of the two elements (L_m) and the luminance of the background (L_b): $BRLC = (L_1 - L_2)/(L_m - L_b)$. In this framework, standard apparent motion (SAM) is considered a special case where L_2 (the lower luminance level of the elements) is equal to L_b ; in other words, when only one visual element is visible at a time. In this case, $BRLC = 2.0$, and motion is typically perceived. When *BRLC* is low (i.e. when the frame-to-frame luminance change of the elements is small), motion is typically not perceived.

Several experiments by Hock et al. (2002) also showed strong evidence that counterchange is the informational basis underlying the perception of apparent motion. They explicitly showed that the sequential order of changes in luminance did not determine the direction of perceived motion. That is, although motion is perceived from

²This distinction between GAM and standard apparent motion (SAM), where only one visual element is visible at a time is especially relevant with respect to the concept of *token-tracking*, which is discussed in context of the motion correspondence problem in Chapter 5.

the location of decreasing activation to the location of increasing activation, decreasing activation did not necessarily have to precede increasing activation. This calls into question accounts of motion perception that appeal to sequential changes at different locations in visual space as being the basis for the perception of motion. They also showed that certain conditions, for example the simultaneous asymmetrical increases in activation at two locations, did not elicit motion percepts, while spatiotemporal correlator models of motion perception (e.g. the elaborated Reichardt-detector) predicted that motion would be perceived.

While this formulation is useful in the case of single element AM, in some instances it is difficult to say a priori what ought to count as foreground and background. For example, drifting sine gratings, an extremely common stimulus in psychophysical studies of motion perception, are composed of (luminance) peaks and valleys, but it is not clear that there are any perceptual boundaries between background and foreground. Hock et al. (2009) nonetheless showed that the counterchange detector is able to detect motion in these conditions.

This problem of defining a background is also evident when it comes to dense random dot displays, where every pixel of the display is filled with either a white or black dot with a probability of 0.5, and there is no obvious figure-ground relation. This problem is addressed in Chapter 3, in which it is shown that it is sufficient to measure increases and decreases in the responses of half-wave rectified spatial filters in order to reliably detect counterchange, and no explicit figure-ground segregation is necessary a priori.

More detailed formulations of a minimal mathematical counterchange detector can be found in Appendix A, and continuous-time dynamical version is instantiated in the neural network described in Chapter 5 and Appendix C.

2.3 OPTIC FLOW AND MOTION ENERGY

As mentioned above, Gibson’s work with military pilots allowed him to recognize the importance of global optic flow patterns for guiding locomotion. What Gibson did not propose was a mechanism to account for this perceptual capacity. Thinking of motion perception as patterns of optic flow emphasized the need for low-level motion detectors tiling the visual field, whereas the perception of AM as in Wertheimer’s experiments could potentially be accounted for by a capacity-limited system that is able to track only a small number of features over time.

Following on Campbell and Robson’s (1968) work which introduced the concept of Fourier decomposition into visual perception of static images, several authors (e.g. Adelson and Bergen, 1985; van Santen and Sperling 1984, 1985; Watson and Ahumada, 1985) have contributed to the conception of motion perception as a process that can be understood in terms of *motion energy* extraction. To calculate the motion energy of a stimulus, one simply transforms the stimulus from the spatiotemporal domain to the frequency domain. In the presence of motion, the distribution of Fourier energy will be biased to certain regions of the space, and the direction (and velocity) can be estimated. However, motion energy by itself cannot explain perception, as it is not clear how the nervous system would accomplish the extraction of this information. Although several models have been constrained by this theoretical framing (e.g. the elaborated Reichardt-detector, van Santen and Sperling, 1985), at best they are only approximating the local Fourier energy and do not correspond to it in a one-to-one mapping.³ As such, the usefulness of the motion energy concept is not evident in the current effort. Instead, emphasis is laid on models of the *processes* of motion detection, rather than any abstract mathematical transformations of the raw

³This is shown explicitly in Appendix B for an elaborated Reichardt detector in response to two-frame polarity-inverted AM stimuli.

stimulus.

2.4 CONCLUSION

The study of motion perception has a rich history in perceptual psychology and continues to be an exciting area of inquiry. Modern technology has opened the door to both probe empirically into the neural tissue that supports perceptual processes in unprecedented ways, as well as build and test computational models whose complexity is beyond the scope of analytic techniques. Research on perception will continue to capture the imagination, as perception is intimately related to what it means to be a conscious agent in the mysterious world we find ourselves in.

Contrasting Accounts of Short-range Motion: Direction and Shape Perception in a Random-dot Cinematogram

This chapter is adapted from an article originally published as Norman, J., Hock, H., & Schöner, G. (2014). Contrasting accounts of direction and shape perception in short-range motion: Counterchange compared with motion energy detection. *Attention, Perception, & Psychophysics*, 1-21.

3.1 INTRODUCTION

In an ecological context, many organisms benefit from minimizing their visual profile via camouflage in order to remain undetected (Stevens and Merilaita, 2009). As a coevolutionary complement, organisms have been selected with visual systems that are, at least in some cases, able to overcome the challenges in detecting and segregating entities whose static visual cues are obscured by camouflage. One basis for the perceptual ‘breaking’ of camouflage entails the detection of coherent motion, which provides the opportunity to group portions of the visual field into connected wholes (as in the Gestalt principle of common fate) and to thereby segregate a moving entity

from its background in order to determine its shape¹from its motion. The short-range motion paradigm (Braddick, 1974), in which portions of a random field of elements are coherently displaced, provides a means for studying this ability to detect and segregate entities from their surrounding environment by virtue of their motion alone.

In the original 2-frame short-range motion paradigm (Figure 1), each square element of a random checkerboard has a 0.5 chance of being white (or black). A segment of the random checkerboard that is presented during the first frame is rigidly displaced and re-presented during the second frame (the coherent figure) while the surrounding elements are independently re-generated (the incoherent ground). Because the figure and background portions are generated in the same manner, the displaced random, incoherently moving background elements, and thereby determine its shape. As the size of the frame-to-frame displacement of the figure is increased, perceptual judgments become less consistent, with subjects reporting a loss in coherence of the moving figure (Braddick, 1974; Sato 1989).²In this article, psychophysical experiments and computational simulations investigate the motion mechanisms that are the basis, in the two-frame short-range motion paradigm, for the perception of motion, the conditions under which it is coherent enough to segregate a moving figure from its background, and the perception of the figure’s shape from the coherent motion.

Short-range motion perception has been considered a paradigmatic case for motion energy detection³(Cavanagh and Mather, 1989; van Santen and Sperling, 1985;

¹By shape we mean the ability to discriminate the orientation of the displaced figure. Although this does not put an explicit emphasis on the boundaries of the figure, they can be perceived at small displacements.

²The focus of this article is on the differential effects of figure displacement for same- vs. inverted-polarity conditions. D_{max} , a measure of the maximum displacement for which motion is perceived, is not determined.

³Rather than focusing on the features of the spacetime Fourier transform of the stimulus per se, our emphasis is on mechanisms proposed to detect Fourier-based motion energy, specifically the elaborated Reichardt detector.

Adelson and Bergen 1985; Marr and Ullman, 1981). A major feature of models of Fourier- based motion energy detection (Adelson and Bergen, 1985; van Santen and Sperling, 1985) is that they predict reverse-phi motion (Anstis, 1970). As shown in Appendix B, motion is predicted in the direction opposite to that of the displacement when the luminance polarity of the visual elements composing a stimulus is inverted

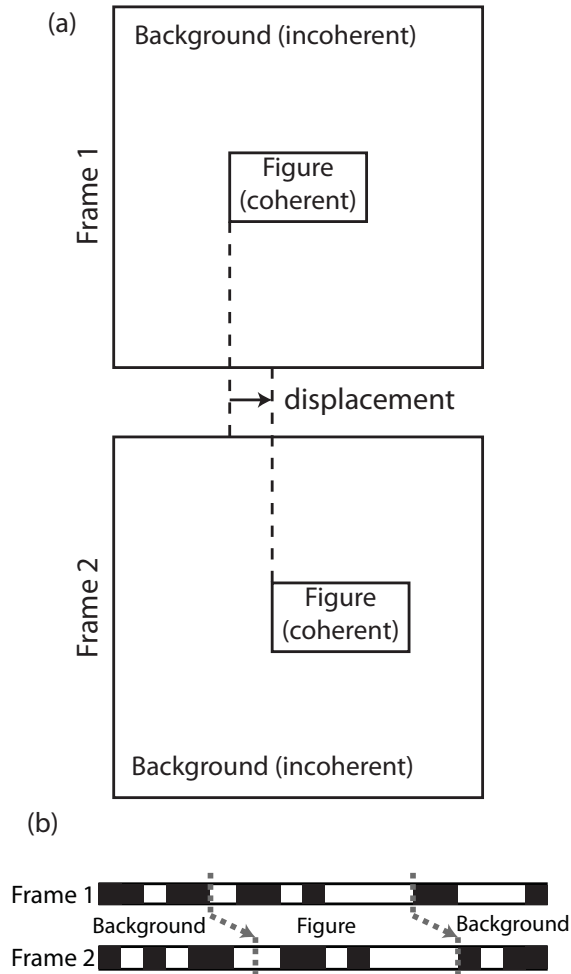


Figure 3.1: A sketch of the two-frame short-range motion stimulus. The figure region is coherently displaced (either left or right) from Frame 1 to Frame 2 while the incoherent dynamic background is updated randomly. (a) shows the layout of the 2-dimensional experimental stimulus, (b) shows a 1-dimensional slice of the random dot cinematogram (with fewer dots than in the experiment). The stimulus used for the simulations below is also of the 1-dimensional form depicted in (b).

between successive frames (i.e. white elements become black and black elements become white). The strength of this reverse-phi motion is identical to the strength of motion in the direction of displacement when luminance polarity remains the same. Consequently, empirical evidence for asymmetry in motion and shape perception between the same- and inverted-polarity stimuli would indicate that motion perception was not determined solely by motion energy detection.

Experimental results relevant to this determination have been reported by Sato (1989) who tested both direction of motion and shape discrimination with both same- and inverted-polarity versions of the short-range motion stimulus. Although he reported that direction discrimination was similar for the same- and inverted polarity stimuli, this symmetry was not consistently obtained in all his experiments. Whenever performance was below ceiling, direction discrimination was poorer for inverted-polarity stimuli.

Moreover, shape discrimination was severely deteriorated for the inverted-polarity stimuli, regardless of the size of the displacement. If these asymmetries were empirically confirmed, it would provide evidence that motion perception and the perception of shape from motion in the short-range paradigm is not primarily determined by 1st-order motion energy detectors. Instead, or in addition, an alternative motion detection mechanism that is sensitive to the difference between same- and inverted-polarity stimuli would be implicated. The alternative mechanism that is evaluated here entails the detection of counterchange; i.e., oppositely signed changes in activation for pairs of spatial filters at different spatial locations (Hock, Gilroy and Harnett, 2002; Hock, Schöner and Gilroy, 2009).

Because the symmetry, or lack thereof, of motion and shape perception in same- and inverted-polarity conditions is theoretically critical, the current study begins with a psychophysical experiment that re-evaluates and extends Sato's (1989) re-

sults. Computational simulations then determine how well the results obtained in the experiment are accounted for by Fourier-based 1st-order motion energy detection (van Santen and Sperling’s [1985] elaborated Reichardt detector, which is based on Reichardt’s [1961] motion detection model) compared with the non-Fourier detection of counterchange (Hock et al., 2009). For both models, investigating shape judgments in addition to motion direction judgments requires addressing the spatial arrangement of motion detectors in addition to their internal structure.

3.2 EXPERIMENT

The results of Sato’s (1989) third experiment came closest to providing evidence for symmetry in direction discrimination for standard (same-polarity) and reverse-phi (inverted-polarity) motion. The possibility that this was due to ceiling effects for highly practiced observers was suggested by the lack of symmetry in his first two experiments, which used the same, though presumably less practiced observers. In addition, in Sato’s second experiment, the advantage in direction discrimination for standard motion compared with reverse-phi motion became more pronounced when reducing the size of the elements lowered discrimination performance from ceiling.

The experiment closely resembles Sato’s (1989) third experiment, in which participants indicated both the direction of motion and the shape of the displaced. In order to reduce the possibility of ceiling effects, testing was done primarily with naive participants who received minimal practice at the task and no feedback regarding the accuracy of their discriminations.

3.2.1 Method

Stimuli

The dynamic random checkerboard stimuli, which were generated with a Mac Mini computer, were centered in a Mitsubishi Diamond Pro 930SG monitor and viewed in a dimly lit room from a distance of 58 cm (maintained by a chin rest). As in Sato (1989), the stimuli were composed of two frames, each with a random checkerboard composed of 120x120 square elements that was presented against a black background. Each square element composing the checkerboards subtended a visual angle of 2x2 min (one pixel per check), and the entire checkerboard subtended a visual angle of 4x4 deg. The luminance of the white elements was 76.6 cd/m² and that of the black elements was 0.0 cd/m².

The first frame of each 2-frame trial was generated by independently assigning each square element of the checkerboard to be either white or black with a 0.5 chance of each. During the second frame, a region (the figure) was selected from the center of the first frame and displaced by either 2, 4, 6, 8, 10, 12, 14 or 16 element-units (4 to 32 min) to the right or left. The rest of the checkerboard (the background) was randomly re-generated, again with a 0.5 probability of each element being white or black. The figure was either a vertically oriented rectangle (60x30 element-units; 120x60 min) or a horizontally oriented rectangle (30x60 element-units; 60x120 min). In the same-polarity condition, the luminance of the square elements composing the displaced figure was the same during both frames. In the inverted-polarity condition, the luminance of the square elements composing the displaced figure was inverted during the second frame; white elements became black and vice versa.

Procedure

To familiarize participants with the task, a version of the random checkerboard stimulus was shown in which all but the left-most and right-most two columns of elements from the entire 120x120 field of elements constituted the figure, which was displaced rightward or leftward by two element-widths (i.e. there was not an incoherent background from which coherent motion had to be segregated). In order to maintain the size of the field for the second frame, the two columns at the leading edge of the figure were removed rather than displaced, and the trailing two columns were randomly re-generated. This was done for both same- and inverted-polarity versions. Participants viewed these demos without feedback for approximately 5 min, until they indicated that they were able to perceive both leftward and rightward motions. Shape discrimination was then explained by means of drawings of the tall-thin and short-wide rectangles, and a demo stimulus composed of ten 138 msec frames, with 2-element displacements during each frame (without polarity change). The figure shapes were easily discernible for this demo. A similar shape demo was not provided for the inverted- polarity condition as it did not make the shapes discriminable and so did not aid in describing the task. Participants other than the first author received no practice with what would become the test stimuli.

As in Sato (1989), each test trial began with the participant fixating in the center of a 8x8 min square arrangement of four 2x2 min white dots, which was presented for 0.5 sec against a black background. This was followed by a blank black screen for 0.5 sec, then the two stimulus frames were presented for 138 msec each, and finally, another blank black screen. After each trial, the participant made two two-alternative forced- choice responses by pressing keys on the computer keyboard to indicate: 1) the direction in which the figure was displaced (either right or left), and 2) the shape

of the displaced figure (either a vertically or horizontally oriented rectangle). There was no feedback.

Design

Blocks of 128 test trials were generated by the orthogonal combination of 2 displacement directions, 8 displacement distances, 2 figure orientations, and 4 repetitions. Order was randomized within sub-blocks of 32 trials. The same- and inverted-polarity stimuli were tested in alternating blocks of trials. Each participant was tested for 7 blocks of trials for each polarity condition for a total of 14 blocks of trials. Participants. In addition to the first author, three students from Florida Atlantic University voluntarily participated in this experiment. They were naive with respect to its purpose. All participants had normal or corrected-to-normal vision.

3.2.2 Results

The results for each of the four participants are presented in Figure 2. Direction discrimination is graphed with respect to the actual figure displacement, regardless of the polarity condition. Thus, reverse-phi perception is indicated by responses which are systematically in the opposite direction of the displacement, and therefore below chance level (i.e. below 0.5). As in Sato (1989), both direction and shape discrimination decreased with increasing displacement of the rectangular figure, with shape discrimination falling to chance at smaller displacements compared with direction discrimination. Most importantly, the results for each of the four participants indicated a clear asymmetry in both direction and shape discrimination between the same- and inverted-polarity conditions; both were superior in the same-polarity condition.

A two-way repeated measures ANOVA performed on the arcsine transformed proportion data indicated that the effects on direction discrimination of displacement

size, $F(7, 21) = 55.74, p < .001$, luminance polarity (same or inverted), $F(1, 3) = 29.25, p < .05$, and the interaction between polarity and displacement size, $F(7, 21) = 14.37, p < .01$ all were statistically significant. (In the inverted-polarity condition, responses in the reverse phi direction were treated as correct, so the complements of the proportion correct responses were used in the ANOVA.) For shape discrimination, the effect of displacement size, $F(7, 21) = 11.93, p < .001$, and the interaction of polarity with displacement size was statistically significant, $F(7, 21) = 5.85, p < .01$. For each participant, shape discrimination was better in the same- than the inverted-polarity condition for the small displacements, but because of floor effects and the small sample size, the effect of polarity fell short of statistical significance, $F(1, 3) = 7.77, p = .069$.

Because there was a consistent trend of shape discrimination being better in the same-polarity condition for all participants, especially evident at the smallest displacement of 2-elements, a log-likelihood ratio test was performed for each participant as well as their pooled scores to evaluate the null hypothesis that the probability correct was identical in the two contrast conditions. That is, let pS (pD) be the proportion correct in the same-polarity (inverted-polarity) condition and p be the pooled proportion correct across both conditions, then the null hypothesis is $pS = pD = p$. If $kS(kD)$ is the number of correct responses in the same- (inverted-) contrast condition and $nS(nD)$ is the number of incorrect responses in the same- (inverted-) polarity condition, then the likelihood for the unconstrained model can be expressed

$$\text{Log}LU = kS\log(pS) + nS\log(1 - pS) + kD\log(pD) + nD\log(1 - pD) \quad (3.1)$$

and the constrained model as

$$\text{Log}LC = (kS + kD)\log(p) + (nS + nD)\log(1 - p). \quad (3.2)$$

Then under the null hypothesis $pS = pD = p$ the test statistic

$$X = 2(\text{Log}LU - \text{Log}LC) \quad (3.3)$$

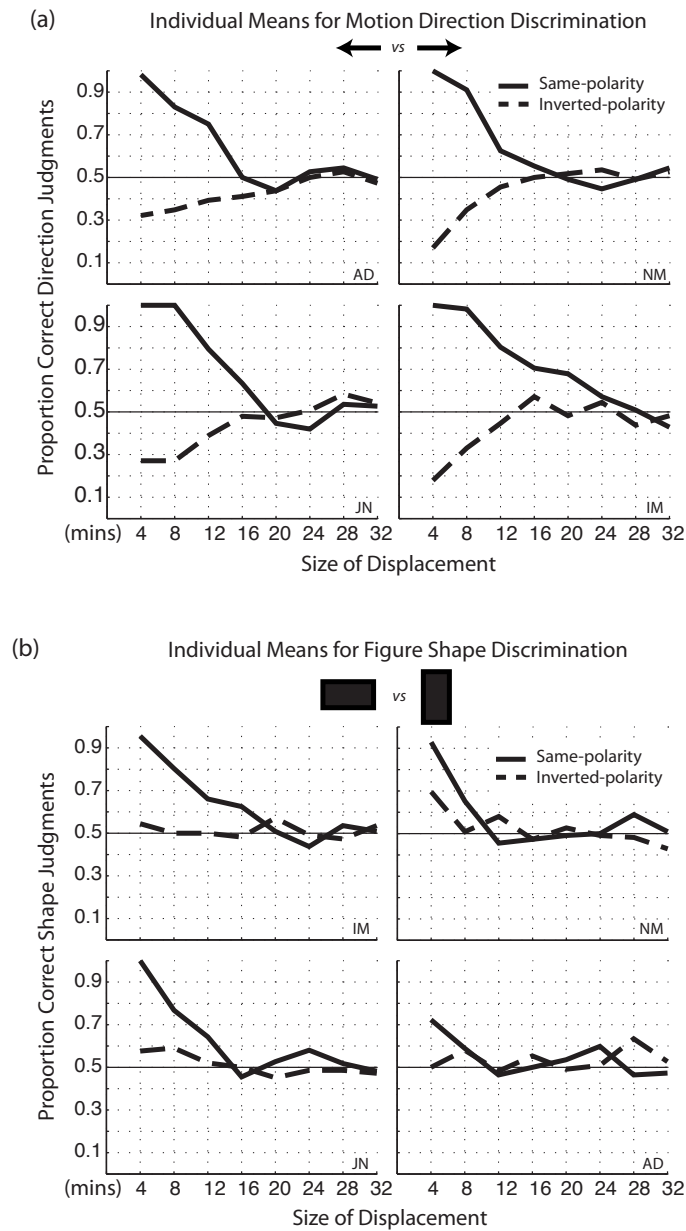


Figure 3.2: Mean experimental results for individuals for (a) direction judgments (left or right) and (b) shape judgments (wide or tall rectangle). Proportion of correct responses are plotted as a function of figure displacement in dot-units. Solid lines indicate the ‘same-polarity’ condition, and dashed lines indicate ‘inverted-polarity’ condition. Data points in (a) that are below chance (0.5) indicate a systematic bias to see motion in the direction opposite to displacement (reverse-phi).

is asymptotically distributed as chi-square with $df = 1$ (degrees of freedom determined by the number of free parameters in the constrained models subtracted from the number of free parameters in the unconstrained model). For each individual and for the pooled scores, the constrained (null) model was rejected in favor of the unconstrained model with $p < .001$ (with the greatest individual p-value = 4.4728×10^{-6} ; individual chi-square values = 56.69, 57.28, 71.86, 21.05; pooled chi-square value = 123.39). These results suggest that the probability of a correct response in the same-polarity condition was significantly different than the probability of a correct response in the inverted- polarity condition at the displacement of 2-elements, for each participant individually and for their pooled responses.

If the effects on direction and shape discrimination were symmetrical, there would have been neither differences between the same- and inverted-polarity conditions nor significant interactions with the size of the figure displacement. Further, the likelihood ratio test would have indicated no difference between the probability of a correct shape response in the same- and inverted-polarity conditions. The results indicate that this was not the case.

3.3 COMPUTATIONAL SIMULATIONS

Computational implementations of van Santen and Sperling's (1985) elaborated Reichardt detector (ERD) and Hock, Schner and Gilroy's (2009) counterchange detector, which are detailed in Appendix A, were compared with respect to their ability to simulate the results of the experiment described above. For the purpose of these simulations, the two-dimensional random checkerboard stimuli were reduced to one-dimensional vertical bars whose luminance, white or black, was randomly determined, as was done by van Santen and Sperling (1985), Adelson and Bergen (1985), and Sato

(1989). Consistent with the stimuli in the experiment described above, a portion of the random-bar stimulus was rigidly translated from the first frame to the second (the ‘figure’) while the rest of the stimulus (the ‘background’) was randomly generated in both the first and second frames. The stimulus was 240 bars long in the simulations. There were two figure lengths, analogous to the two figure shapes in Experiment 1: a figure that was 60 bars long represented the thin-tall rectangle, and a figure that was 120 bars long represented the wide-short rectangle. In the inverted-polarity condition, bars within the figure that were white during the first frame were black during the second frame and vice versa. The figure was displaced by 2, 4, 6, 8, 10, 12, 14, or 16 bar-widths, the same displacements that were probed in the experiment. The ‘random bars’ provided the input stimulus to the motion detector ensembles.

Coincidence detection and directional selectivity

Both models use the multiplication of activity patterns in pairs of spatially separated, one-dimensionalized edge filters (an excitatory zone and an adjacent inhibitory zone) to establish a correspondence between them⁴. However, the nature of the patterns whose coincidence is detected is different in the two models.

The ERD is sensitive to sequential changes in edge-filter activation; i.e., instantaneous edge-filter outputs are compared at different points in time. This is achieved by delaying the output of one edge filter in order to temporally align activation that occurs at its location at one moment in time with the pattern of activation at a paired location at a later moment in time so that the patterns can be compared. At the level of the subunits where multiplication occurs (before the difference between the two subunits is taken), positive products signal motion from the location of the edge filter

⁴The scale of the edge filters for the ERD was determined by the quadrature constraint of the model. The edge filters for the counterchange model were selected to be most responsive to the size of the checks in the checkerboard stimulus.

whose activity has been delayed to the location of the edge filter whose activation has not been delayed, while negative products signal motion in the direction from the location of the non-delayed edge filter to the delayed one. ⁵

Although temporal coincidence is also central to the counterchange motion detector, a temporal delay is not required in order for it to be directionally selective. This is because the counterchange detector is sensitive to a particular pattern of simultaneous changes in edge filter activation: a decrease in the activation of one edge filter and a simultaneous increase in the activation of a paired edge filter. Rather than deriving a directional asymmetry from sequentiality, as in the ERD, an asymmetry in the direction of activational change in local spatial filters is established, with motion beginning from a location of a decrease in spatial filter activation and ending at a location of an increase in spatial filter activation. This is irrespective of the sequential order of the stimulus events producing the decreases and increases in activation (Gilroy and Hock, 2009; Hock et al., 2009).

Edge filter polarity

In the ERD model, the multiplication of instantaneous outputs of the paired edge filters occurs irrespective of whether they are positive (excited) or negative (inhibited). On this basis, it is sufficient to have only one edge filter polarity for the ERD model (e.g., excitatory zone on the left, inhibitory zone to its right) as the entire range of positive and negative edge filter outputs take part in motion computation. In other words, both edge-types are represented, one by positive values and the other by negative values. For example, if more white elements fall in the positive lobe

⁵Typically, Reichardt-type detectors are described as detecting motion in the direction from the delayed input toward the non-delayed input. This, however, is not strictly true in the ERD formulation, as each subunit may carry information about two (opposite) motion directions (Adelson and Bergen, 1985; Lu and Sperling, 2001).

than in the negative lobe of an edge filter during Frame 1 (positive response), and more white elements also fall in the positive lobe of a paired edge filter during Frame 2 (another positive response), the product of the two positive responses is positive. Further, if more white elements fall in the negative lobe than the positive lobe of the same edge detector during Frame 1 (negative response), and more white elements fall in the negative lobe of the edge filter with which it is paired during Frame 2 (another negative response), the product of the negative responses is also positive for the ERD. Thus, nothing would be added to the computations by including edge filters with reversed positive/negative polarity. It also is noteworthy that if a negative edge-filter response in Frame 1 is multiplied with a positive response in Frame 2 (or vice versa), a negative response is elicited, indicating motion in the opposite direction than that of a positive response. Importantly, this is the basis for the ERD model signifying motion in the reverse-phi direction (although negative-valued products are also produced with non-inverted- polarity stimuli). These edge-filter products occur at the level of the ERD subunits, from which the difference is taken to determine the final motion-detector output. In contrast, in the counterchange model the activation values of edge filters are half-wave rectified, so only positive outputs are subject to the subsequent change- detection that leads to motion detection. This is in line with the principle of counterchange motion pairing ‘like’ edges, detecting their disappearance at one location and appearance at another location (this is discussed in more detail in the discussion). For this reason, the model includes two edge-filter polarities. The filter with its excitatory zone on its left side captures inputs in which there are more white elements falling on the filter’s left side, whereas the filter with the excitatory zone on the right captures inputs in which there are more white elements falling on the filter’s right side. The two edge-filter polarities compute motion in parallel.

Opponency

The ERD is an opponent system; it takes a difference between its two component subunits for its final output. Each subunit can carry information about both leftward and rightward motion because they each can have negative or positive values. Taking the difference between the subunits gives the final motion output. Net positive outputs signal motion in one direction (i.e. rightward) and net negative outputs in the opposite direction (i.e. leftward). Furthermore, opponency is necessary to prevent the ERD from signaling motion in response to stationary patterns. For purposes of comparing the two models, the counterchange model was arranged in a similar opponent fashion with leftward motion signals being subtracted from rightward signals. This is not a necessity for the counterchange model because unlike the ERD, leftward and rightward motion signals are separable, and motion cannot be signaled for stationary stimuli. Therefore, by convention, rightward motion is represented in both models by positive values and leftward motion by negative values at each location along the detector arrays.

Spatial arrangement of motion detector arrays

For both ERD and counterchange motion detectors, the distance between the centers of the pair of edge filters that provide input to each motion detector is referred to as that detector's span. (This is illustrated in Figure 3, which shows the general layout of both the ERD and counterchange detectors.) Both models included arrays of detectors with spans of 2, 4, 6 and 8 bar-widths. Within each array, the detectors densely covered the entire stimulus. Edge filters that served as input to the motion detectors were located every $\frac{1}{4}$ of a bar width across both the displaced figure and its background. Following van Santen and Sperling's (1985), there were multiple layers

of motion detectors, each layer corresponding to a particular span. In the current simulations this meant that there were four layers. Direction-discrimination. In order to simulate the direction discrimination task, for each trial all motion signals were summed across space and across layers and the sign of the sum indicated the motion direction decision (as rightward motions were positive and leftward motions were negative). Within each layer, responses were summed across all motion detectors covering the 240 random bars constituting the entire stimulus (not just the 60 or 120 random bars corresponding to the displaced figure). Summing activation over the entire field of random bars was significant because it meant that motion direction

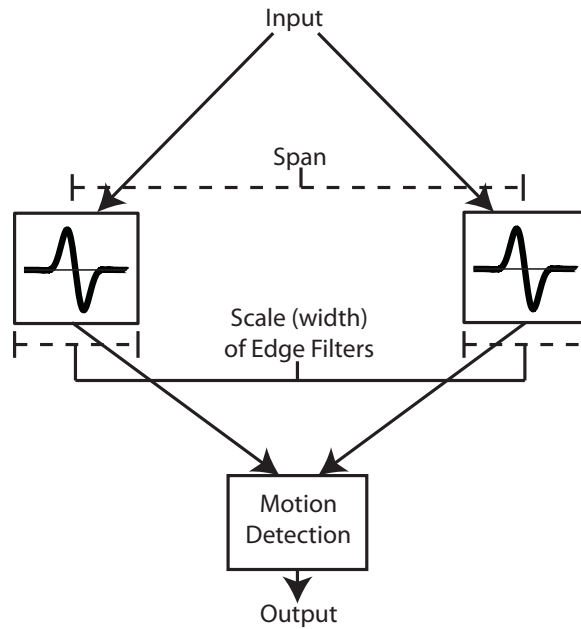


Figure 3.3: General layout of both motion detectors. A pair of edge filters separated in space serve as inputs to subsequent motion detection, the distance between the center of their receptive fields is referred to as the detector's span. For the ERD, the size of the span and the scale (width) of the edge filter co-vary in order to maintain an approximate quadrature relationship (i.e. so there is approximately a 90-degree phase shift with respect to their preferred spatial frequency). The counterchange detector has no such constraint, and in the current model, the scale of the edge filters is held constant over a range of spans. For both models, detectors are arranged in layers, and each layer corresponds to a specific span.

was being discriminated by the models without pre-determination of the shape of the figure. That is, figure segregation was not considered a prerequisite for direction detection. This is consistent with the shape of the figure being derived from the motion rather than vice versa. Motion detector responses were also summed across all layers (spans). That is, all spans contributed equally to the determination of motion direction. This implies that direction discrimination does not depend on motion signals being concentrated at a particular span or in a particular image location.

For each trial, therefore, a positive sum (the positive component is greater than the negative component) signifies rightward motion perception, whereas a negative sum signifies leftward motion perception. In this way, both models make the same kind of forced-choice responses as the participants in the actual experiments. The proportion of trials that motion perception was signified in the direction of the displacement was determined for 224 repetitions (matching the aggregated number of experimental trials for the four participants in the experiment). Proportions in the direction of the displacement that were less than 0.5 indicated that a majority of the simulated responses were in the so-called reverse-phi direction.

Shape-discrimination. The ability of participants in the experiment to discriminate the shape of the displaced figure indicates that the detected motion could be used to segregate the figure from its background and determine its shape. This was simulated for both the ERD and counterchange models with templates that corresponded to the width of the two figures. The two templates functioned as filters whose inputs were the spatial distribution of motion signals along the stimulus array.

The simulations for the experiment were based on two principles of coherent motion supporting the perception of shape-from-motion. Accordingly, coherent motion arises from regions of activated motion detectors that: 1) are in the same direction, and 2) are of the same span. A high density of such signals within a template's

positive area compared with its negatively weighted flanking regions would result in a positive template output. The same-span constraint on motion coherence was consistent with the two-dimensional percepts elicited by the rigidly translating figures in the experiment. (The possibility of relaxing this constraint to account for recovery of depth information is addressed in the discussion.) One template was composed of a positive interior region matching the relatively short one-dimensional size of one figure (60 bar-widths) and another template was composed of a positive interior region matching the relatively long one-dimensional size of the other figure (120 bar-widths). All the detected motions within the figure region were summed with equal positive weight. Negative regions flanking the positive interior regions extended to the boundaries of the random-bar stimulus, which was 240 bar widths in length. All detected motions within the flanking regions were summed with equal negative weight. The templates were normalized such that their positive interior region integrated to 1 and their negative exterior regions integrated to -1. For each trial, the output of each template was determined for each direction (leftward and rightward) and for each of the four spans. The figure size (either long or short) with the greatest template response was taken as the shape decision for a trial. (As in the experiment, shape-discrimination required forced-choice decisions by the models.)

3.3.1 Simulations Based on the Elaborated Reichardt Detector

A diagram of the ERD can be seen in Figure 4a. As in van Santen and Sperling's (1985) ERD model, the edge filters in its current implementation model are band-pass. Space-time filters in the Fourier domain are approximated by establishing a quadrature relationship between pairs of filters constituting a motion detector. Thus, pairs of edgefilters, implemented as one-dimensional real-valued Gabor filters, are modulated by sine waves that are 90-degrees out of phase with on another. Larger

spatial filters are therefore required to approximate the quadrature relationship among motion detectors whose component receptive-field centers are further apart (i.e., have larger spans).

Results

Single trial simulations. As indicated above, rightward motion was signified by positive values and leftward motion by negative values. In the single trials presented in Figure 5, the displacement of the figure is to the right. When the figure's displacement is small (e.g. 2 bar-units rightward, Figure 5a), much of the activity is concentrated within the figure at the span that corresponds to the actual displacement, with most motion signals in the correct direction (rightward). In the background regions there is also a fair amount of activity, though weaker on average and directionally incoherent, as would be expected for responses that are driven by noise. At larger spans, directional responses are generally consistent with the actual displacement direction within the figure region, but are spread across several spans for all displacement sizes, with

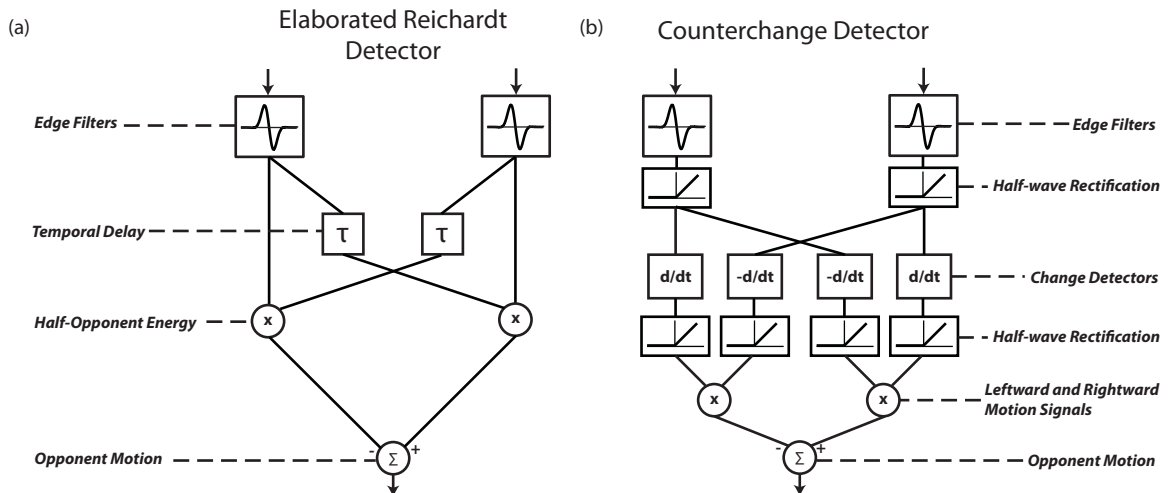


Figure 3.4: Block diagrams of the (a) elaborated Reichardt detector and (b) counterchange detector. Only one polarity channel of the counterchange detector is shown here, the other one operates in parallel.

the average strength of the response decreasing with greater spans. This weakening of the response is a consequence of the larger spatial filters required by larger-span detectors due to the ERD's quadrature constraint. When the figure's displacement is larger (e.g. 6 bar-units, Figure 5b), the span corresponding to the displacement shows a directionally consistent but relatively weak response within the figure region. The responses of nearby spans also are directionally consistent within the figure, and with similar strength. Therefore the directional motion information for the figure region is again spread across several spans for all displacement sizes. Furthermore, small-span detectors that are driven almost entirely by noise respond strongly due to their filters responding more strongly to the spatial structure of the stimulus.

Regardless of the size of the displacement, symmetrically opposite results were indicated for the inverted-polarity conditions when the second frame is the exact inverse of the second frame in the same-polarity condition. Motion was most often signaled in the leftward, reverse-phi direction within the figure, with the same strength and spatial distribution across all locations and spans, both within the figure and the background, as in the same-polarity condition (Figure 5a, dashed curve).

Simulation of experimental results. ERD-determined simulations of direction and shape discrimination in the short-range motion paradigm are presented in Figures 6a and 6b, along with the averaged results for the four participants in the experiment. It can be seen that the ERD successfully simulated the effect of displacement size; direction and shape discrimination were poorer for the larger displacements.

The ERD also simulates the perception of reverse-phi motion in the inverted-polarity condition, but incorrectly predicts that it is quantitatively equal to motion in the direction of the displacement in the same-polarity condition; in the experiment, both direction discrimination and shape discrimination were significantly poorer for motion in the reverse-phi direction. It could be concluded, because of its inherent

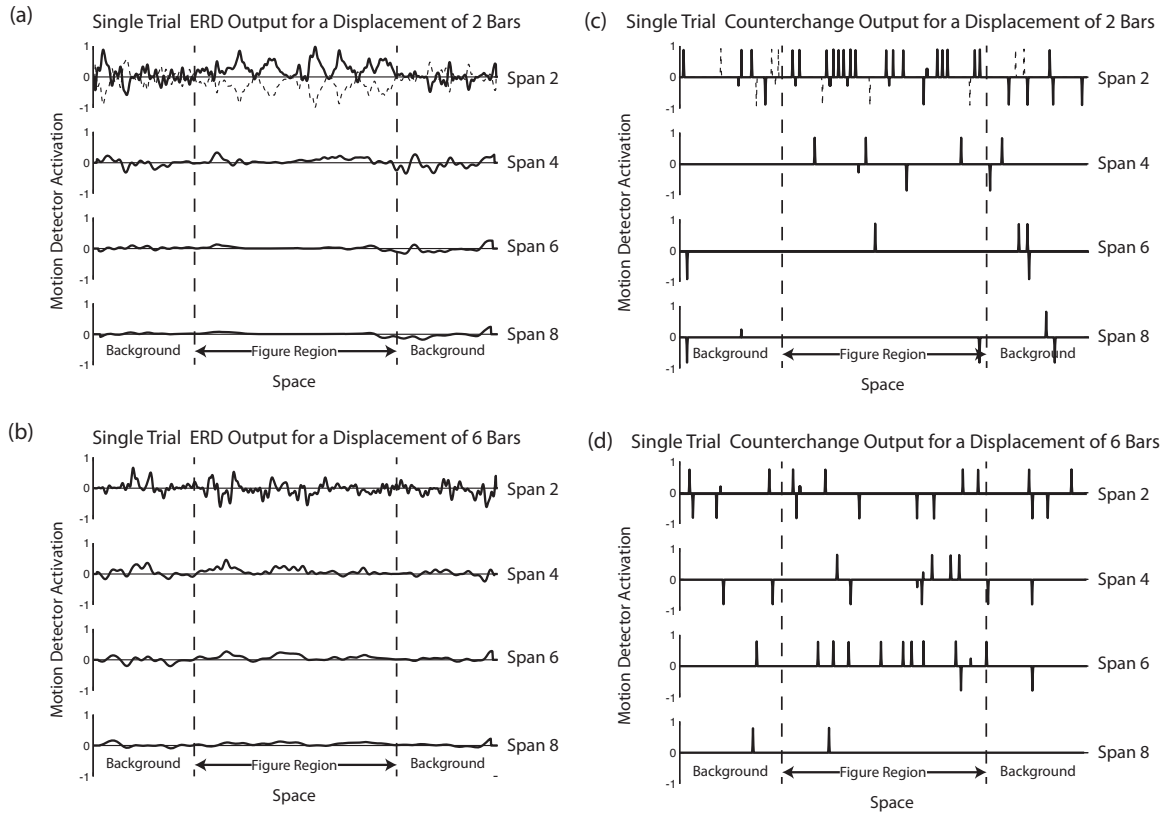


Figure 3.5: Figure 5 - Single trial simulation outputs of the ERD (a & b) and counterchange detector (c & d). (a & c) show a rightward displacement of the figure by 2 bar-units, and (b & d) show a rightward displacement of 6 bar-units. Solid curves represent the local motion detector output across space for each of 4 layers of motion detectors with various spans. Activations above 0 signal rightward motion, and activations below 0 signal leftward motion. The figure occupies the region between the dashed vertical lines, and the flanking background regions fall outside of it. The dashed curve in the first detection-layer in panels (a & c) depicts the response to the inverted-polarity version of the same stimulus. Note the ERD's symmetry around 0 with respect to the same-polarity stimulus (reverse- ϕ). Although not depicted, the same symmetry is obtained for all the span layers of the ERD. Also noteworthy is the indication that ERD activation is spread across span layers rather than being concentrated at the span corresponding to the displacement, particularly for larger displacements. In contrast, the inverted-polarity condition does not elicit a symmetrical response from the counterchange detector. This is true at all span-layers, despite the dashed curve only being shown for the smallest span-layer in (c).

symmetry with respect to the same- and inverted-polarity stimuli, that the detection of 1st-order motion energy by the ERD is not sufficient in order to account for short-range motion perception.

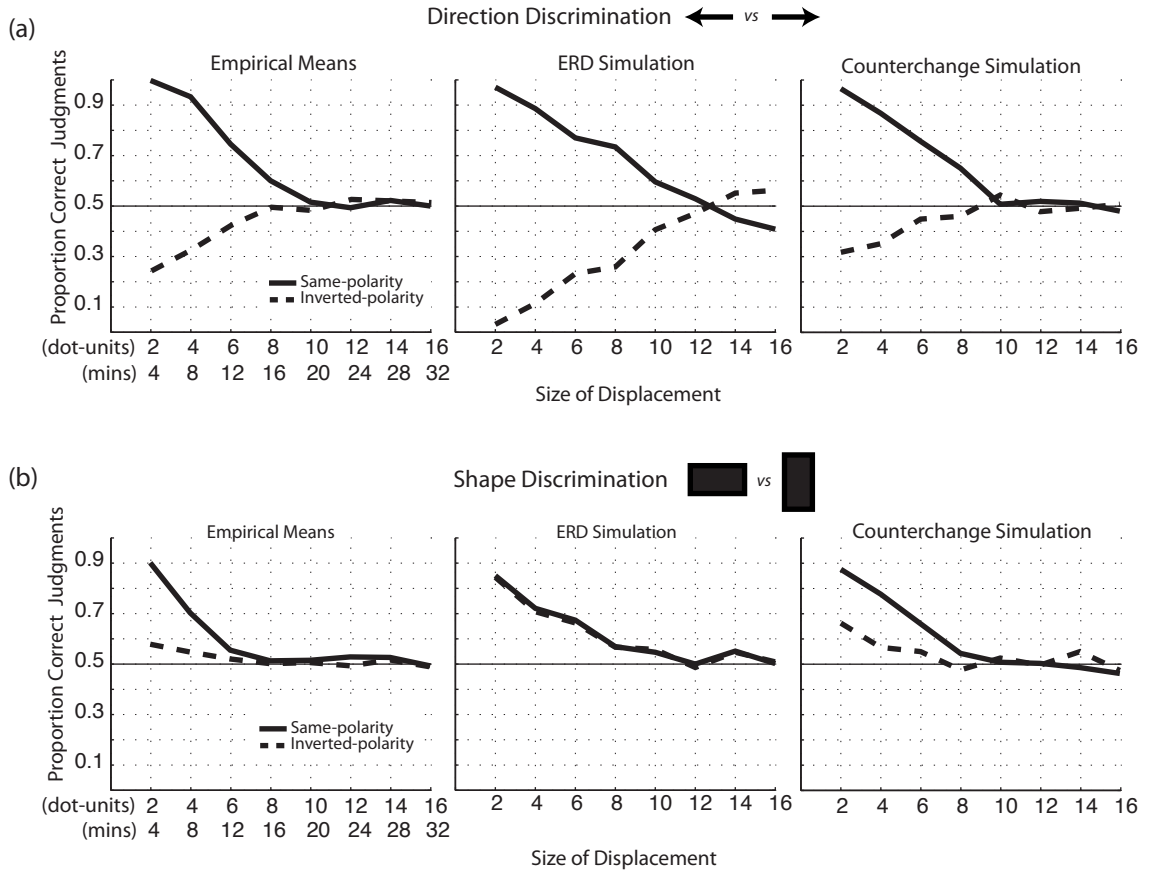


Figure 3.6: Results from the experimental simulations alongside the empirical means for (a) the direction discrimination task (left or right), and (b) the shape discrimination task (wide or tall rectangle). Solid curves represent mean scores from the same-polarity conditions, dashed curves represent mean scores from the inverted-polarity condition. Because of symmetry in its response to the same- and inverted-polarity stimuli, the ERD overestimates performance in the inverted-polarity condition for both direction judgments (corresponding to reverse-phi percepts) and shape judgments. The Counterchange detector is very similar to the empirical data both qualitatively and quantitatively. The empirical asymmetry between same- and inverted-polarity percepts as evidenced in both direction and shape judgments is clearly evident in the counterchange simulation.

Second-order motion energy

Also considered was the possibility that the perception of motion and shape entails 2nd-order motion energy extraction (Lu and Sperling, 2001). Full-wave rectification of the edge filters' activation in the 2nd-order system would make all negative activation values positive, so inverting luminance-polarity would result in the output of the edge filters being the same as in the same- polarity condition. The simulation of 2nd-order motion energy therefore would result in motion perception being signified in the direction of the displacement, regardless of whether or not the luminance polarity of the elements is inverted during the second frame of the two-frame trials. Reverse-phi motion percepts would not be predicted.

The effect of contrast

van Santen and Sperling (1984, 1985) have reported that their experimental support for the ERD as the basis for motion perception was obtained only for low contrast gratings. They argued that the perceptual invariance of suprathreshold motion is evidence of motion detectors' early saturation. It might be argued, therefore, that our empirical evidence, which was contrary to the predictions of the ERD, might have been due to testing short-range motion perception with high contrast (black and white) elements. An experiment was therefore conducted in order to determine whether the ERD's prediction of symmetry with respect to the effect of same- vs. inverted-polarity would be obtained at very low (barely visible) contrast levels. The results, which are presented in Figure 7a, are very similar to those obtained in the primary experiment. That is, both better direction and shape discrimination were obtained for the same-polarity than the inverted-polarity stimuli. A likelihood ratio test of the same form used to analyze the results of shape discrimination in the

primary experiment was here used to test the significance of the difference in both direction and shape perception at the smallest displacement of 2 dot-units. For the direction discrimination task, the chi-square value was 90.18, $p < .001$; for the shape discrimination task, the chi-square value was 319.81, $p < .001$. Symmetry with respect to luminance polarity was not obtained for low contrast short-range motion stimuli, which might have been expected on the basis of van Santen and Sperling's (1984, 1985) evidence that the ERD functions properly only for low contrast motion stimuli.

The effect of frame-rate

Another possibility is that the ERD functions properly only for fast frame rates that more closely approximate continuous motion, so the lack of symmetry found in the main experiment may have been due to the relatively slow frame rate of the stimulus (138 ms/frame). A variant of the experiment was run with much faster frame rates (35 ms/frame). The same likelihood ratio test as in the low-contrast variant above was again run. For the direction discrimination task, the chi-square value was 69.29, $p < .001$; for the shape discrimination task, the chi-square value was 483.28, $p < .001$. Again, asymmetry with respect to polarity-inversion was found for both direction and shape discriminations (Figure 7b).

3.3.2 Simulations Based on the Counterchange Motion Detector

A diagram of the counterchange detector can be seen in Figure 4b. The counterchange motion detector is sensitive to simultaneous and oppositely-signed changes in activation for pairs of spatial filters at separate locations (Hock et al. 2009), motion being signaled from the location of the decrease to the location of the increase in activation. Decrease subunits respond with excitation to decreases in their activational input and increase subunits respond with excitation to increases in their activational input.

Counterchange-determined motion is indicated when the product of the ‘decrease’ and ‘increase’ excitation is greater than zero.

Although the perception of short-range motion has typically been attributed to the detection of motion energy (e.g. Adelson and Bergen, 1985; van Santen and Sperling, 1985; Cavanagh and Mather, 1989), it was shown by Hock et al. (2009) that it could plausibly be accounted for by the detection of counterchanging activation. Their account, which is recapitulated below, was based on the distribution of excitatory and inhibitory effects on spatial filters by the randomly arranged white and black

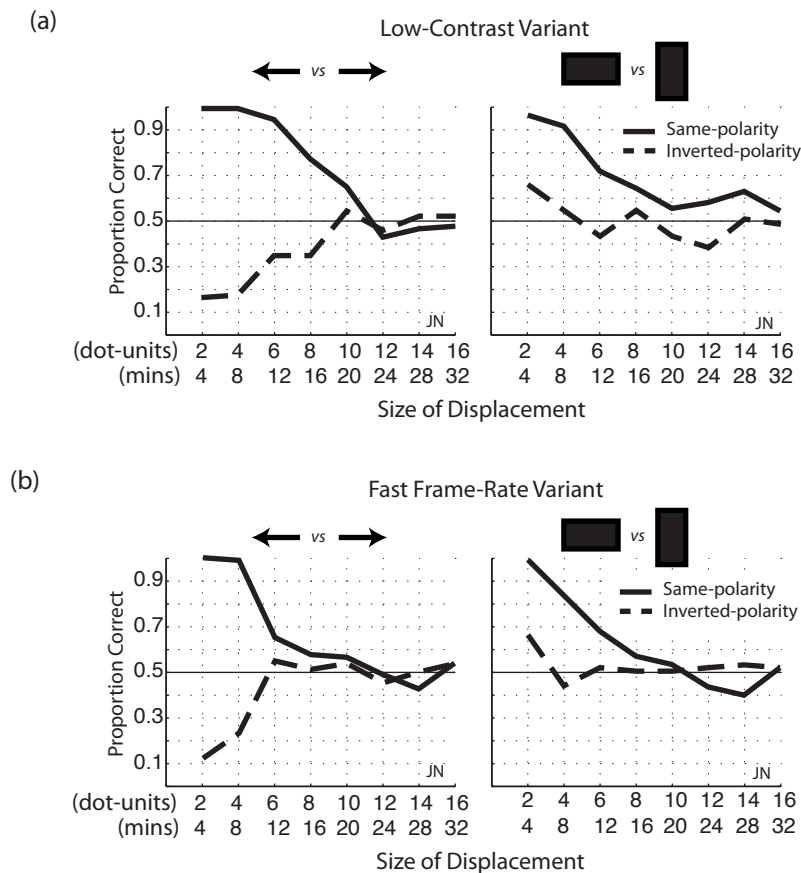


Figure 3.7: Two variants of the experiment in order to test the effects of (a) low-contrast and (b) fast frame-rates (35 ms) on the empirical asymmetry. Both conditions show the same asymmetry as the main experiment in both direction and shape judgments.

elements constituting the short- range motion stimulus (Figure 8).

Among the many edge filters that are stimulated by the figural portion of a random checkerboard, there are some that are (by chance) positively activated during the first frame of each two-frame trial (Figure 8b). When the figure is displaced to a new location during Frame 2, the filters that were excited during Frame 1 will be stimulated by a distribution of elements that is more likely to produce a decrease than an increase in activation. (It is illustrated in Figure 8a that there is a greater range of possible excitation and inhibition levels that would lead to decreases compared with increases in activation.) At the same time, the elements of the figure that had produced an excitatory effect on an edge filter during Frame 1 are exactly displaced to a new location during Frame 2, where they will produce similar activation of another, paired edge filter with the same excitatory/inhibitory polarity. It is likely that this filter was more weakly activated during Frame 1, so its activation is likely to increase. A counterchange motion detector spanning these two locations within the figure will be activated by the multiplicative combination of decreased activation at one edge filter location and increased activation at another edge filter location. There is no constraint for the non-Fourier counterchange model that requires a quadrature relationship between the sizes of the edge filters and their span, so the size of the edge filters was the same for all spans. As indicated earlier, the outputs of the edge filters are half-wave rectified, so only positive activation levels are passed forward. Likewise, the outputs of the decrease and increase detectors are half-wave rectified before they are multiplied to yield a directionally-selective motion computation. The reasons for the inclusion of half-wave rectification after each stage of processing are twofold: for reasons of neural plausibility and for conceptual soundness of the counterchange principle. These issues are addressed in more detail in the discussion.

In order to detect the motion of both white-black and black-white edges, two

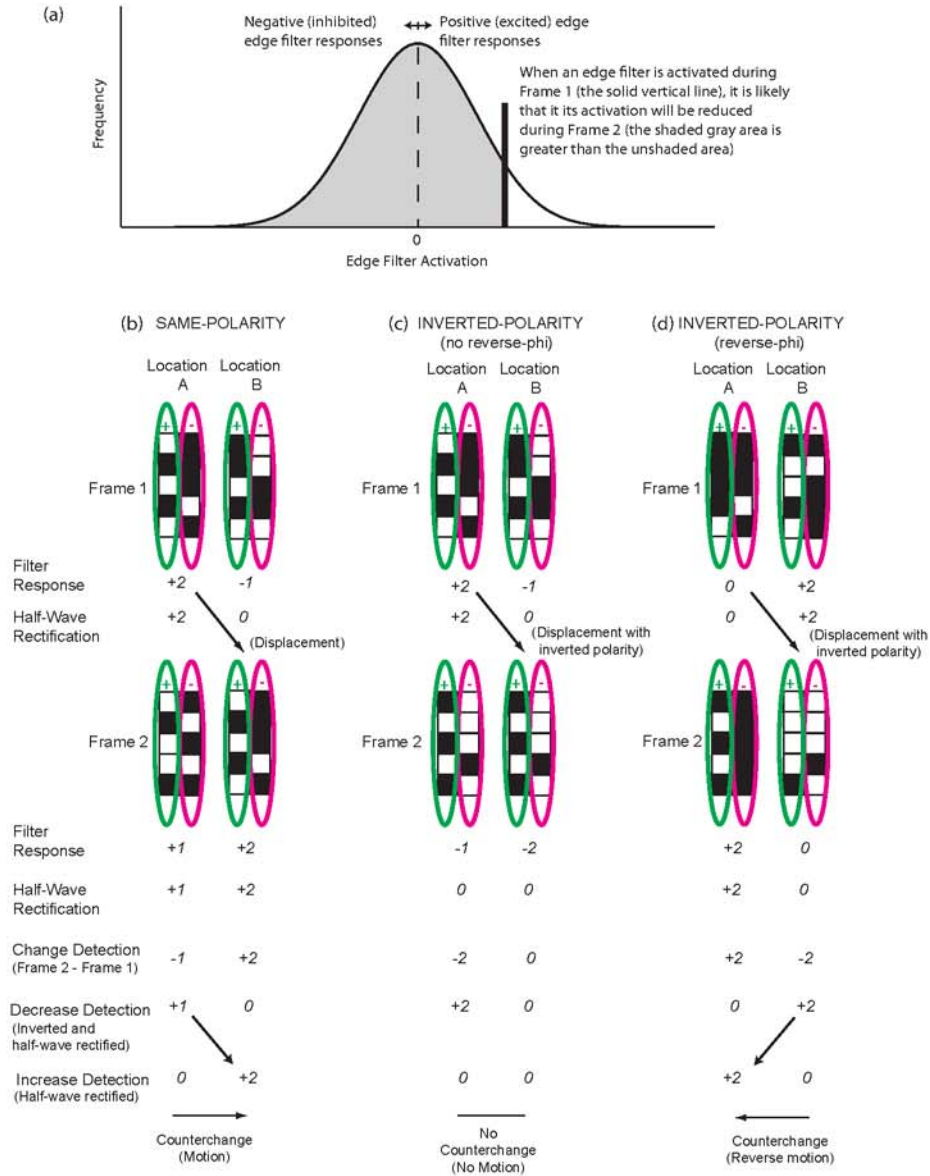


Figure 3.8: Sketch of counterchange detection of motion in random-dot cinematogram, restricted to one polarity for simplification. Panel (a) shows why positively activated edge filters are likely to undergo a decrease in activation when the pattern is displaced out of their current location. For the same reason there is likely to be an increase in activation at the location the pattern is shifted to. (b) shows an example of counterchange motion being detected for a stimulus in the same-polarity condition, (c) shows the typical nullification effects of polarity-inversion on counterchange motion, and (d) shows an arrangement of dots that elicit reverse-phi counterchange motion under polarity-inversion.

channels detect counterchange motion in parallel. One channel is responsible for edge filters with their excitatory zone on the left and the other channel for those with their excitatory zone on the right. The motion computations for the two channels are then combined and the leftward signals subtracted from the rightward to yield a single array of motion responses⁶.

Finally, the counterchange model assumes that any decrease in edge filter activation can contribute to only one motion signal. Shorter-path motions beginning at the location of the activational decrease are preferred over longer-path motions, and in the case of conflicting directions of the same span, the stronger motion is preferred (in the case of equal strength, one motion or the other is chosen with an equal chance).

Results

Single trial simulations. For small displacements (e.g. 2 bar-units rightward) in the same-polarity condition, rightward motions (in the direction of the displacement) were most strongly activated within the figure for the span corresponding to the size of the displacement (i.e. the motion signals were coherent; Figure 5c). Responses in the background regions were sparser than in the figure, with inconsistent directionality.

For larger displacements (e.g. 6 bar-units, Figure 5d), there was still activity within the figure region at the span corresponding to the displacement. However it was less consistent than for the small displacements, with the distribution of motion responses spread across other, especially shorter, spans. Again, the background regions are sparsely activated and directionally incoherent.

In the inverted-polarity condition, motion signals were generally very sparse, both

⁶Although we used both polarity channels in the current simulation, virtually identical results are obtained when only one polarity channel is employed. However, because the ERD utilizes both edge polarities, a more direct comparison was achieved by including both channels. Additionally, including both channels shows they do not interfere with one another.

within and outside of the figure. At the span corresponding to the displacement, there are no motion signals generated within the figure in the displacement direction, and a small number in the reverse direction. The latter skews the response distribution in favor of a leftward total response. This is indicated in Figures 8c and 8d and addressed in the General Discussion.

Simulation of experimental results. The counterchange model does a very good job of simulating the averaged experimental results for direction and shape discrimination (Figures 6a and 6b). It successfully simulates the effect of displacement size (both direction and shape discrimination were poorer for the larger figure displacements), and also simulates the weaker direction and shape discrimination obtained in the inverted-polarity condition.

These results contradict the general view that short-range motion is perceived via motion energy detection, and that the perception of reverse-phi motion in particular is necessarily the result of motion energy detection. They show that a much different, non-Fourier model entailing the detection of counterchanging activation can fully account for both the perception of short-range motion as well as motion in the reverse-phi direction.

Spatial pre-filtering

Whereas the scale of the edge filters for the ERD model were determined by the quadrature constraint of the model, the edge filters for the counterchange model were the same for all spans and selected to be responsive to the intrinsic scale of the checkerboard stimulus. The filters for the counterchange model therefore were relatively small. Morgan (1992), however, has argued for a stage of spatial low-pass filtering prior to motion processing in order to account for how effects of displacement size vary with the size of the elements and the spatial frequency content of the image.

Implementing this low-pass pre-filtering did not produce major deviations from the simulation results obtained with the counterchange model without pre-filtering. (This also was the case for the ERD model.)

3.4 GENERAL DISCUSSION

Any mechanism that yields symmetrical responses to same- and inverted-polarity two-frame stimuli cannot, by itself, account for asymmetrical data in either motion or shape discrimination for the short-range motion paradigm. In order for a motion detector to potentially account for the observed asymmetry, its polarity channels must either function in a completely segregated manner or contain a parameter that enables between-polarity interactions to be weighted differently than within-polarity interactions. The ERD, which in this article served as a representative model for the detection of first-order motion energy, does not segregate its polarity channels, nor does it contain a parameter which could weight the interactions of the polarity channels differently, and therefore necessarily gives symmetrical responses to same- and inverted-polarity conditions. Moreover, symmetry with respect to polarity inversion is not unique to the ERD. It is intrinsic as well to Adelson and Bergen's (1985) motion energy detector, which replaces the multiplication scheme of the ERD by a sum-or-difference-then-square scheme. Despite such internal differences, it is formally equivalent on output to the ERD.

Both the ERD and the motion energy detector are comparator-type detectors that call for a quadrature arrangement of filters in order to approximate a region in the spatiotemporal Fourier domain. However, this quadrature arrangement is not a necessary condition for obtaining symmetrical responses to same- and inverted-polarity stimuli. Rather, the symmetry that these detectors exhibit results from treating both

positive and negative spatial filter responses in the same manner; that is, the output values of spatial filters are treated arithmetically (e.g. multiplying negatives to get a positive response), rather than as representing a biophysical quantity in the nervous system. Consequently, when luminance polarity is inverted, the sign of the spatial filter response is also inverted, but retains the same magnitude. Regardless of whether one uses the multiplication scheme of van Santen and Sperling's (1985) ERD or the sum-or- difference-then-square scheme of Adelson and Bergen's (1985) motion energy detector, this inversion of the local spatial filter responses on the second frame results in a change in the sign (direction) but not the magnitude (strength) of the final motion detection output, leading to reverse-phi motion of equal magnitude to the same-polarity condition. Moreover, the symmetry that results from this multiplicative interaction is not unique to comparator-type detectors. Gradient detectors that evaluate motion at zero- crossings (Marr and Ullman, 1981) exhibit symmetry for the same reason. That is, inverting polarity on the second frame changes the sign of the temporal derivative, consequently inverting the sign of local motion signals while preserving their magnitude and spatial distribution (Sato, 1989). The contribution and interaction of negative values in comparator-type (and gradient) detectors raises questions with respect to their biological plausibility. Neural systems generally communicate via action potentials, where only positive activation is transmitted to post-synaptic units (Heeger, 1993). Inhibition of a neuron reduces the amount of output, but chemical synapses cannot transmit less-than-zero values. The less-than-zero contributions entailed in the ERD (and other models) makes a one-to-one mapping from the model to the nervous system doubtful, as negative values are not treated as inhibitory. In contrast, the counterchange detector, which successfully accounts for the asymmetrical effect of luminance polarity on direction and shape discriminations, is neurally plausible as only positive activation values contribute to motion-detection

computations.

3.4.1 Source of asymmetry and reverse-phi in counterchange model

The half-wave rectification of edge filter outputs also is responsible for motion being asymmetric in the same- and inverted-polarity conditions. Because motion is computed within polarity channels and not between them, stimuli that would have signaled motion in the same-polarity condition in most cases have their motions nulled rather than reversed in the inverted-polarity condition. An example in Figure 8 is restricted to one polarity channel for simplicity. In the same-polarity condition (Figure 8b), a pattern of elements that is positively stimulating edge filter A in Frame 1 is shifted to edge filter B in Frame 2. This shift causes a decrease in response in A and an increase in response in B, signaling motion in that polarity channel from A to B. In the inverted-polarity condition (Figures 8c and 8d), the response of B in Frame 2 is necessarily of the opposite polarity. Its response is therefore negative, and the half-wave rectification leads to an output of zero. A zero output during Frame 2 implies that over the course of the two frames at B, the only possible responses are a decrease or no response (i.e. there cannot be an increase to zero, as it is the lowest possible value for a half-wave rectified signal). In the inverted-polarity condition, some arrangements of stimulus elements lead to counterchange detection in the direction opposite to displacement (reverse-phi motion). Figure 8d shows an example of such an arrangement. In Frame 1, a near-zero response is elicited in an edge filter at Location A and a stronger positive response is elicited in an edge filter at Location B. In Frame 2, the near-zero response from Location A has shifted to Location B and been inverted, causing a decrease in activation (the inverted response of a near-zero output is also near-zero), while new elements are shifted into Location A that happen to cause an increase in that polarity channel, eliciting a (reverse) counterchange

response. This reverse-phi signaling is rare compared to counterchange detection in the same-polarity condition in the direction of displacement as most responses are zeroed and don't lead to a reverse-phi signal. This leads to the observed asymmetry between same- and inverted-polarity counterchange detection.

3.4.2 Half-Wave Rectification in the Counterchange Model

Half-wave rectification at each stage of processing (only positive activation levels are passed forward) is an essential feature of the counterchange model. In addition to its previously discussed biological plausibility (a given neuron can transmit more or fewer action potentials, but never less-than-zero), half-wave rectification ensures that inhibitory activation states have no role in signaling the presence of counterchange, which entails a motion event that is detected by virtue of the (effectively) simultaneous decrease in a feature at one location and increase in that same feature at another location. In the current case, the features are white-black (and black-white) edges that are formed by chance within a random cinematogram: motion is signaled from the location of a decrease in edge filter activation to the location of an increase in edge filter activation. Such features can be (more or less strongly) present, or not present, but not negatively present. Moreover, if half-wave rectification were removed prior to the detection of decreases and increases in spatial filter activation, the resulting negative values would introduce ambiguities into the conceptual framework of counterchange. For example, the response of a BW filter would be positive to a black-white (BW) edge, negative to a white-black (WB) edge, zero to a black-black (BB) non-edge, and zero to a white-white (WW) non-edge. If the BW edge filter is exposed to a two-frame sequence in which it is stimulated first by a WB edge followed by a WW non-edge, its activation will have gone from a negative value to zero, so it would have increased (assuming no rectification). However, in order to conform to

the principle of counterchange, this event is more appropriately registered as a decrease in the presence of a feature (WB edge), rather than as an increase in a feature (BW edge). Introducing half-wave rectification eliminates this ambiguity, treating the increase of a BW edge as non-symmetrical with respect to the decrease of a WB edge (and vice versa). In other words, the increase in one feature does not imply an equivalent decrease in its polar opposite feature. By including separate channels for each of the polar opposite edge filters, what would be a negative value for one channel (without rectification) constitutes positive values for the other channel. Removing rectification before the outputs of the increase and decrease subunits of the counterchange detector are multiplied also leads to violations of the counterchange principle, eliminating directional selectivity. That is, instead of motion occurring exclusively from the location of a positive response for a decrease subunit to the location of a positive response for an increase detector, the opposite motion could also be signaled from the location of the activation increase because the negative output from a decrease detector (indicative of an increase in activation) can be multiplied by a negative output from an increase detector (indicative of a decrease in activation) yielding a positive motion detector output, erroneously signaling motion from an increase to a decrease in local activation.⁷

3.4.3 Dual Motion Pathways

It is well known that the nervous system is segregated into two parallel pathways that respond with excitation to opposite luminance-contrast polarities. The so-called ON

⁷It would be feasible to remove one of the rectifiers after change-detection as long as the other was still present and achieve reasonable behavior from the detector; as long as the negative outputs of the motion detector were ignored and only positive outputs signal motion (if one channel can never go below zero a positive product cannot result from multiplying two negative values). However, the motion-opponency scheme employed here to evaluate the final motion detection output would demand half-wave rectification on output of the motion detector, effectively displacing a rectifier, but not eliminating it.

and OFF channels respond to luminance increments and decrements, respectively. Here we use the terms ON and OFF pathways to refer to two parallel channels opposite luminance polarity sensitivity, and do not intend to imply a specific type of spatial filter (e.g. center-surround, edge-detector, etc.). The two segregated polarity channels in the counterchange model can be interpreted as corresponding to these two pathways, each computing motion independently. Our simulations show that these two segregated counterchange channels (or either one by itself) are sufficient to account for both the standard and reverse-phi percepts in the current stimulus. Further, other studies have shown evidence for the independence of these channels in computing motion by demonstrating similar asymmetries (e.g. Edwards and Badcock, 1994; Wehrhahn and Rapf, 1992; Sato, 1989; Doshier et al., 1989). In contrast, Bours et al. (2009), using a sparse random-dot display in which individual motion signals were spatially and temporally uncorrelated, aimed to show that motion detection thresholds were symmetrical for same- and inverted-polarity dot-pairs. They argued that this suggests that motion is computed by correlating (with equal weighting) signals both within and between the ON and OFF polarity channels, with between-channel correlations signaling reverse-phi motion. Such an architecture could account for the symmetry observed in the ERD without appealing to the interaction of negative activation values (an example of such a detector can be seen in Eichner et al. [2011]).⁸

Although most of the parameter space probed in Bours et al.'s (2009) experiments

⁸Eichner et al. (2011) have also presented a '2-quadrant' Reichardt detector model in which only ON-ON and OFF-OFF spatial filter pairings are established to account for physiological findings in the visual system of the fly. This model showed weakened responses to inverted-polarity as compared to same-polarity stimuli. However, it included front-end elaborations whose introduction is not currently justified for the human visual system. Nonetheless, it would be valuable for future studies to compare the response characteristics of this Reichardt-variant detector to the counterchange detector under conditions which could clearly distinguish the models (i.e., stimuli in which no counterchange information is present but a clear autocorrelation is not, and vice versa).

was not indicative of symmetry (detection thresholds were higher for inverted-polarity stimuli), symmetry with respect to luminance inversion was consistently obtained for brief frame durations and small displacements. Because they also are the spatial and temporal conditions that are optimal for the perception of two-frame short-range motion (Braddick, 1978), it is worth considering the implication of these results for motion detection. That is, they indicate that for fast motions over short distances, direction discrimination is based on a motion mechanism that correlates within- as well as between-polarity channels, which is implied by motion energy models. Further, the spatially and temporally uncorrelated nature of the motion signals generated by Bours et al.'s (2009) stimuli implies that the integration of motion signals does not depend on their being simultaneous or spatially contiguous. In contrast, the short-range motion paradigm studied in the current article constrains coherent motion signals to occur simultaneously and within a spatially defined region (i.e. the displaced rectangle) where all dots undergo the same frame-to-frame translation. These conform to natural constraints of a rigidly translating surface, where motion signals are necessarily generated simultaneously and are in close proximity to one another by virtue of physical connectedness. Under these constraints, there is convergent evidence that spatial structure is not recoverable when luminance polarity is inverted, while it is recoverable when polarity is held constant. Evidence obtained in the current study, Sato's (1989) and Doshier et al.'s (1989) are consistent in indicating that same-polarity motion correspondences are essential for the perception of shape from motion.

Overall, these results are consistent with the existence of dual pathways, one entailing within-polarity counterchange mechanisms for the perception of motion for displaced objects, surfaces, and shapes, and the other entailing within- and between-polarity motion energy mechanisms for the perception of objectless global motion, without the individuation of particular objects, surfaces and shapes.

The distinction between these two kinds of motion pathways has its origin in Wertheimer's (1912) distinction between beta (object) and phi (objectless) apparent motion. More recently, Sperling and Lu (1998) asserted that object motion entails the detection of motion via their 3rd-order, salience-based motion system, whereas objectless motion is perceived when motion is signaled only by 1st- or 2nd-order motion energy systems. Further evidence for dual pathways has come from Azzopardi and Hock (2011), who found that motion direction can be discriminated in the cortically blind hemifield of an individual with unilaterally damaged visual cortex (and thus no object or shape perception) on the basis of detected motion energy, whereas motion direction was discriminated on the basis of changes in shape in the unimpaired hemifield. Finally, Hock and Nichols (2013) and Seifert and Hock (submitted) have provided evidence linking the perception of a surface's motion with the detection of counterchange and the detection of changes in luminance (without the perception of surface motion) to the detection of motion energy.

It is likely that these two motion systems, sensitive to different stimulus patterns, subservise different behavioral functionalities; e.g. the counterchange pathway to perceive changes in position of objects and the motion energy pathway to perhaps detect optic flow patterns that guide locomotion (Pelah et al., submitted). The motion energy pathway, which leverages both within- and between polarity-channel correlations, subserves 'global' motion perception, while the counterchange pathway, detecting only same-polarity patterns, subserves 'form/motion' perception, which can include the derivation of a figure's shape from the spatial relationships among counterchange-determined motion signals (Figure 9). Further empirical work to identify the spatial and temporal limits for the perception of spatial structure in the counterchange pathway and to determine what, if any, spatial localization is possible in the motion energy pathway would help to further distinguish these two systems.

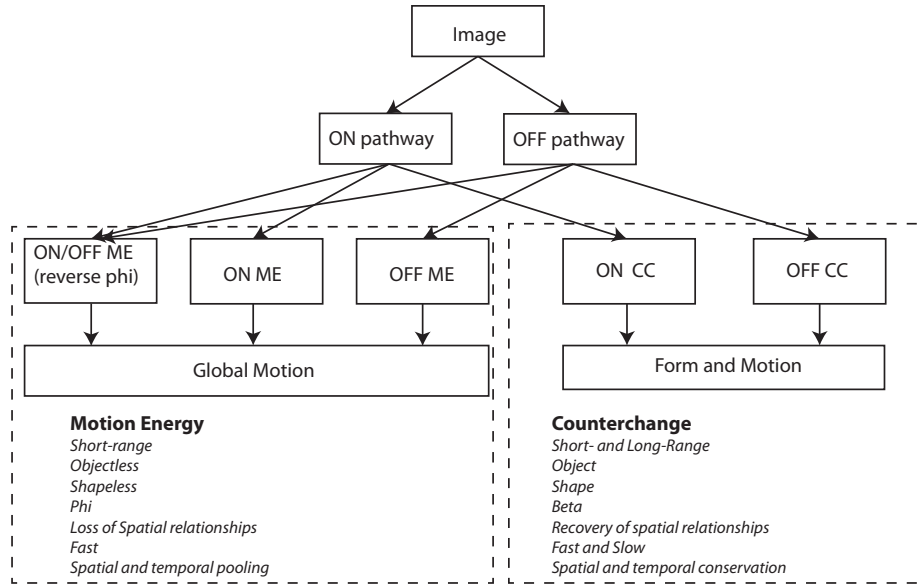


Figure 3.9: A conceptual model of a dual pathway motion system. ON and OFF here designate two channels with opposite luminance polarity sensitivity and do not necessarily imply a particular type of spatial filter. Both within- and between-polarity interactions subserve a motion energy (ME) system that detects global motion. Only within-polarity interactions subserve a counterchange (CC) system in which spatial relations of motion detectors are preserved allowing for recovery of form from motion.

To summarize, several speculative conclusions can be drawn from the relevant literature:

1. Although asymmetry in motion direction discrimination between same- and inverted-polarity stimuli is observed under most experimental conditions, evidence for symmetry is obtained for very fast motions over small distances in Bours et al. (2009). This parameter range is typically associated with the short-range paradigm, suggesting that the presence of spatial structure among motion signals, which is absent in Bours et al.'s paradigm but present in Brad-dick's short-range paradigm, can affect motion detection. The evidence for symmetry obtained by Bours et al. (2009) is consistent with motion energy as the basis for motion direction discrimination in the absence of spatial structure.

2. The presence of temporal simultaneity and spatial contiguity among motion signals is not necessary to obtain asymmetry with respect to luminance inversion; e.g. Bours et al. (2009) have obtained evidence for asymmetry with a stimulus for which motion signals are spatially and temporally uncorrelated (this was the case for slow motions over relatively long distances). However, when simultaneity and spatial contiguity are present, as in the short-range motion paradigm, asymmetry with respect to luminance inversion is obtained (as in the present study) even when fast motions are perceived over small distances (see Figure 6b).
3. Spatial structure and form, including depth structure, is recoverable only in same-polarity conditions (likely through the detection of counterchange) and is decimated in inverted-polarity conditions (Figure 2 in the current study; Sato, 1989; Doshier et al. 1989).
4. To the extent that ON and OFF channels (or other opposite-polarity channels) are correlated in motion detection, local spatial relationships are lost, and the motion percept could be called ‘global’. Spatial and temporal pooling in the motion energy pathway could be responsible for this loss (as suggested by the nature of the Bours et al. [2009] stimulus).

3.4.4 The Source of Shape from Coherent Motion

The dual pathways dichotomy described above proposes that the detection of counterchange is basis for the derivation of shape from coherent motion which has been defined as occurring when multiple motion detector responses agree in direction and span. When there is a high density of coherent motion signals within some region of the moving image, that portion of the image is perceived as moving together as a

continuous ‘surface’. In order to segregate the moving surface from the background, coherent motion signals must be relatively dense within the figure, and relatively sparse and/or incoherent outside the figure. This difference in coherence and density between the moving figure and the background is essential for successful segregation and the recovery of shape, as it is the only cue to the boundary of the figure.

This definition of coherence is at odds with how coherence is typically framed in terms of motion energy (Simoncelli and Heeger, 1998; Sato, 1989). The general motion energy approach entails taking local velocity estimates of oriented sinusoid components across a dynamic image. The output of a given motion detector is then considered a time-varying velocity estimate at a given location, where the sign of the output signifies the direction of motion, and the magnitude signifies the speed. In this view, multiple motion signals across some area of the image would be considered coherent if their direction and speed were sufficiently similar (Yuille and Grzywacz, 1998). In other words, among motion detectors of the same scale and directional selectivity, a low variance across the response magnitudes (speeds) would constitute evidence of coherent motion. This presents a challenge for the ERD account of shape-from-motion for short- range motion stimuli. For small displacements, single trial simulations for ERD detectors

(Figure 5a) indicate strong directional agreement, but with a high degree of variance in terms of magnitude (and therefore speed).⁹The current approach using the counterchange detector assumes a different role of motion detector responses. Rather than the magnitude of the response representing a velocity estimate, detector responses are conceived of as providing evidence for a given displacement (corresponding to the span of the detector). While the phase- invariant responses of motion energy detectors signal luminance-defined motion at a single location, counterchange detectors signal motion of an image feature (e.g., an edge) from one location to another.

A strong motion detector response indicates strong evidence for a given displacement corresponding to the detector’s span. Weaker responses, which could occur for multiple reasons (pattern details, smaller contrast change, etc.), are not indicative of slower speeds, but instead as reduced evidence of motion between two locations. There are two consequences of this approach: 1) counterchange-determined motion marks spatial distances, providing a direct basis for the recovery of shape from motion, and 2) rather than a homogenous (i.e. low variance) response magnitude across a given direction and span, a sufficient density of responses for a given span is required for coherence within an image region.

While the interpretation of the outputs of ERD and counterchange detectors differ in general, in the current article they both simulate shape judgments with the same template-matching scheme. This scheme does not take into account the variance of motion detector magnitudes and the criterion for coherence is the same for both models.

3.4.5 Theoretical Framework for the Recovery of Depth from Counterchange Motion

Although the definition discussed above limits motion coherence to motions of the same direction and span, this restriction can be relaxed to account for coherent motion patterns that give rise to the impression of depth structure in moving images. The framework follows from the idea that motion direction and shape discrimination entail patterns of activation within and across layers of motion detectors with the same directional selectivity, with each layer composed of a spatially distributed, densely packed array of motion detectors. The defining feature for each layer is that the same span separates the pairs of edge filters that compose its constituent detectors.

When the directionally-consistent motion detector activation within a displaced

surface is concentrated in a particular span-layer, it indicates that the detected motions all are in the same depth plane, as must be the case for two-dimensional surfaces oriented perpendicular with respect to one's line-of-sight. However, if motion signals within some local neighborhood occur at different, but similar, span-layers, these motions may be interpreted as belonging to a single surface that is non-uniform in depth. For example, if a one-dimensional slice were taken along the direction of motion from the front face of a rotating cylinder composed of moving dots, all the dots would be moving in the same direction, but would stimulate different span-layers depending on the speeds of the dots (the speeds are constrained by the three-dimensional structure of the cylinder). Dots near the outer edges of the cylinder would be moving relatively slowly, therefore activating small-span detectors. Towards the center of the cylinder the speed of the moving dots would increase, leading to the activation of larger-span detectors, with a maximum span reached at the center. With a sufficient density of dots, this cross-layer activation pattern would be smooth, with neighboring detectors differing only minimally in span. Templates similar to the ones used in the current simulations for single-depth motion could respond to sufficiently smooth patterns across span-layers, signaling depth structure in the moving image.

The single trial simulations in Figure 5, which were the basis for the discrimination of motion direction and shape in the short-range motion paradigm, made it possible for the counterchange model and the ERD motion energy model to be compared with respect to their compatibility with this theoretical framework for deriving depth structure from image motion. Two features of the simulations are relevant: 1) the extent to which motion detector activation for displaced surfaces is concentrated within the same span-layer, and 2) the spatial resolution of the activation patterns.

Concentration of activation within a span

It can be seen for the ERD simulations in Figure 5 that directionally-consistent motion is most strongly concentrated within the displaced surface for the detector span that corresponds to the surface's displacement. However, directionally consistent activation is evident for other spans. The latter occurs because the ERDs are Fourier-based, so their edge-filters are constrained to maintain a quadrature relation between filter size and span. As a result, the detectors composing different span-layers overlap significantly in terms of their spatiotemporal frequency response. Thus, a motion detection response in a given layer is likely to be accompanied by similar responses in layers with similar spatiotemporal frequency sensitivities (i.e. with similar spans). Because of the Fourier character of motion energy detectors like the ERD, this sort of diffusion across multiple span-layers is unavoidable for most displaced objects.

This 'muddling' of span-layer activation for the ERD does not occur for the counterchange model because directionally-consistent motions are concentrated within the displaced surface only for the detector-span that corresponds to the surface's displacement (particularly for small displacements). Because the counterchange model does not require a quadrature relationship between the span and size of the edge filters composing the motion detectors, detectors have the same size edge filter for every span. The consequence is that the spatiotemporal stimulus patterns that a detector is sensitive to are more dissimilar across span-layers than for the Fourier-based ERD. Because the edge filters for each span respond to the same stimulus information, the detectors whose span corresponds to the actual figure displacement will generally signal more strongly and more often than displacement-inconsistent spans. In addition to this, the shortest-path selection constraint in the counterchange simulation minimized further the incidences of multiple motion-signals occurring across multiple

span-layers at a given location.

Spatial resolution

It also can be seen in Figure 5 that essentially all ERD detectors composing a span-layer are activated for virtually every location across the short-range motion stimulus, regardless of whether the detectors' edge filters are responding to changes in element luminance occurring within the displaced figure or within the background. In contrast with this spatially continuous distribution of activation, the distribution of counterchange detector activations within the figure is dense but discontinuous, and outside the figure responses are very sparse (Figure 5c). This is due to the counterchange detectors being much more selective than the ERD motion energy detectors (and not to the difference in spatial filter inputs to the two models). That is, counterchange detectors are responsive to a much smaller number of random dot patterns than are motion energy/comparator models like the ERD. This is because counterchange detectors are activated only when their edge filters are affected by changes in element luminance that result in decreases in edge filter activation at one location and increases in edge filter activation at another location, whereas nearly any change in edge filter response will result in a motion signal for the ERD. A discontinuous but dense distribution of activated motion detectors is important for the spatial resolution of the shapes that are derived from detected motion, especially when such a pattern indicates depth-structure. That is, recovering depth would be exceedingly difficult if it were unclear which span-layer was optimally stimulated at a given location, as the relation between neighboring motion signals at different spans would need to be differentiated in order to discern differences in depth.

3.4.6 Conclusion

In this article we have demonstrated the insufficiency of comparator-type motion energy detectors such as the ERD in accounting for motion direction perception and shape-from-motion segregation in the short-range motion paradigm. As an alternative, we have shown the plausibility of a counterchange-based mechanism in accounting for these experimental results. It is argued that the detection of counterchange-determined motions mark spatial distances, providing a direct basis for the perception of spatial shape from motion. In addition, we have suggested how counterchange detection could be extended to account for the recovery of depth from motion. Finally, non-Fourier counterchange detection can potentially account for other phenomena (e.g. the correspondence problem) that do not conform well to motion-energy formulations without necessitating high-level token-trackers or centralized cost-function calculations (Dawson, 1991; Morgan, 1992).

Dynamical Preliminaries

4.1 STABLE FIXED-POINT MODEL OF A SIMPLE NEURON WITH INPUT

Neurons are coupled to one another via synapses. Synapses release neurotransmitter under the influence of action potentials. When a neuron receives excitatory (inhibitory) input from other neurons, it becomes depolarized (hyperpolarized). Under no synaptic input, many neurons return to a baseline 'resting level' around which they fluctuate within a limited range.¹ Many physiological and biochemical processes are responsible for the maintenance of a neuron's resting level (e.g. ion channels, local chemical gradients). However for the purposes of modeling perceptual processes, these details are (arguably) irrelevant.

Below, a brief overview of modeling perceptual systems as continuous-time dynamical neural models is presented. For more on the potential role of dynamical stability in the formation and stabilization of percepts in general, see Hock et al. (1993) and Hock et al. (2003).

¹Of course, there are multiple neuron types that may behave differently; for example, bursting neurons may continue to oscillate continuously under no synaptic input.

4.1.1 Stable fixed-point

What is relevant is the emergent property of the neuron's *resting level*. This property is captured and modeled as a simple dynamical system with negative feedback; that is, a system with a single stable fixed-point.

$$\tau \dot{u} = -u$$

The variable u is used to represent the *activation state* of the neuron. Roughly, the activation can be mapped onto the membrane potential of the neural unit. The activation state is a continuous real variable that can take on both positive and negative values. Because the neural element is modeled as a stable fixed point, the system returns to the fixed point when it is perturbed away from it (i.e. the system forms a single stable attractor). The parameter τ establishes the timescale of the dynamical evolution of the equation (in milliseconds by convention).

It is easy to see why this behavior emerges. When the system is in a positive state ($u > 0$), the change in the system state is negative ($\dot{u} < 0$), and the state of the system is lowered. In a symmetrical fashion, when the system takes a negative state, the change of the state is positive, and the state of the system is increased. Thus, the system tends to return to 0.

4.1.2 Resting Level

One can offset the location of this stable fixed-point simply by adding a constant (h) to the equation.

$$\tau \dot{u} = -u + h$$

Setting $\dot{u} = 0$ it is easy to see that the fixed point is now at $u = h$. Henceforth, h will be referred to as the *resting level* of a neural element (Hock et al., 2003). In the models, h will typically take a negative value. This is an arbitrary convention, as

what actually has an effect on the evolution of the neural network is the relationship between the resting level and the threshold for interaction (defined by the interaction function), which will be discussed later. The resting level of a neuron is generally below the interaction threshold.

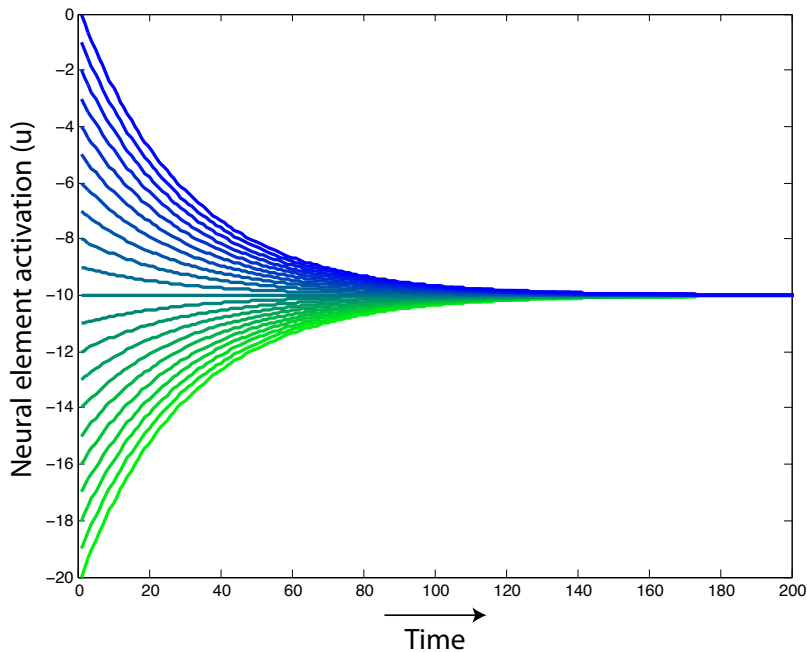


Figure 4.1: Relaxation to stable resting level $h = -10$ from various initial conditions

4.1.3 Neuron with simple input

Neurons receive input, both from stimulus and other neurons. In general an input can be modeled as another additive term, here denoted as S , which moves the stable attractor to $u = h + S$.

$$\tau \dot{u} = -u + h + S$$

In the simulations to follow, stimuli will typically be represented as spatiotemporal functions of the form $S(x, t)$ where x is space and t is time. Input from other neurons

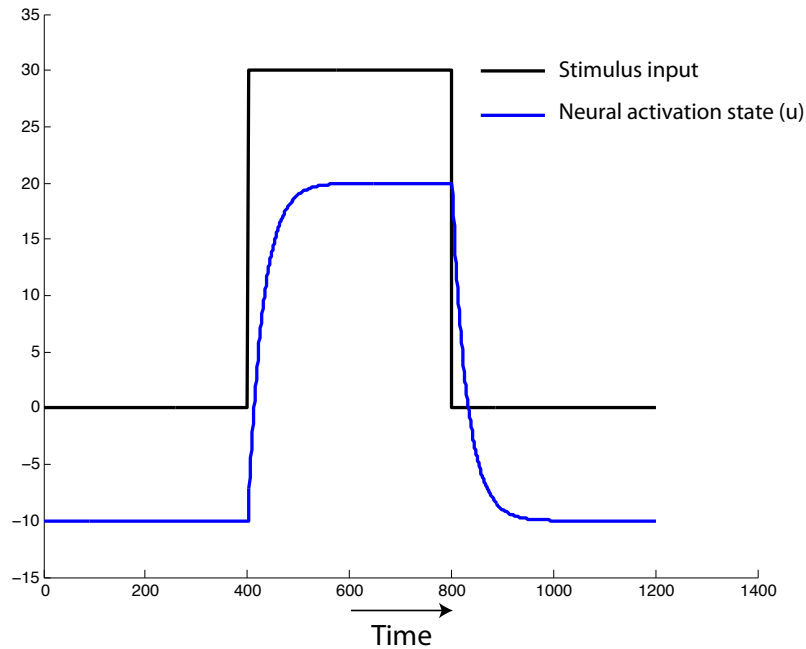


Figure 4.2: Neuron response to simple time-varying input

take a more specific form.

4.1.4 Stochastic fluctuations

In addition to the deterministic dynamics of the fixed-point neuron, a stochastic noise term ξ can be added to the equation. In this dissertation, the noise-term is normally distributed with a mean 0. Adding the noise term serves two purposes: 1) including random fluctuations helps to validate the stability of the obtained solutions and robustness of the model, showing the results are not highly-dependent on the details of the numerical integration, and 2) as will be shown in the next chapter, random fluctuations can promote and induce perceptual switches in multi-stable stimuli.

The stable fixed-point dynamic of the model neurons reduced the noise-induced variance of the neuron's activation state. Figure 4.3 compares the behavior of a simple

stochastic system with negative feedback $\tau \dot{u} = -u + \xi(t)$ with a random-walk (with no feedback) $\tau \dot{u} = \xi(t)$ generated with the same pseudorandom noise vector. Without stabilizing feedback, the system may wander arbitrarily far away from its initial condition, whereas the stabilizing feedback keeps the random fluctuations within a small range around zero.

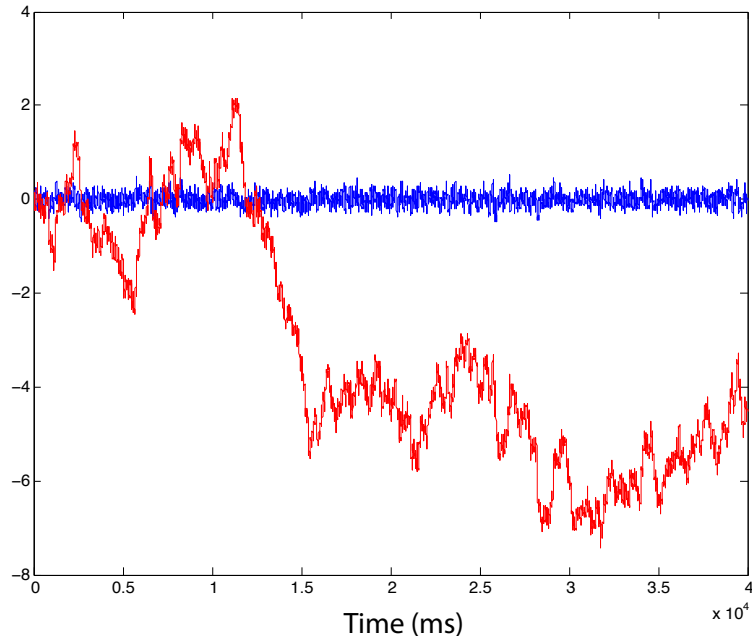


Figure 4.3: A random-walk vs. a neuron stabilized with negative feedback

4.2 CHANGE DETECTION NEURONS

The counterchange motion detection principle is based on local oppositely-signed changes in spatial filter activation. Therefore, it is important to be able to (separately) measure local increases and decreases in a neural-dynamic fashion. For this purpose an elaborated version of the fixed-point neuron model is developed into both an increase and decrease detection neuron variant. Equations of this form for local

change detection were originally presented by Berger et al. (200x). First, an increase detector variant is considered.

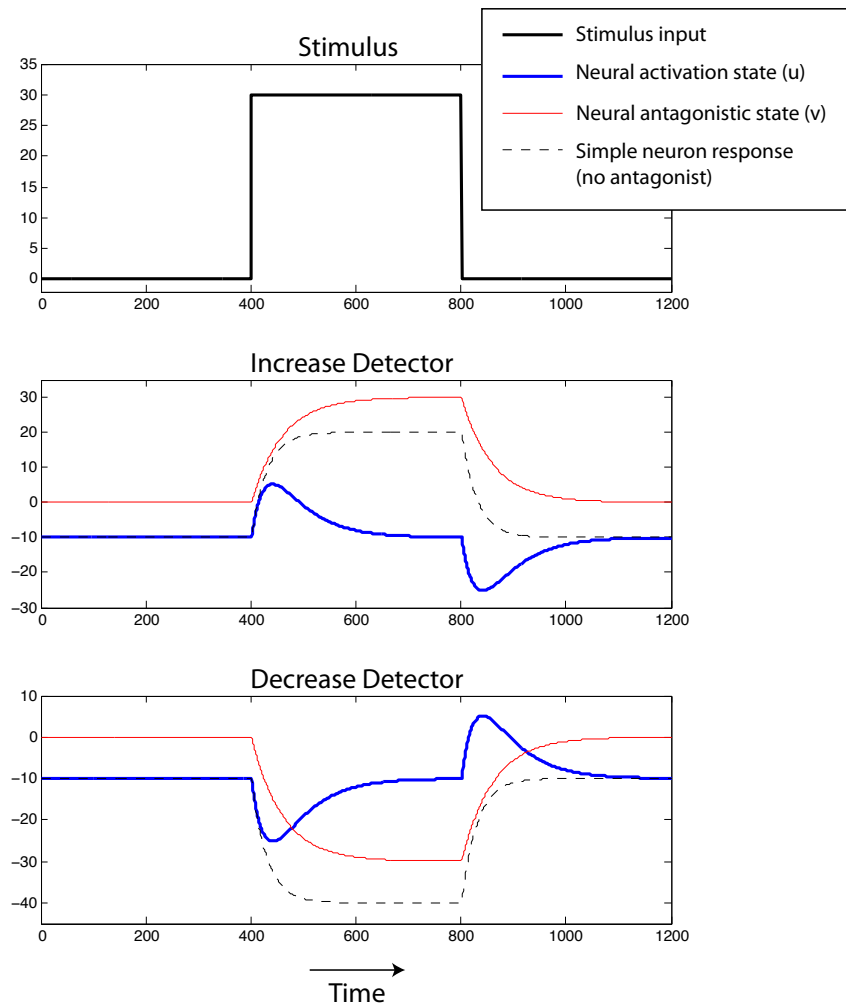


Figure 4.4: Increase and decrease detector responses to simple time-varying input

4.2.1 Increase detection

An increase detector is defined by the following two-equation system:

$$\tau_{fast}\dot{u} = -u + h - v + S$$

$$\tau_{slow}\dot{v} = -v + S.$$

The elaborated change detection neuron functions by having two state variables that operate on different timescales. Both variables are impinged on by the same stimulus (input) S . A state variable v represents a slower antagonistic dynamic component of the faster state variable u ($\tau_{slow} > \tau_{fast}$). For simplicity, assume $h = 0$. When a stimulus is first presented to an increase detector (e.g. S increases from zero to a greater-than-zero value), both state variables begin to evolve toward a newly formed attractor. Momentarily, $v \approx 0$, and therefore both u and v begin to evolve towards the stable fixed-point $u = v = S$. The activation variable u evolves more quickly than does v due to its faster intrinsic timescale, and thus its activation begins to rise above resting level. After a sufficient amount of time, $v > 0$ and the (transient) stable fixed-point for the activation variable u lowers to the value $u = -v + S$. Eventually, v evolves to the stable fixed-point value $v = S$, and therefore u evolves toward the attractor $u = -v + S = -S + S = 0$. Until a further change in input, u will remain at the stable fixed-point at 0. More generally, when $h \neq 0$, the system evolves towards the attractor $u = -S + S + h = h$.

4.2.2 Decrease detection

A decrease detector takes a similar form to the increase detector, except that the sign of the stimulus input is inverted.

$$\tau_f\dot{u} = -u + h - v - S$$

$$\tau_s\dot{v} = -v - S$$

Consider a (temporarily) stationary positive-valued input $S = 10$. For the same reason as for the increase detector, after a sufficient amount of time the activation variable u evolves towards the resting level. Now consider what happens when the input S is removed, i.e. decreases in value from 10 to 0. When the stimulus is removed,

the slow variable v that had previously evolved to $v = -S$ has an excitatory effect on the fast variable u . This is because the value v is negative, and the term containing v in the equation for u is negative. Thus the sign is inverted to become a positive influence on the evolution of u . In the short-term u therefore grows, exhibiting excitation in response to the decreased input. Eventually, v evolves back towards $v = -S = 0$, and therefore $u = -V - S = 0$. Again, in general when $h \neq 0$ the system evolves towards $u = h$.

4.2.3 Biphasic inhibition

It should also be noted that the change detection neurons are not only excited by their preferred stimulus, but are also inhibited by their non-preferred stimulus. That is, increase detectors are inhibited by decreases in input, and decrease detectors are inhibited by increases in input. This is referred to as *biphasic inhibition* and it has been previously shown to account for classic effects of inter-stimulus-intervals on the perception of standard apparent motion. Here, the implications of biphasic inhibition are not explored in detail. For a thorough discussion, please see Hock, Schönér & Gilroy (2009) and Gilroy & Hock (2009).

4.3 NEURAL INTERACTION

Neurons are not isolated entities but exist in networks, and as such neurons receive synaptic input from one another. As mentioned above, neurons typically interact via chemical synapses at which neurotransmitters are released in response to action potentials traveling down the axon of a neuron. Firing rates of action potentials are, by definition, positive. While a neuron is in an inhibitory state, it may fire few or no action potentials, in an excited state it may fire many in a short period of time.

Therefore, a neuron in an inhibitory state can't influence a neuron it projects to, while an excited neuron can. This inherent neural asymmetry is expressed in the model as a sigmoidal *interaction function* (i.e. a soft-threshold).

$$f(u) = \frac{1}{1+e^{-u}}$$

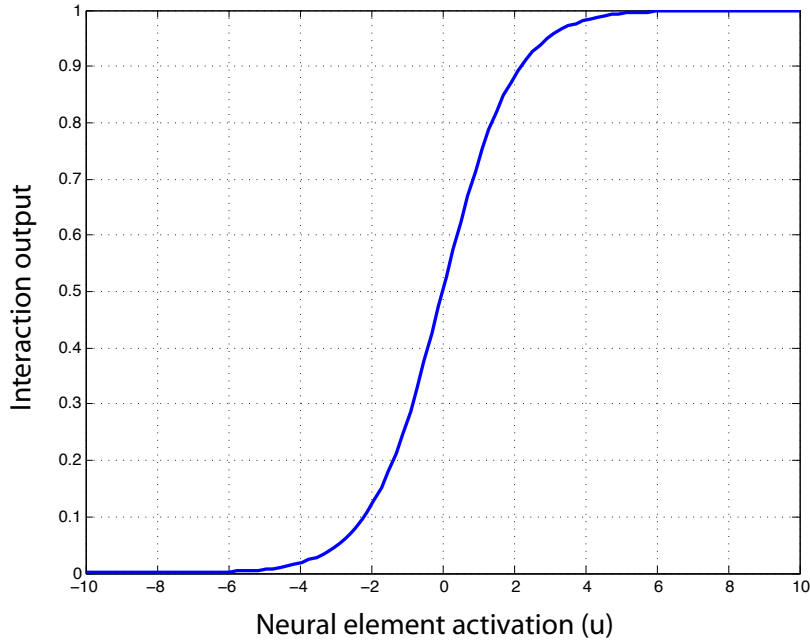


Figure 4.5: Sigmoidal Interaction Function

The output of this threshold function is always between 0 and 1 ($0 < f(u) < 1$). That is, it is a non-negative saturating function. Each neuron-to-neuron coupling term is of the form $w \cdot f(u)$ where w is the *synaptic strength*, which can take on positive (excitatory) or negative (inhibitory) values.

Consider the following system of equations:

$$\tau \dot{u}_1 = -u_1 + h + S$$

$$\tau \dot{u}_2 = -u_2 + h + w_{12} \cdot f(u_1)$$

These equations represent two individual neural elements, with subscripted indices

1, 2. Neuron 1 has a stimulus input S . Neuron 2 does not have a stimulus input, but instead has a synaptic input from neuron 1. The synaptic weight w_{12} is multiplied by the thresholded activation state of neuron 1, $f(u_1)$. Because the function $f(u)$ is bounded between 0 and 1, the (absolute) maximum effective input from neuron 1 to neuron 2 is w_{12} .

Here, an example is shown in which the input, $S = 30$, is turned on at 100 ms, before which it is zero. The resting level for both neurons is set to $h = -10$. The synaptic coupling from neuron 1 to neuron 2 is set to $w_{12} = 20$.

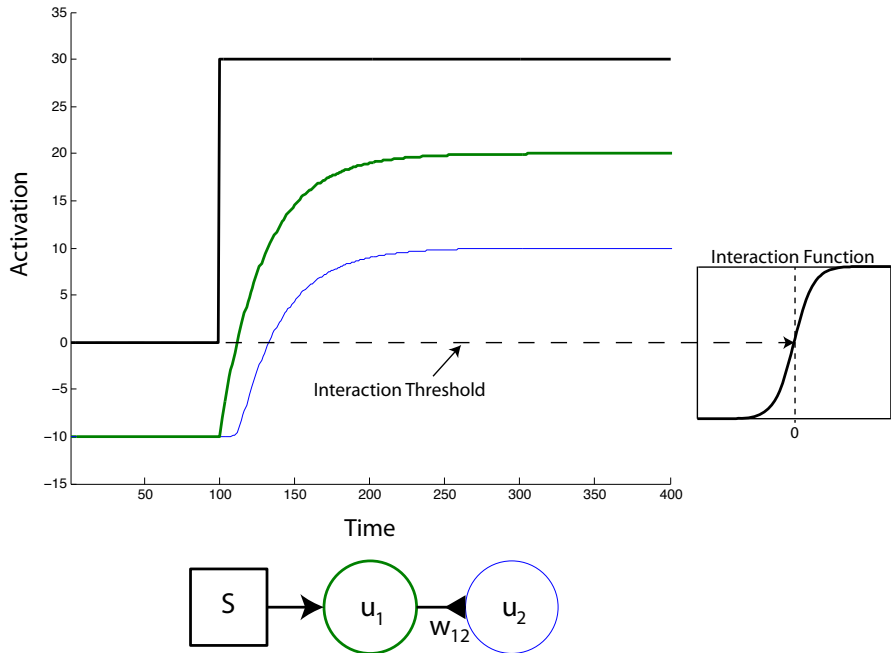


Figure 4.6: A simple two-neuron feedforward network demonstrates neural interaction through a sigmoidal function. A neuron u_1 receives stimulus input and forms an excitatory synaptic connection with a second neuron u_2 to which it passes on activation.

As can be seen in Figure 4.6, up until the input is switched on at 100 ms, both neurons sit at their resting levels. When the input is turned on, neuron 1 begins to evolve towards its newly-formed attractor at $u_1 = h + S = 20$. As the activation state

rises, neuron 2 begins to receive effective input from neuron 1 through the interaction function $f(u_1)$.

Because the stimulus input is strong enough for neuron 1 to essentially saturate (i.e. $f(u_1) \approx 1$), the effective input to neuron 2 becomes $w_{12} \cdot f(u_1) \approx 20$. This implies that an attractor is formed for neuron 2 at $u_2 = 10$, which it can be seen to be approaching asymptotically in the figure.

This example shows a single connection from one neuron to another with a simple static input turned on at 100 ms. Some possible patterns of coupling, not shown here, include multiple inputs/outputs to/from a single neuron, feedback (reciprocal connectivity) between neurons, and self-excitation. However, the basic premise, that neurons couple through (asymmetrical) synapses via a sigmoidal interaction function, is invariant throughout the remainder of the dissertation.

4.3.1 Additive and Multiplicative Synapses

Typically, neural networks are modeled with additive synapses. That is, each projection from one neuron to another is an additive term which contains a synaptic weight w multiplied by the output of the projecting neuron $f(u)$. So, if a single neuron receives inputs from multiple neurons, the terms are added together, e.g. $w_1 \cdot f(u_1) + w_2 \cdot f(u_2)$. However, neural interactions can also be modeled as multiplicative. That is, rather than summing the (weighted) output of multiple projecting neurons, they could be multiplied, e.g. $w \cdot f(u_1) \cdot f(u_2)$ (a single synaptic weight is used for simplicity). There is substantial physiological (e.g. Gabbiane et al., 2002) and psychophysical (e.g. Gilroy and Hock, 2004) evidence that cortical responses can display multiplicative-like behavior, although the mechanism for this behavior remains unclear.

Multiplication also serves as a natural mathematical mapping for concepts in

which multiple conditions must be true for some induced condition to be met. In the case when there are two conditions which must be true, multiplication essentially functions as a logical AND gate. In the next chapter a neural network for solving the motion correspondence problem is developed in which multiplicative synapses play two roles, both of which correspond to a logical AND in the conceptual schema. The first is the detection of counterchange motion itself (the co-detection of an increase and a decrease), the second is the (cooperative) competition among motion signals that constrains the subset of motion paths that are perceived. It is shown that an additive inhibitory alternative is not sufficient to account for all classes of patterns presented to the model. Both of these functions, particularly the latter, will be described in detail in the following chapter.

In order to illustrate some of the dynamical advantages of multiplicative synapses in a simple pattern detection task, consider the following neural network defined by the set of equations 4.1 and diagrammed in Figure X. Two pattern-detection neurons x and y both receive input from two sub-pattern-detection neurons p and q , where the projections from p and q terminate in an additive synapses at x and a multiplicative synapse at y . Each synapse has a weight associated with it denoted as w_{add} for each of the additive synapses on x and w_{multi} for the multiplicative synapse on y . Assume the function of pattern-detectors x and y is to detect the coincidence of detection of sub-patterns p and q ; and further that it is a false alarm if a pattern-detector shows significant activation when either p or q are not active. Let the (arbitrary) *absolute* perceptual threshold $\alpha = 0$ be the value that over which the percept associated with the detector is formed, and let α^* , the *effective* perceptual threshold, be the difference between the resting level and the perceptual threshold. Thus, when $\alpha = 0$, $\alpha^* = -h$.

The maximum effective input (assuming positive synaptic weights) to additive neuron x is $2 * w_{add}$ and to multiplicative neuron y is w_{mult} . To make comparisons as

direct as possible the simulations below will be constrained by the relation $2 \cdot w_{add} = w_{mult}$ except where otherwise noted.

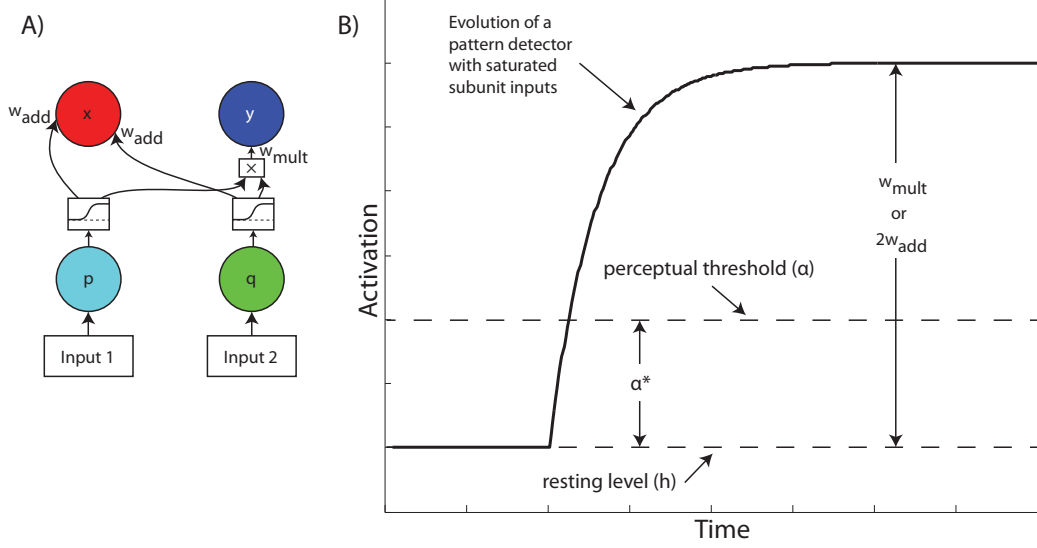


Figure 4.7: A) A wiring schematic for a small network to compare the dynamical consequences of additive and multiplicative synapses on pattern detection. Sub-pattern neural elements p and q synapse in an additive and multiplicative fashion to pattern-detector x and y , respectively. B) Diagram of symbols used in the discussion of the relationship between synaptic strength and pattern-detector response.

Consider the system of equations 4.1

$$\begin{aligned}
 \tau \dot{u}_p &= -u_p + h + S_1 \\
 \tau \dot{u}_q &= -u_q + h + S_2 \\
 \tau \dot{u}_x &= -u_x + h + w_{add} \cdot f(u_p) + w_{add} \cdot f(u_q) \\
 &= -u_x + h + w_{add} \cdot (f(u_p) + f(u_q)) \\
 \tau \dot{u}_y &= -u_y + h + w_{mult} \cdot f(u_p) \cdot f(u_q)
 \end{aligned} \tag{4.1}$$

In the case where both subunits p and q receive input, both the additive and

multiplicative pattern-detectors are also activated. Assuming all neurons essentially saturate, and that $2w_{add}, w_{mult} > \alpha^*$, both of the neurons successfully detect the pattern (Figure 4.9).

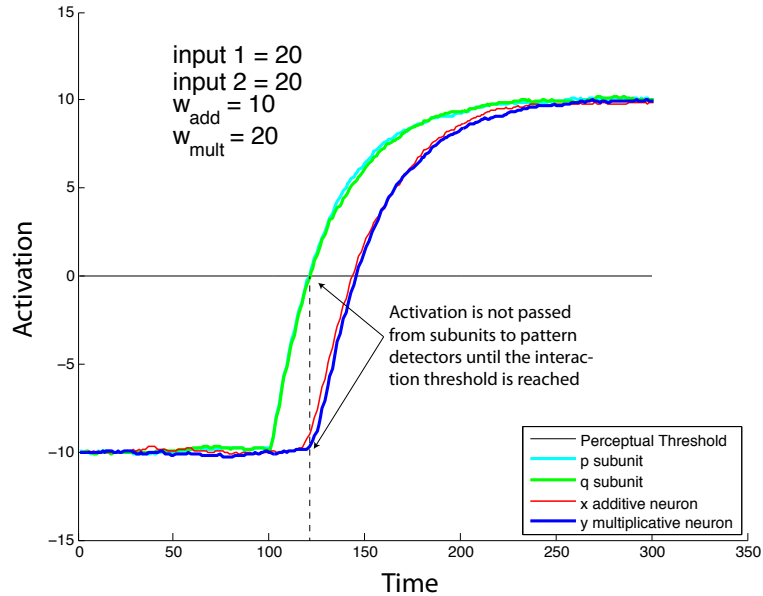


Figure 4.8: Additive and multiplicative pattern-detectors responding correctly to input. Both inputs are turned on at 100 ms. Subunits p and q are driven to saturation. When they cross the interaction threshold, both the additive and multiplicative pattern detectors, x and y respectively, are driven above perceptual threshold.

Using the same synaptic weights, when only the subunit q receives input, the additive neuron reaches only half of its previous activation. In this case, $w_{add} < \alpha^*$, so the additive neuron x does not signal that a pattern has been detected. The multiplicative neuron remains essentially at its resting-level. This is because the output of neuron p is so close to zero that the synaptic product is also very close to zero. In this case, both pattern-detectors correctly produce a ‘no-detection’ response.

Increasing (e.g. doubling) the synaptic weights produces some notable effects. When both inputs are on, both the pattern-detection neurons correctly show activa-

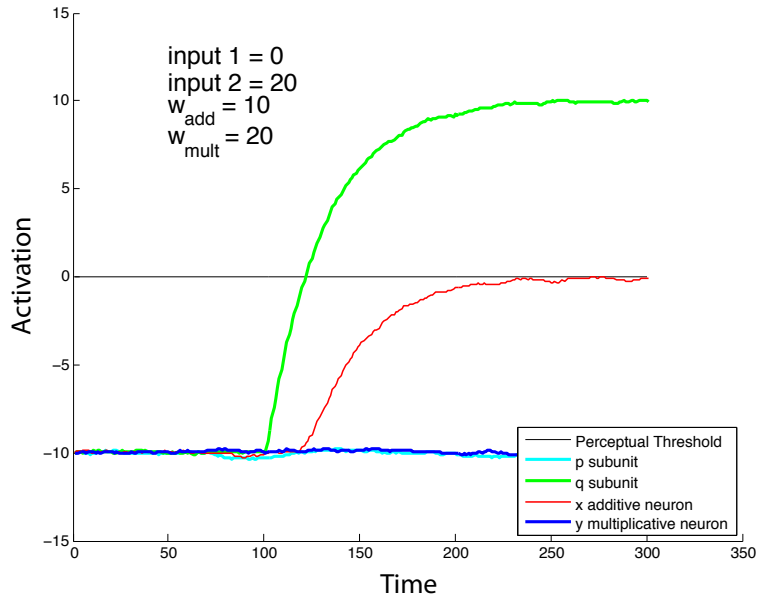


Figure 4.9: Additive and multiplicative pattern-detectors correctly detecting no pattern when only one input is present. The additive neuron receives effective activation from one of its subunits, but relaxes below the perceptual threshold. The multiplicative neuron remains at resting level.

tion values above α . Additionally, relative to the trial with smaller synaptic weights, the latency of the response (the time from when the stimulus turns on until the pattern detectors cross the perceptual threshold α) is shortened for both the additive and multiplicative neurons. This is a result of the newly formed attractor being further from the resting-level, and therefore inducing a faster dynamical change.

With these same synaptic weights, removing input to one of the sub-pattern detectors p highlights a fundamental difference between the additive and multiplicative pattern-detectors. Because $w_x > \alpha^*$, the single additive input is enough for activation of x to rise above α . This is considered a false-positive, as only one of the two necessary pattern components is present. The multiplicative neuron y once again remains essentially at resting-level.

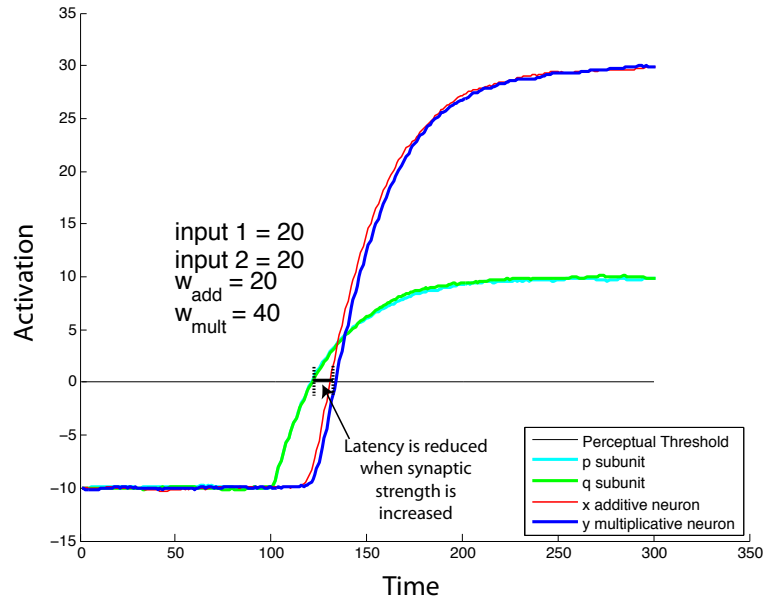


Figure 4.10: Additive and multiplicative neurons with increase synaptic weights. Latency of response after the subunits cross the interaction threshold is reduced in both pattern detectors as compared to the lower synaptic weights used in the simulations above. As will be shown in the next simulation, the additive neuron’s synaptic weight causes false positives when only one subunit is active.

Thus, we can say that for a functional pattern-detector comprised of two sub-patterns, the synaptic weight w must be less than α^* so that a single subunit input is not enough to drive the pattern detector over its perceptual threshold. Additionally, w_{add} must be greater than $\frac{1}{2}\alpha^*$ so that together the sub-patterns provide enough feedforward activation to drive the pattern-detector above α . Therefore, we can say that for an additive pattern-detector with two sub-pattern inputs, $\frac{1}{2}\alpha^* < w_{add} < \alpha^*$.

In contrast, the only logical constraint on the multiplicative synaptic weight is that it must be large enough to cross the perceptual threshold, thus $w_{mult} > \alpha^*$. The fact that w_{mult} can be (much) larger than w_{add} confers a potential dynamical advantage to the multiplicative synapse with a large weight. Consider the case that is no longer constrained by the relation $2 \cdot w_{add} = w_{mult}$. If we set w_{add} to its maximum

functional value, and w_{multi} to some arbitrary value greater than w_{add} , it can be seen that the latency of the multiplicative neuron y is much less than the latency of the additive neuron x .

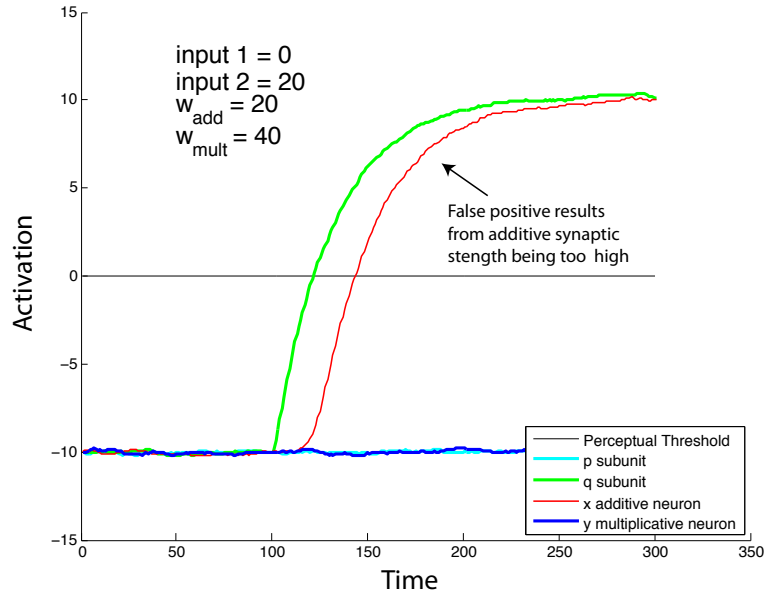


Figure 4.11: With increased synaptic weights, when only one subunit is active, the additive neuron incorrectly crosses the perceptual threshold. The multiplicative neuron remains at resting level.

Multiplicative synapses confer several advantages for the purposes of the present modeling. Large synaptic weights increase the rate of change of the neural dynamics, an extremely desirable characteristic for perceptual systems. The maximum synaptic weight for pattern detectors with additive sub-pattern inputs is bounded at a relatively low value. Multiplicative synapses can have large synaptic weights without causing false alarms when an insufficient number of sub-patterns are stimulated to constitute perceiving a pattern. For the same reason, the range of (synaptic) parameter values for which the dynamical behavior is qualitatively the same is much larger. In other words, the model needs less tuning. Finally, in what follows it is

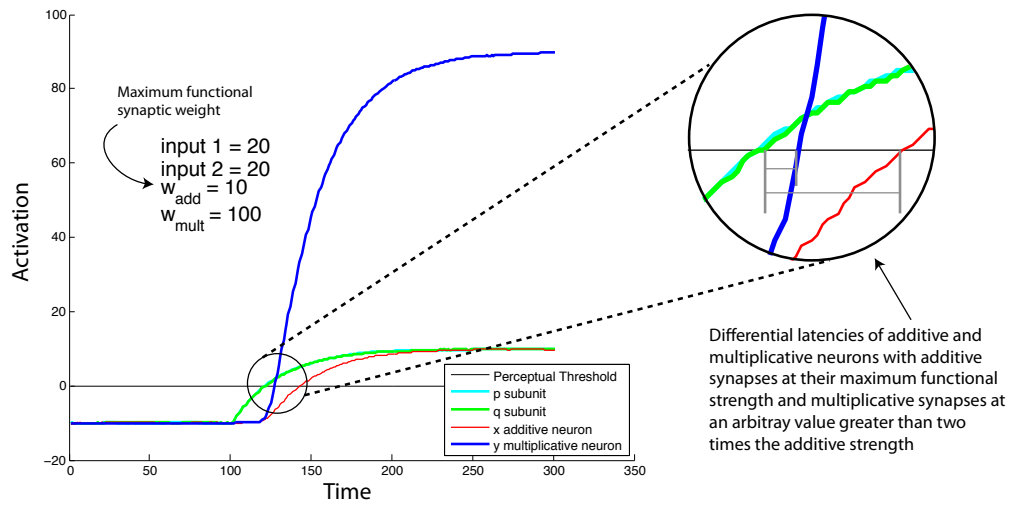


Figure 4.12: Setting the additive neuron to its maximum functional synaptic strength (i.e. the value at which any increase in synaptic strength would result in false positives) and the multiplicative neuron at an arbitrary value greater than two times the additive weight shows the difference in latency for the two (functional) pattern detectors.

shown that employing an additive inhibition scheme analogous to the multiplicative one employed in the main model is not sufficient to account for all of the motion patterns presented to the model.

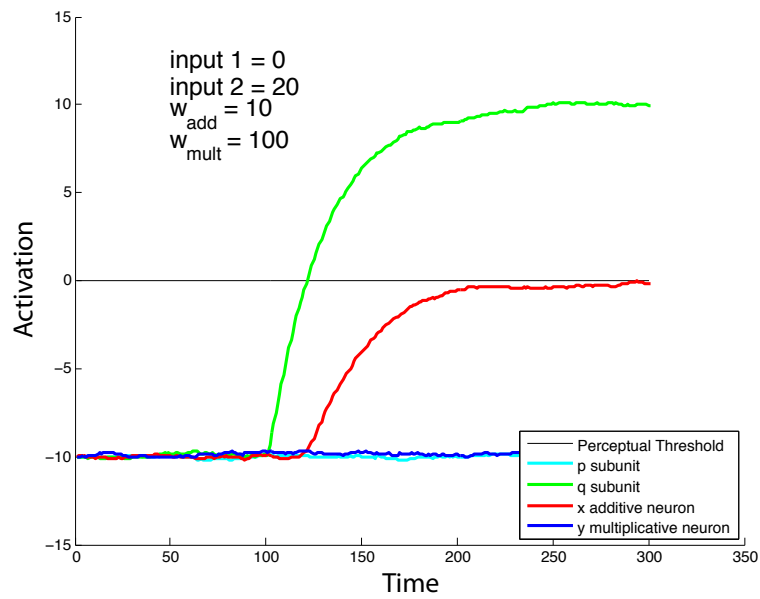


Figure 4.13: Removing one of the inputs verifies that both neurons remain functional and do not elicit false positives. Again, the additive neuron is effectively activated by its lone input (below perceptual threshold), and the multiplicative neuron remains at resting level.

A Dynamical Neural Network for Solving the Correspondence Problem

5.1 INTRODUCTION

The motion correspondence problem is a general problem the visual system faces whenever ambiguities occur in the changing location of visual features over time, whether these ambiguities are due to noise, to features being discretely displaced via an artificial display, to their occlusion, or to any other visual anomaly. In addition to disruption of the continuous availability of optical information, spatiotemporal context might lead to one perceptual interpretation being preferred over another (i.e. previously induced percepts can influence current ones). In other words, the visual system must solve the problem of ‘what moved where’ in the face of uncertainty.

The correspondence problem is intimately related to the coherent motion perceived in the random-dot cinematograms in the Chapter 3. In terms of paradigms, they differ in that ‘correspondence problems’ typically 1) contain fewer and larger visual elements with a sparser distribution, 2) are not embedded in noise, and 3) may induce more intricate motion percepts (e.g. splitting and fusing of elements). However, there is no principled dividing line between the two, and the term is often used loosely. In random dot cinematograms, coherent motion (that is, motion of the same direction

and span) of many elements is necessary to disambiguate the signal from the noise, as the number of potential mismatches is overwhelming. Because of this, theoretical analyses of random-dot stimuli generally approach the correspondence problem from a statistical point of view (e.g., Read, 2002). That is, rather than evaluating whether *specific* motion paths are perceived or not between individual visual elements, it is generally sufficient to show that there is a statistical reliability in the perception of motion direction; in other words that the majority of generated motion signals correspond to the veridical displacement. This is evident in the treatment of the short-range motion paradigm in the context of counterchange detection in Chapter 3. In this chapter, we are more interested in the specific element-to-element motion paths that are perceived when a (relatively) small number of visual elements are present on each frame. Reducing the necessity of strong coherence for motion signals to be reliably generated and disambiguated from noise allows a more nuanced look at the class of typically perceived correspondence patterns.

A classical correspondence display consists of two discrete frames that contain some salient visual elements whose location and possibly number differ on frames 1 and 2. Assuming there are n visual elements on frame 1 and m visual elements on frame 2, there are a possible $n \times m$ correspondence matches, and if each possible correspondence match can either be perceived (1) or not (0), there are a total of $2^{n \times m}$ possible global solutions.

This classical two-frame approach presents some limitations when it comes to characterizing how correspondence matches are formed in typical perception. For example, Ramachandran and Anstis (1987) showed an effect they termed *visual inertia* in which the first frame-change in a three-frame sequence strongly constrained the percept induced by the second-frame change. Namely, the motions perceived in the latter frame-change were typically the ones that were collinear with the perceived mo-

tions in the first frame-change. In this dissertation, both classical two-frame stimuli as well as displays with three or more frames (including but not limited to the visual inertia display) will be explored in relation to the model.

The rest of this chapter is structured as follows. First, a brief review of the most well-known and successful previous approaches to the problem will be presented. The limitations of these earlier approaches will be highlighted and compared with the current approach. Second, a perceptual principle, the unique split/fusion principle, will be proposed and described. Third, two versions of a dynamical neural network model for solving the correspondence problem will be developed and explored. The first is a simplified version of the model that only accounts for one-dimensional motion patterns, for which all stimuli fall on a line. The second model generalizes the first in order to account for two-dimensional motion patterns. The neural networks' response to both novel and benchmark correspondence displays will be discussed in order to evaluate the degree of agreement between the model and human perception, as well as to identify which aspects of the model are responsible for the emergent percepts formed under various conditions (i.e. human-like responses to some displays necessitate all the features of the model, while other displays highlight only a subset of the features). Finally, a discussion of the relation to other models as well as testable predictions are discussed.

5.2 BRIEF REVIEW

5.2.1 Ullman's Minimal Mapping Theory

The most well-known treatment of the motion correspondence problem is probably Ullman's (1979) *minimal mapping theory*. Ullman formulated the problem in terms of minimizing a cost function associated with forming competing correspon-

dence matches. The cost function applied two constraints to the problem: the nearest neighbor principle and the good cover principle. The nearest neighbor principle assigns larger costs to longer-range motions (such that shorter-range motions are preferred), and the good cover principle requires that each element on each frame has at least one correspondence match associated with it (this prevented the otherwise degenerate minimal-cost solution of no motion paths).

Ullman’s model accounted for a set of simple correspondence displays, but quickly fails when the complexity of the stimulus increases. This is especially evident in ‘group motion’, when multiple elements move together in an invariant formation. The failure occurs because each motion path is treated independently when minimizing the cost function. For a more detailed overview of the model’s shortcomings in response to specific correspondence displays, see Dawson (1991).

Additionally, the model did not propose a neural process by which the cost-function is minimized, nor did it specify the process by which visual elements are extracted from the stimulus and their locations measured continuously. This is a general problem with the concept of *token-tracking*, which is discussed in more detail below.

5.2.2 Dawson’s Hopfield Network

Dawson (1991) developed a neural network model (a Hopfield network; Hopfield, 1982; Hopfield and Tank, 1985) in order to overcome some of the limitations of Ullman’s minimal mapping theory. In Dawson’s model, three constraints were simultaneously applied via synaptic weights: the nearest neighbor principle, the relative velocity principle, and the element integrity principle. In the Hopfield network, each neuron represents a motion path, and there is all-to-all connectivity between every neuron. The strength of the synaptic connectivity between any pair of neurons is a com-

combination of the three principles, where parameters can vary the weighting of each constraint (i.e. how influential it is on the evolution of the system). For each correspondence display, the network is initialized with all potential motion paths set above threshold (by the system operator). The network then evolves according to a discrete dynamical updating rule, and eventually a stable solution is reached (guaranteed by the symmetrical connectivity inherent in the Hopfield net). After a sufficient number of iterations, the state of the network determines the particular solution to a given classical correspondence problem.

In many ways, the model developed here shares similarities with Dawson's (1991) correspondence network. Firstly, his was the first thorough treatment of the motion correspondence problem in terms of a neural network solution. This is a step towards biological plausibility from the abstract cost-minimization approach of Ullman. Additionally, his three constraints have significant conceptual overlap with the current model which specify 1) a constraint on the extent of splitting and fusing, 2) a tendency for shorter-path motions with all other things being equal, and 3) a collective effect of group motion when multiple visual elements can be interpreted as being translated rigidly.

There are, however, several notable theoretical advances in the current formulation. Whereas Dawson's solution to the problem necessitated that a new network be instantiated for each problem, the present model is formulated as a general solution, new networks are not needed for each problem. Furthermore, no front-end was specified in Dawson's model, which takes for granted the existence of 'token-trackers' that can extract and track the locations of figural elements over time (this is discussed in more detail below). A front-end is fully-specified in the current model along with the mechanism for detecting motions from the dynamic stimulus array. Another consequence of Dawson's formulation is that his model must be re-set for each cor-

respondence display, even if the stimulus is the same. This is because the network evolves to a stationary solution, and remains there indefinitely until it is externally modified. In the model below, solutions are transient events that are not stabilized indefinitely, but persist on the timescale of hundreds of milliseconds. This means that the model can address cases in which multiple motion-events (i.e. frame-changes) are presented in sequence without running into conceptual predicaments.

5.2.3 Token-tracking

Both Ullman (1979) and Dawson (1991) make use of the notion of *token-tracking* in order to explain the generation of the set of potential motion signals out of which the perceived motions are a subset. Essentially, the idea is that at each ‘moment’ the visual system is segregating a set of figures from their background, and isolated figures are treated as *tokens*, whose locations are cognitively ‘tagged’, continuously monitored, and updated. When a token appears at a new location, it is (under appropriate conditions) perceived as sharing an identity with a token at a previously occupied location, and a motion percept from one token to the other is formed. There are several shortcomings to this approach.

In this view, a stationary visual token must have its location continuously reaffirmed in order to be perceived as stationary. At each moment, a token takes place in a correspondence calculation and, under appropriate conditions, it is measured to identify with itself. This is not necessarily a problem, but it must be noted that it would be a rather costly process when viewing a stationary image.

Further, token-tracking approaches do not specify *how* visual tokens are to be extracted and tracked; they take these processes for granted. This manifests as a theoretical disconnect between token-tracking and what is typically considered low-level motion detection.

The concept also breaks down when one considers the case of generalized apparent motion (GAM). In GAM, two visual elements are simultaneously visible on each frame, yet motion is seen to originate from one element location and end on another under conditions of counterchange (Johansson, 1950; Hock, Kogan & Espinoza, 1997). It has been demonstrated by Hock, Gilroy and Harnett (2002) that other simple low-level feature tracking mechanisms cannot account for the perception of motion in GAM, for example keeping track of the element with higher absolute contrast. Motion is perceived even if the location of maximum contrast does not change during counterchange. Thus, GAM presents a considerable challenge to accounts reliant on token tracking.

5.2.4 Hard and soft constraints

Most approaches to the correspondence problem have sought solutions by applying constraints to a set of potential motion paths, but how these constraints are imposed varies from model to model. When applying constraints to search for a solution to a problem, a distinction is often made between *hard* and *soft* constraints. Hard constraints describe those conditions which must be met by the system of interest (i.e. given absolute priority) while soft constraints describe conditions which are preferred, but are not always met and will yield to the demands of hard(er) constraints. Soft constraints are more like tendencies or biases of a system, rather than rules which are always obeyed. A soft constraint will always be violated in the interest of conforming to a hard constraint. They allow malleability in the kinds of solutions reached compared with hard constraints, and are more plausible in the face of counterexamples. A single counterexample calls into question the viability of a hard-constraint. In the visual domain, a hard-constraint would correspond to something *never* seen. Evaluating hard-constraints puts an emphasis not only on what percepts tend to be seen

when exposed to a given stimulus, but also what set of percepts are possible (even if infrequent) and, importantly, seemingly impossible for that stimulus.

Three constraints were instantiated in the present model. There are two soft constraints: a bias for nearest neighbor matchings and mutual enhancement motion signals with the same span (where span in two dimensions refers to both the distance and direction of a motion path), and one hard constraint: the unique split/fusion principle. The unique split/fuse principle is described conceptually below, and the implementations of all three constraints are discussed in detail in the modeling section.

The unique split/fusion principle

In this chapter, a hard-constraint is proposed in the context of ambiguous motion perception: the unique split/fusion principle. Solutions to a number of correspondence problems are found that are in agreement with human perception when coupling this constraint with other empirically-supported tendencies (soft constraints).

Some terminology will be introduced in order to aid in the description of the principle. As discussed, a counterchange motion signal is (potentially) generated from the location of a decrease in (spatial filter) activation to the location of an increase in (spatial filter) activation. It is assumed that multiple motion detectors may share a change-detection subunit; i.e. multiple counterchange motion paths may originate or terminate at a common location. When multiple (perceived) motion paths share a decrease subunit (i.e. they originate at the same location) it is referred to as *splitting*. When they share an increase subunit (i.e. they terminate at the same location) it is referred to as *fusing*. Two motion paths are said to be *independent* if they share neither a decrease nor an increase subunit, and conversely are said to be *dependent* if they do share a change detection subunit.¹ The unique split/fusion principle asserts

¹This is not to be confused with *statistical* dependency. Here, we refer to *hierarchical dependency*,

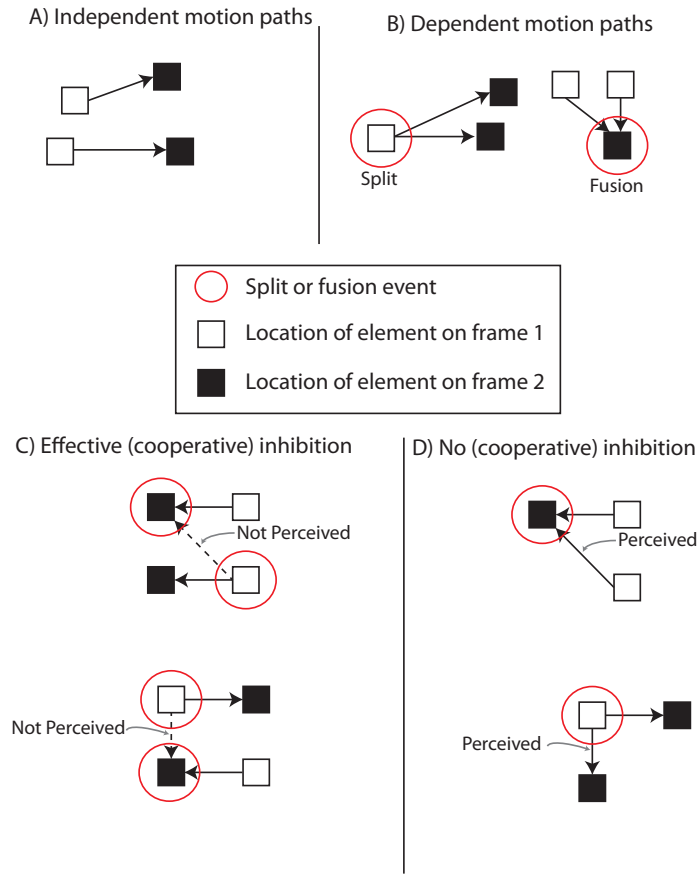


Figure 5.1: Some local motion patterns and examples of the split/fusion principle. Panel A) shows two independent motion paths; they share no subunit (increase- or decrease-detector). Panel B) shows two pairs of dependent motion paths. The two on the left (right) share a decrease (increase) subunit, resulting in a split (fusion) event. Panel C) shows two patterns that invoke the unique split/fusion constraint. The solid arrows represent perceived motion paths. The dashed arrow represents an unperceived motion path that, if it were perceived, would result in a splitting event in the lower right and a fusion event in the upper left, both circled in red. The example below it is arranged differently but follows the same logic. Panel D) shows variants of panel C) with a single element removed on frame one. No motion path connects a split and a fusion, circled in red, and thus all paths are perceived.

that *no perceived motion path can both originate from a location of splitting and terminate at a location of fusion* (for a single motion event). A motion path may

where two (pattern detection) elements are considered to be independent if they do not depend on common subunits.

be seen to take part in splitting *or* fusing during each motion event, but not both. In other words, no motion perceived path *connects* a splitting event with a fusing event, making splitting and fusing events *unique*. Figures 5.1 and 5.2 help to describe this principle. The potential functionalities of such a constraint are explored in the discussion at the end of this chapter where it is argued that the perceptual exclusivity between approaching and retreating objects may provide ecological benefit.

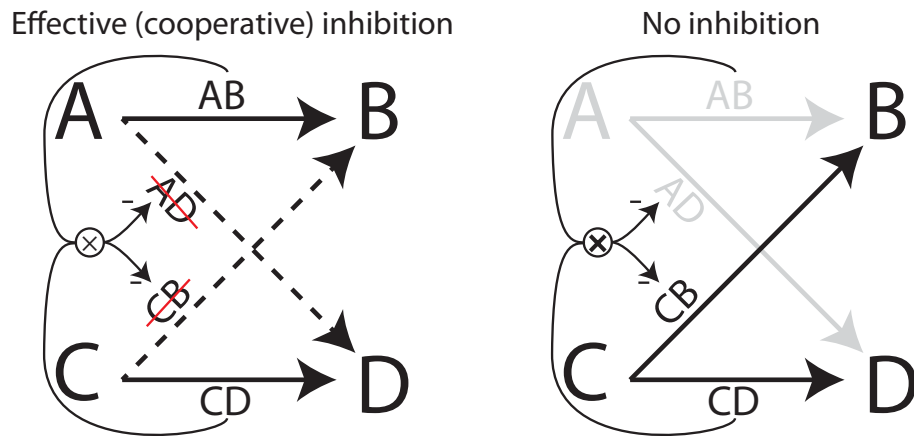


Figure 5.2: Diagram of cooperative inhibition in a simple correspondence problem. On the lefthand panel, independent motion paths AB and CD cooperatively inhibit motion paths AD and CB, each of which is mutually dependent with the pair AB and CD. The righthand panel shows a case where no cooperative inhibition is invoked. The thick solid arrows represent active counterchange motion paths, the dashed arrows represent inhibited paths where there would otherwise be counterchange motion. Motion paths in light grey are inactive.

This is an abstract principle, and requires further specification as a mechanism if it is to be embodied in a model. A general mechanism is proposed to account for the principle: two independent motion paths *cooperatively inhibit mutually-dependent motion paths* with which they each share a subunit. Here, cooperative inhibition implies that there is only effective inhibition when both independent motion paths are sufficiently stimulated.²

²The concept of cooperative inhibition does not appear to be widely considered in neuroscience.

A given motion path is said to be mutually-dependent with a pair of independent motion paths if it shares a subunit with each of them. By definition, two independent paths cannot share a subunit. Therefore, because the motion detector for each path is fed by two subunits, the mutually-dependent path shares one of its subunits (e.g. the decrease detector) with one of the pair of independent motions and the other subunit (e.g. the increase detector) with the other.

In other words, a motion path that is mutually-dependent with two independent motion paths originates at the same location as one of the independent paths and terminates at the same location of another. It logically follows from this that each pair of independent motion paths (potentially) inhibits exactly two (mutually-dependent) motion paths: from the decrease-detector of one (independent) path to the increase-detector of the other, and vice versa.

The left panel of Figure 5.2 shows an example of the unique split/fusion constraint in the context of a relatively simple motion correspondence problem. Two decrease events occur at locations A and C, and two increase events occur at locations B and D. Potential counterchange motions therefore exist from A to B (AB), C to D (CD), A to D (AD), and C to B (CB). AB and CD are independent from one another because they do not share any subunits, and AD and CB are independent from one another as well. The pair of paths AB and AD are dependent because they share a decrease subunit (split), as do the pair CD and CB. Additionally, the pair AB and CB and the pair AD and CD each share a location of increase, and are therefore dependent pairs (fusion). The independent paths AB and CD (potentially) cooperatively inhibit the mutually dependent paths AD and CB, and complementarily, AD and CB have the potential to cooperatively inhibit AB and CD. Assuming some competitive advantage

However, it has been observed and conceptualized in physiological research; see, for example, Cardiello et al., (1998); Cloutier, (1999); Murakami et al. (2003).

is conferred to AB and CD (e.g. a shorter-path advantage), the pair of motion paths AD and CB (each of which is mutually dependent with AB and CD) will both be cooperatively inhibited.

The right panel of Figure 5.2 shows a case with no effective (cooperative) inhibition. It shows a variant of the same stimulus without a decrease event at A (i.e. a decrease at C and two increases at B and D). The cooperative inhibitory circuit would not have any significant causal influence and motion paths CB and CD would both remain activated.

The implementation of this cooperative inhibitory mechanism is detailed in the section describing the modeling below and in Appendix C. Importantly, the scheme is general and it is not necessary to pre identify where pattern elements and potential motion paths might occur a priori. Cooperative inhibition emerges where appropriate out of the architecture of the network.

5.3 THE MODEL

The neural network is composed of four interconnected sub-network components, the two change-detection arrays, the motion detection array, and the biasing array (Figure 5.3). The input to the network consists of a spatially-filtered dynamic gray-scale image. The input feeds into parallel pathways, one pathway feeding-forward to an array of decrease detectors, and the other pathway feeding-forward to an array of increase detectors. These two arrays feed-forward to the motion-detection network where counterchange motion signals are both generated and undergo competitive selection (entailed by the unique split/fusion principle). The motion-detection network is reciprocally connected to a network that integrates global information about the distribution of (potential) motion signals, and biases the generation of the motion

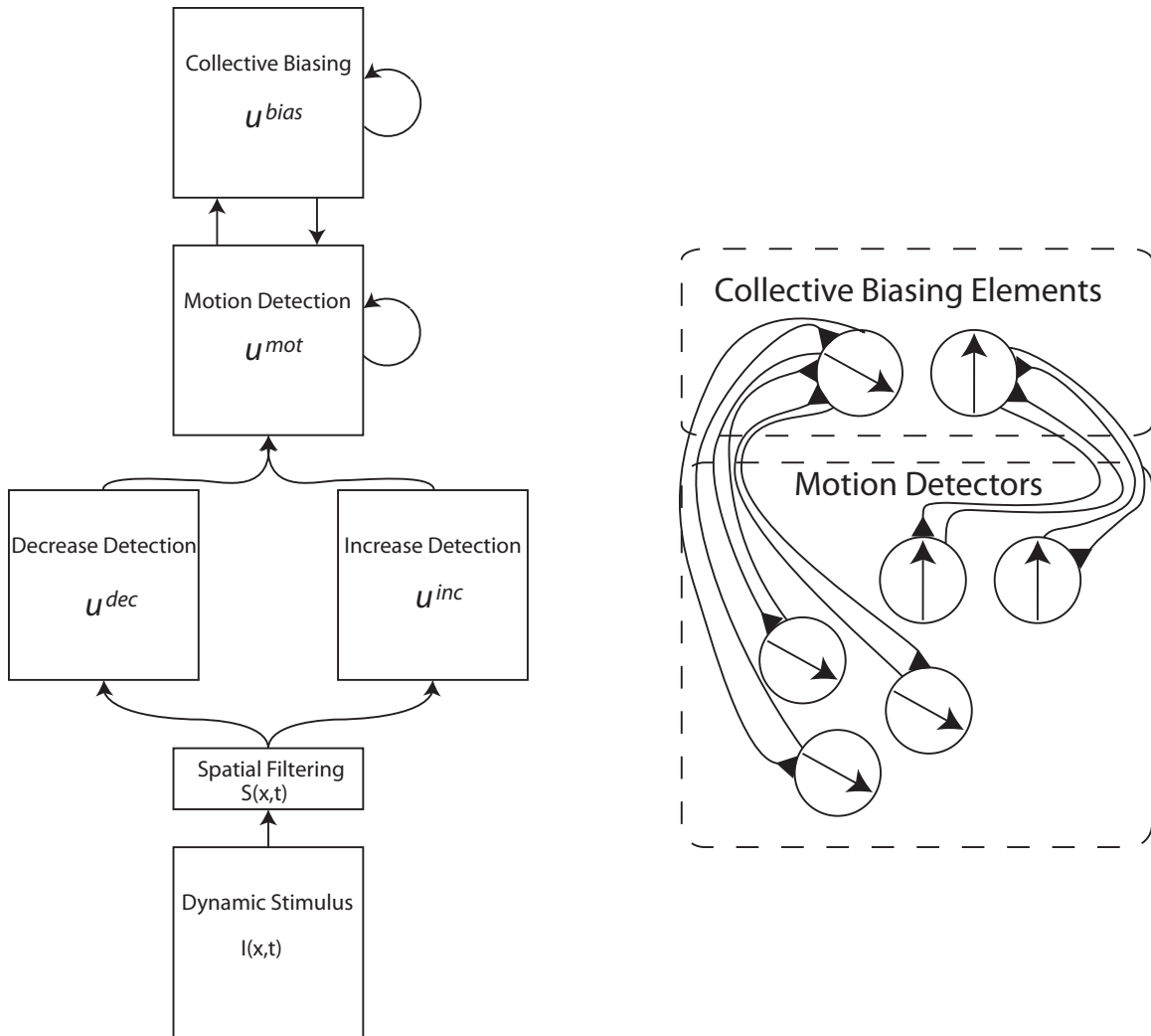


Figure 5.3: A wiring schematic of the correspondence network. A dynamic stimulus is spatially filtered and fed in parallel to increase- and decrease-detection arrays. These arrays are combined into a counter-change motion-detection network where motion signals are generated and competitive selection takes place. The motion-detection network is reciprocally connected to a biasing network. Each class of motion-detector (defined by their direction and magnitude, ie. ‘span’) converges on a detector that feeds excitation back to all the motion-detectors of that class that biases selection in their favor.

signals on the motion-detection network, referred to as the *biasing array*. The biasing array also undergoes a process of competitive selection. It is internally structured as a soft winner-take-all network where each neuron inhibits all others in the biasing

array sub-network³. Excitatory feedback to the motion-detection array is therefore limited to the ‘winners’ of this competition. Each of these components will be described in more detail below along with illustrative examples. Their interaction and emergent behavior will be discussed on a case-by-case basis, highlighting theoretically critical aspects. Except where otherwise noted, all model parameter values are held constant for each simulation (listed in Appendix C). For some simulations, certain components and interactions will be deactivated in order to address the necessity of each component, it will be stated explicitly when this is the case.

Two versions of the dynamical model are presented below. Mathematically, they take almost exactly the same form, with the major difference being whether one- or two-dimensional motion correspondences can be computed. The one-dimensional motion model is presented first as it is easier to conceptualize and visualize and therefore instructive of the behavior of the model in general. Several one-dimensional stimuli will be discussed on a case-by-case basis, highlighting relevant features of the model where appropriate. The model will then be formulated for the case of two-dimensional motion, and again several cases will be discussed.

5.3.1 One-dimensional correspondence network

Stimuli

The dynamic image input $I(x, t)$ consists of N locations (pixels) $x = [x_1, x_2, \dots, x_N]$ whose local intensity can vary from 0 (i.e. black) to 1 (i.e. white) and T time samples $t = [t_1, t_2, \dots, t_T]$ (each time sample representing one millisecond). Spatial intensity patterns remain static for multiple time samples before undergoing a discrete change to a different pattern, generating correspondence ambiguities and potentially eliciting

³The winner take-all network is said to be *soft* because it is not guaranteed that only one neuron will stabilize above the interaction threshold, as can be seen in some of the example cases below.

apparent motion percepts. In the following simulations all display ‘frames’ last for 200 time-samples. The input pattern $I(x, t)$ is convolved with a simple 1-dimensional center-surround spatial filter $c(x) = [-.5 \ 1 \ -.5]$ and scaled by a factor ϵ , which produces the time-dependent spatial function $S(x, t) = \epsilon(I(x, t) * c(x))$ (in all of the simulations, the result of the convolution is truncated in order to maintain the number of locations N in the original input vector $I(x, t)$).

Change-detection arrays

The detection of motion is achieved via a counterchange mechanism that detects motion from locations of decreasing spatial filter activation to locations of increasing spatial filter activation. The spatially-filtered stimulus is passed in parallel to two neural arrays, one which responds with excitation to decreases in activation, and the other which responds with excitation to increases in activation. They evolve according to the systems of equations 5.1 and 5.2, respectively.

$$\begin{aligned}\tau \dot{u}_m^{dec} &= -u_m^{dec} + h^{dec} - v_m^{dec} - S(m, t) + \xi^{dec}(m, t) \\ \tau_{slow} \dot{v}_m^{dec} &= -v_m^{dec} - S(m, t)\end{aligned}\tag{5.1}$$

$$\begin{aligned}\tau \dot{u}_n^{inc} &= -u_n^{inc} + h^{inc} - v_n^{inc} + S(n, t) + \xi^{inc}(n, t) \\ \tau_{slow} \dot{v}_n^{inc} &= -v_n^{inc} + S(n, t)\end{aligned}\tag{5.2}$$

Where each location indexed by subscripted variables m and n on each array, respectively, receives input from the corresponding location x along the spatially-filtered stimulus array $S(x, t)$. For example, the increase detection unit u_4^{inc} receives input from $S(4, t)$ for all t . The superscripted indices *dec* and *inc* refer to the decrease and increase sub-network components.

Feedforward counterchange detection

The two 1-dimensional change-detection arrays project feed-forward connections to the 2-dimensional motion-detection array. Each decrease detection element is paired with each increase detection element, delivering input via a multiplicative synapse to a motion-detection element. The motion-detection lattice is indexed such that the motion detector $u_{m,n}^{mot}$ receives input from u_m^{dec} and u_n^{inc} where the superscripted index *mot* denotes the motion-detection sub-network. In other words, the decrease detection array projects its synaptic outputs across each row of the motion-detection lattice, and the increase-detection array projects its synaptic outputs across each column of the motion-detection lattice. For example, the decrease detector u_2^{dec} projects synaptic input to $u_{2,n}^{mot} \forall n$ and the increase detector u_3^{inc} projects input to $u_{m,3}^{mot} \forall m$. If both of these change-detectors are active their synaptic projections will coincide at the motion-detector $u_{2,3}^{mot}$. Figure 5.4 shows this example in graphical form. Equation 5.3 describes the motion-detection network with only feedforward connections (and a noise term ξ) scaled by the synaptic weight A which is the same for all motion detectors.

$$\begin{aligned} \tau \dot{u}_{m,n}^{mot} &= -u_{m,n}^{mot} + h^{mot} \\ &+ A \cdot f(u_m^{dec}) \cdot f(u_n^{inc}) \\ &+ \xi^{mot}(m, n, t) \end{aligned} \tag{5.3}$$

Because of the multiplicative nature of the synapses, effective input is only delivered to a motion-detector unit if it is receiving input from both of its subunit inputs. In the example above (Figure 5.4), $u_{2,4}^{mot}$ would not receive effective input because although it is receiving input from u_2^{dec} , it receives no input from u_4^{inc} . The parameter A sets the feedforward synaptic weight for all counterchange motion-detectors.

The two-dimensional motion-detection lattice can be thought of as a neural prod-

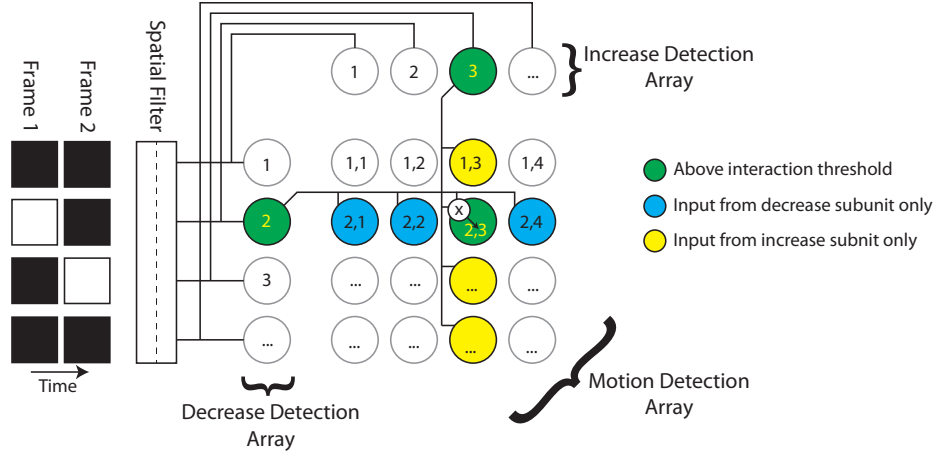


Figure 5.4: Feedforward architecture with some connections shown. The spatially-filtered image is fed-forward in parallel to the decrease and increase detection arrays. Each motion detector receives input from one decrease and increase detector which are multiplied, computing motion from the location of its decrease detector input to its increase detector input. In the example illustrated in the diagram, the blue nodes represent motion detectors receiving input from a decrease detector but no increase detector, and the yellow nodes represent motion detectors receiving input from an increase detector but no decrease detector, and therefore there is no effective input. The green motion detector ($u_{2,3}^{mot}$) is the only active node as it receives input from both a decrease (u_2^{dec}) and an increase detector (u_3^{inc}).

uct space of the two one-dimensional change-detection spaces. Each motion-detector $u_{m,n}^{mot}$ associates a decrease at location m and an increase at location n ; in other words, a counterchange motion from m to n . In this sense, the motion-detection lattice could be conceived as embodying a space-code, where a neuron's location within the lattice signals its perceptual interpretation.

When only one decrease and one increase are detected at a time, a single counterchange motion signal is generated. When there is transient activation in more than one decrease or increase detector at a time, multiple counterchange detection signals are generated, from each decrease detection location to each increase detection location (assuming there is sufficient activation). Typically, only a subset of these

motion paths are perceived. The correspondence problem makes itself evident here: a counterchange motion is detected from each motion-detection neuron where a row intersects with a column that are each receiving inputs from their corresponding decrease and increase detection subunits, respectively. A selection process is therefore needed for the model to reflect human perception.

Competitive selection via cooperative inhibition

In order to select among motion signals, a competitive scheme is instantiated that embodies the unique split/fusion principle described in Section 5.2.4. Recall that each row of the motion-detection lattice receives input from a single decrease detector, and each column from a single increase detector. This implies that motion-detectors in the same row are related to one another by virtue of the fact that they share a common decrease-detection subunit, and similarly motion-detectors in the same column are related by their common increase-detection subunit.

Recall that two motion paths are said to be *dependent* if they share a subunit and *independent* if they do not. Thus, a synaptic neighborhood is established for each motion-detection neuron based on its shared subunits. A given motion-detector's neighborhood is the collection of all motion paths with which it is dependent and is therefore composed of two sub-neighborhoods: all the motion-detectors it shares a decrease-detector with (i.e. all elements in the same row) and all the motion-detectors it shares an increase detector with (i.e. all elements in the same column). In complement, all of the neurons not within a neuron's neighborhood represent motion paths that are *independent* from the given neuron's motion path.

Recall further that according to the unique split/fusion principle, a given motion path is co-inhibited by two independent motion paths, both of which it is dependent with. In other words, two independent motion paths co-inhibit motion paths that

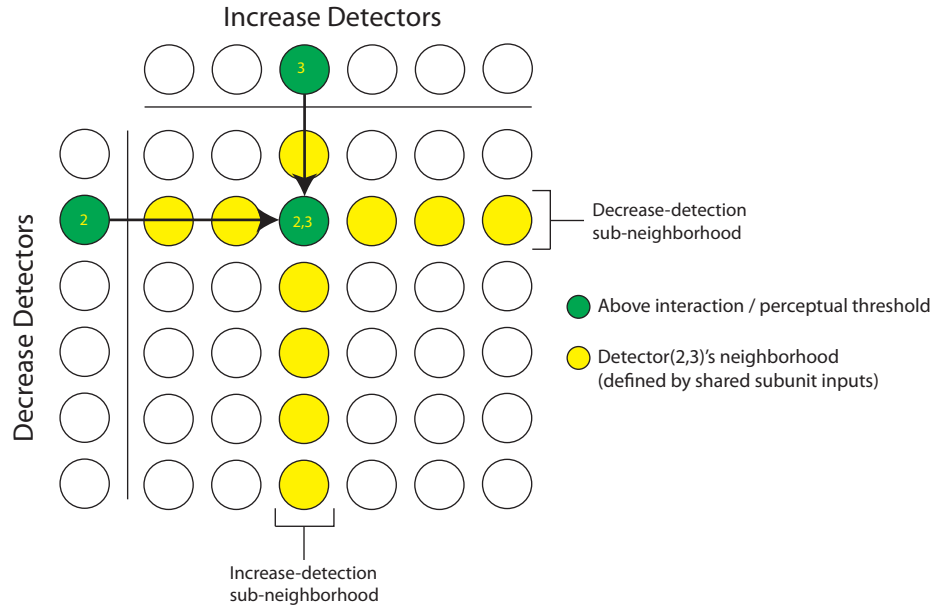


Figure 5.5: A motion detector’s neighborhood is composed of all of the motion-detectors that it shares an increase or decrease subunit with. It is further subdivided into its increase- and decrease-detection sub-neighborhoods.

are mutually-dependent. In terms of counterchange this implies that a motion path from a given decrease-location to a given increase-location will be inhibited by the coincidence of two separate motion paths, one of which shares the location of decrease (origination), and the other the location of increase (termination).

This scheme also implies that sets of dependent motion paths do not inhibit one another by themselves. In the motion-detection lattice space, this means that for a neuron to be effectively inhibited, it must sit at the intersection of the neighborhoods of two active, independent motion-detectors. For this to be the case, it must receive synaptic input from at least one neuron in both the same row *and* the same column. This will be shown to be an essential feature of the model.

This interaction is embodied in the network by a term in the equation that is the product of two sums, each sum representing a sub-neighborhood of a neuron (Equation 5.4). The first sum represents all the summed interaction state of all the

neurons that share a decrease detector with the neuron of interest (i.e. all the neurons in its row except for itself); the second sum, all the neurons that share an increase detector subunit (i.e. all the neurons in its column except for itself). These sums are multiplied with one another and scaled with a negative (inhibitory) synaptic weight $B < 0$. If either is close to zero the product is also close to zero and the inhibition is not effective. Additionally, because the inhibitory interactions take place through the sigmoidal activation function $f(u)$, it is only when they are supra threshold that competitive interactions can take place. This fact will be shown to be important for achieving the desired dynamics on the network and is related to the application of the unique split/fusion principle as a hard constraint. Equation 5.4 shows the (feedforward) motion-detection network with the cooperative inhibition term added.

$$\begin{aligned}
\tau \dot{u}_{m,n}^{mot} &= -u_{m,n}^{mot} + h^{mot} \\
&+ A \cdot f(u_m^{dec}) \cdot f(u_n^{inc}) \\
&+ B \cdot \sum_{p \neq m} f(u_{p,n}^{mot}) \cdot \sum_{q \neq n} f(u_{m,q}^{mot}) \\
&+ \xi^{mot}(m, n, t)
\end{aligned} \tag{5.4}$$

Biasing Array

By itself, the unique split/fusion principle can account for some motion displays, however it lacks two critical aspects that are evident in motion perception: 1) a preference, all else being equal, for shorter-path over longer-path motions, and 2) a tendency to see group motion when an interpretation allows for a rigid translation of several elements (i.e. multiple counterchange motions of the same direction and span), even if the shortest motion paths are not entailed. The biasing array, which is reciprocally connected to the motion-detection array, integrates activation from

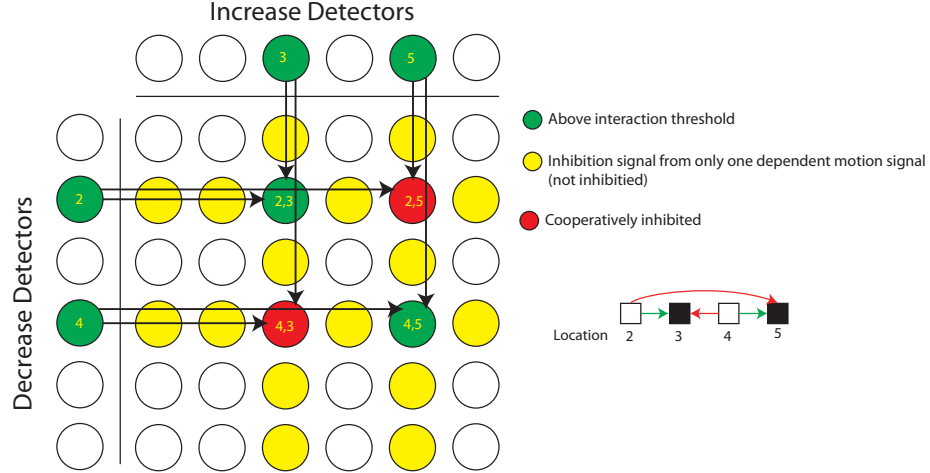


Figure 5.6: An example of the effect of cooperative inhibition. The decrease detectors u_2^{dec} and u_4^{dec} are active along with the increase detectors u_3^{inc} and u_5^{inc} combining to form four (potential) motion paths. In this example, the pair of motion detectors $u_{2,3}^{mot}$ and $u_{4,5}^{mot}$, by virtue of some bias, cooperatively inhibit motion detectors $u_{2,5}^{mot}$ and $u_{4,3}^{mot}$. Under different circumstances, the alternate pair might win out.

motion detectors across the stimulus array and, after undergoing an internal soft winner-take-all competition sends excitation to subsets of motion detectors that gives them a competitive advantage compared to their non-biased counterparts.

The biasing array is organized such that all motion-detectors of a given span (where span refers to both direction and euclidean distance of a motion path) send convergent projections to a single neuron in the biasing array which represents the global activation of motion detectors of that span. The span of a detector is referred to as δ_x , which neurons on the biasing array are indexed according to $(u_{\delta_x}^{bias}$; Equation 5.5). The superscripted index *bias* denotes the biasing array sub-network. δ_x can take on both positive and negative values where, by convention, positive values refer to detectors of rightward motions and negative values to detectors of leftward motions. In general, $\delta_x = n - m$ where m is the index corresponding to the location of decreasing activation and n is the index corresponding to the location of increasing

activation. Thus, a given neuron in the biasing array has its (additive) inputs defined by the term $D \cdot \sum_{n-m=\delta_x} g(u_{m,n}^{mot})$ where the function $g(u)$ is again a sigmoidal activation function, differing from $f(u)$ only by its inflection point.⁴ Projections from the motion-detection array to the biasing array can be visualized as a synaptic projection along the diagonals of the lattice (i.e. the lines parallel to $m = n$), where each diagonal projects convergent input to a single neuron on the one-dimensional biasing array.

The nature of the summed input means that those neurons receiving multiple concurrent inputs from activated motion-detectors of the same span will gain a dynamical advantage as the activation begins to rise compared with those neurons with fewer activated inputs. This is due to the induced attractor being further away from the resting level for those neurons with a greater total input, and thus the rate of change is greater. Because of the winner-take-all dynamics (defined by the term $E \cdot \sum_{r \neq \delta_x} f(u_r^{bias})$, where E is the parameter that sets the strength of the competitive interaction), those with an activation advantage have a better chance of crossing the interaction threshold and providing feedback to motion detectors.

The same-motion detectors that feed into each biasing neuron are the ones that receive excitatory feedback from it (i.e. all motion-detectors of the same span). This is defined in the equation for the motion-detector array by the term $C \cdot f(u_{n-m}^{bias})$. The maximum amount of feedback to any individual motion detector is therefore equal to C . Crucially, this value must be below the value of the effective input needed to induce a motion percept, as the biasing array functions only to, as implied by its namesake, bias potential motion signals such that they gain some advantage during competitive selection. The biasing array should never, by itself, be able to

⁴Namely, the interaction threshold for the projections from the motion-detection array to the biasing array are lower than the other interaction thresholds in the network. This is a necessary feature of the model to achieve proper functioning, which will be discussed in more detail below.

induce a motion percept without corresponding bottom-up evidence originating at the stimulus (i.e., it should not hallucinate). This fact is also one of the reasons that the biasing of same-span motions necessitates a separate subnetwork to mediate the interactions between same-span motion detectors. If there were direct facilitation between motion detectors on the motion detection array, the synaptic weighting of each connection would have to be very small to ensure that self-stabilizing excitation does not form in the network when a handful of motion signals become activated. The saturation of the interaction function eliminates this problem by limiting the amount of facilitation between same-span motion detectors to well-below their interaction threshold, regardless of how many motion detectors are simultaneously active.

$$\begin{aligned}
\tau \dot{u}_{\delta x}^{bias} &= -u_{\delta x}^{bias} + h^{bias}(\delta x) \\
&+ D \cdot \sum_{n-m=\delta x} g(u_{m,n}^{mot}) \\
&+ E \cdot \sum_{r \neq \delta x} f(u_r^{bias}) \\
&+ \zeta^{bias}(\delta x, t)
\end{aligned} \tag{5.5}$$

The biasing array also embodies the nearest neighbor principle.⁵ Simply, the resting levels of each neuron is a function of the euclidean distance of the corresponding motion span such that longer-path motions are further from their interaction threshold at resting levels. This is described by the function $h^{bias}(\delta_x) = \beta + \gamma \|\delta_x\|$ where β (negative) represents a global offset and γ (negative) represents the rate at which the resting level decreases as the euclidean distance of the corresponding motion paths which can be measured as the vector norm $\|\delta_x\|$.⁶ The consequence of this is

⁵This essentially local shortest-path bias could also be expressed directly in the motion-detection network. However the current approach cleanly separates the biases from the processes of detection and selection. Future work should look at the implications of organizing this otherwise.

⁶In the one-dimensional case presented here, $\|\delta_x\| = |\delta_x|$, but this is not true in the two-

that, when all else is equal, biasing neurons that integrate activity from shorter-path motion-detectors have a slight competitive advantage over their longer-path counterparts. As will be shown, this bias is often outweighed by the advantage gained by biasing neurons that receive a greater number of activated inputs during a given stimulus presentation.

Adding a term to the motion-detection network to represent the input of the biasing array leads to an equation of the form:

$$\begin{aligned}
\tau \dot{u}_{m,n}^{mot} &= -u_{m,n}^{mot} + h^{mot} & (5.6) \\
&+ A \cdot f(u_m^{dec}) \cdot f(u_n^{inc}) \\
&+ B \cdot \sum_{p \neq m} f(u_{p,n}^{mot}) \cdot \sum_{q \neq n} f(u_{m,q}^{mot}) \\
&+ C \cdot f(u_{(n-m)}^{bias}) \\
&+ \xi^{mot}(m, n, t)
\end{aligned}$$

Typical network operation

When all the subnetworks are connected, as described above, many seemingly disparate motion percepts can be accounted for. For some correspondence problems, all of the aspects of the network play a role in determining the emergent solution. For others, only some of the dynamics are crucially relevant. Here, a typical sequence of events is described that lead to the network forming a solution.

1. When the (spatially-filtered) stimulus is first presented to the network, the initial set of elements elicits excitation of increase-detection subunits at the corresponding locations. There are only stimulus onsets, so only increases in

dimensional case, so the distinction is necessary.

spatial filter activation, thus no counterchange motion is detected. Given a sufficient amount of time, all detectors return to their resting levels.

2. The stimulus then changes discretely to a new pattern of elements. Some stimulus elements are eliminated and new elements are presented. A set of decrease-detections and increase-detections feed activation forward into the motion-detection array. Assuming sufficient stimulation, the activation of motion-detectors representing paths from each decrease location to each increase location is initiated. Initially, the only bias for a given motion detector is by virtue of noisy fluctuations.
3. When the stimulated motion-detection neurons cross the bias threshold $g(u)$ that feeds excitation from the motion-detection array to the biasing array, the global biasing neurons begin to integrate input from active neurons of their corresponding span. The biasing neurons undergo an internal competition, and the winners that emerge begin to feed back excitation to the motion-detectors of their corresponding span.
4. The motion-detectors receiving feedback from the biasing array gain a competitive advantage over the motion-detectors receiving no feedback. Finally, the motion-detectors that are now potentially receiving both feedforward and feedback activation cross the interaction threshold $f(u)$, and the first subset of signals that satisfy the unique split/fusion principle (i.e. that cooperatively inhibit other potential motion signals) cross the perceptual threshold and become the global percept.
5. If more frames follow with sufficient temporal proximity (i.e., if too much time does not pass allowing all neurons to return to baseline), the feedback from the

biasing array may continue to bias following percepts. Importantly, because of the inhibition of the unperceived motion signals, the biasing array tends to stabilize previously perceived patterns by virtue of ‘future-shaping interactions’, which will be discussed below.

5.3.2 One-dimensional Cases

Except where otherwise noted, simulations were run a minimum of ten times. This is because the equations are stochastic, so not every trial is the same, and often multiple solutions exist to a single stimulus. Where multiple solutions are found to exist, the proportion of time each solution was reached is shown. When not explicitly discussed, it can be safely assumed that the same solution was reached for each individual simulation.⁷

Single Element Apparent Motion

Single element apparent motion is essentially a degenerate case for the network, as no selection of motion signals is entailed (i.e. there is no difficult correspondence problem. Regardless, it is worthwhile to ensure the network works as anticipated, especially for cases that are not necessarily trivial for other proposed solutions to the correspondence problem.

Standard apparent motion. For the network to perceive SAM, only the feedforward pathway is necessary. A decrease at one location occurs when the element turns to the background color (by definition in SAM), a decrease is detected. At the same time, an increase is detected at the location where the element turns from the background

⁷Because the detector responses are transient in nature, an explicit criterion must be used to establish what counts as a ‘solution’. In all of the following simulations, a motion path is considered to be part of the network’s solution if its activity is at least 80% of the maximum interaction output (i.e. $f(u_{m,n}^{mot}) > 0.8$) for at least 20 ms during the interval after the most recent frame-change and before the following one.

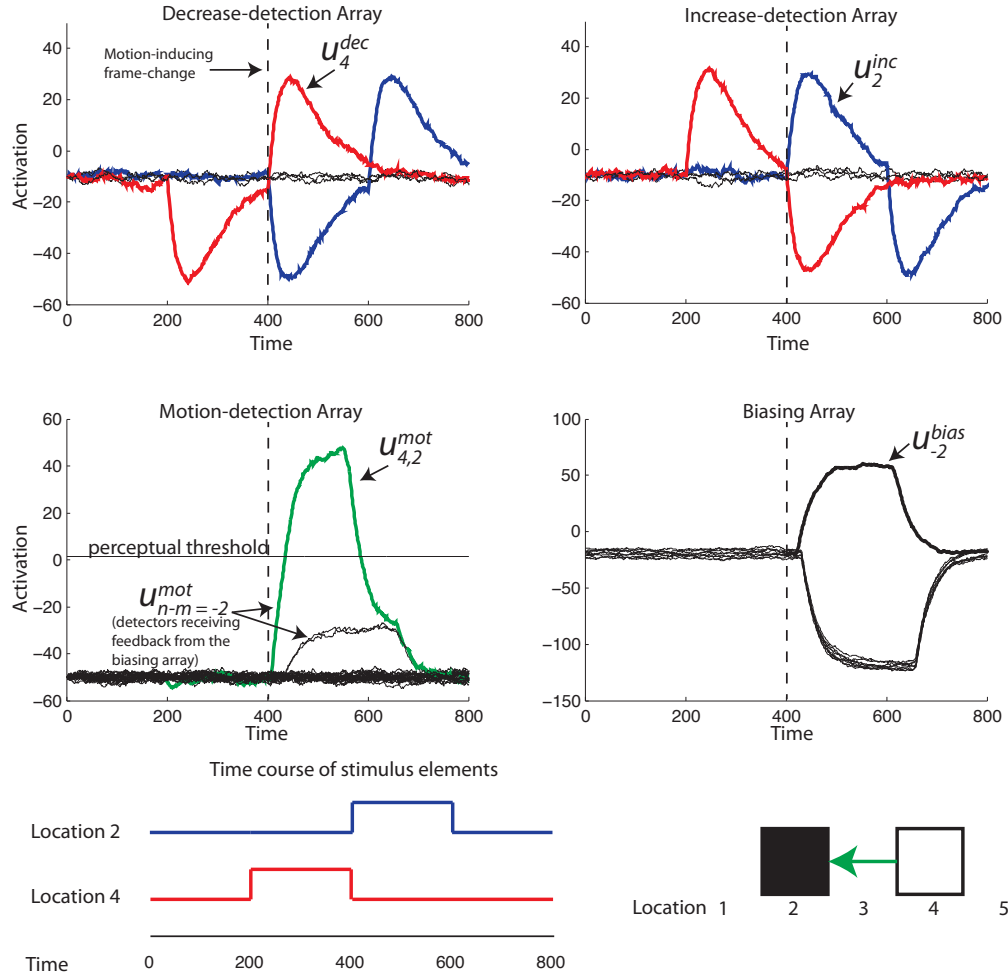


Figure 5.7: Time courses of sub-networks for a single trial of standard apparent motion. An increase-detection at location 4 coincides with an increase-detection at location 2, combining to form a counterchange motion, signaled from location 4 to location 2 (and therefore with a span of -2). The biasing element corresponding to the span of the activated detector feeds back and pre-activates all of the motion detectors with that span, which can be seen rising above the resting level of the other motion detectors.

color (black) to a luminance with high-contrast with the background (white). Given sufficient change in local increase and decrease activation and sufficient time for the neural dynamics to unfold, a motion from the decrease to the increase is always perceived. Despite its irrelevance to the present solution, it is worthwhile to note that

the biasing neuron corresponding to the motion-detectors span δ_x becomes activated and inhibits the other neurons in the biasing array. This activation feeds back and biases the motion-detectors with the same span as the perceived motion; this can be seen in Figure 5.7 where a subset of neurons that do not correspond to any perceived or stimulated motion-detectors sitting above their initial resting level. Given sufficient time after any frame change, all neurons on all the subnetworks will return to resting level.

Generalized apparent motion. In GAM, both elements are visible during both frames, but change their contrast with the background (in opposite directions) during frame changes. The network treats GAM in essentially the same fashion as SAM; as long as there is enough of a change in contrast to elicit sufficient excitation, motion is perceived. The results look nearly identical to the network's response to SAM in Figure 5.7.

Correlational motion

In this display, a stationary high-contrast element is visible on both frames and does not change contrast, during the second frame, a new high-contrast element appears at a remote location (Figure 5.8). It is referred to here as correlational motion because this case (further) differentiates the counterchange detector from spatiotemporal correlators. Because there is an element during Frame 1 at Location A and an element during Frame 2 at location B, the correlational detector will perceive there to be motion from A to B. In contrast, because there is no decrease in activation elicited by a decrease in contrast at Location A, no counterchange motion is specified. Furthermore, depending on the parameters chosen, Dawson's network as well as Ullman's minimal mapping solution will sometimes predict motion to be perceived from the stationary element to the appearing one. Typically in perception, if the elements are

not contiguous, no motion is perceived to originate from a stationary feature. However, Dawson (1991) did note that if the appearing elements were placed *very* close to the stationary ones, motion could be perceived. This is accounted for in the following subsection on the line motion illusion.

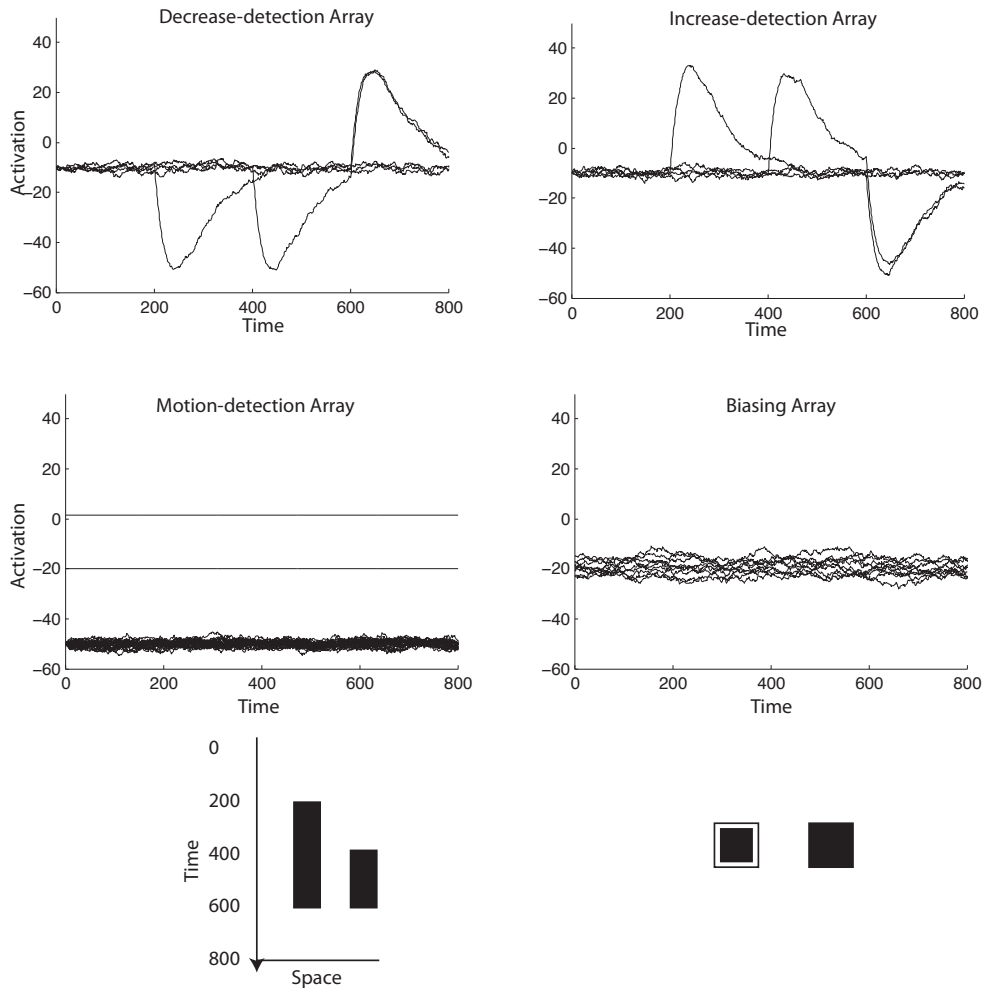


Figure 5.8: Time courses of sub-networks for a single trial for correlational motion. Because there is no decrease in spatial filter activation at the location of the stationary element, no counterchange motion is entailed.

Line Motion Illusion

The line motion illusion (LMI) occurs when a stationary element has its shape altered, typically such that the perceived shape of the element is altered (e.g. stretched). In other words, an element of the same luminance as the stationary element is adjoined to it, with essentially no intervening space between the two. This stimulus typically induces a percept of motion from the stationary surface to the edge of the newly-formed surface that is furthest away from the stationary element.

It has previously been argued that the LMI can be accounted for by counterchange detection (Hock & Nichols, 2010). Here the model demonstrates this account computationally. When the element on Frame 2 appears in the adjacent location to the element on Frame 1, the center-surround filter at the location of the stationary element is somewhat inhibited by the appearance of the new element in its surround-field (Figure 5.9). This inhibition causes an excitatory response in its corresponding decrease-detector, while the newly-appearing element causes an excitatory response in its local increase-detector. The two responses combine to form a counterchange motion signal from the stationary to the newly-appearing element.

Splitting and fusing

When one element is present on Frame 1, and two elements are presented in new locations on Frame 2, counterchange motion is generated from the location of the element on Frame 1, to both of the elements on Frame 2 (Figure ??). Because the two motions are *dependent*, the unique/split fusion principle does not come in to play, and no (cooperative) inhibition suppresses either motion signal; a *splitting motion* is perceived.

One example of a splitting motion is an expanding motion, in which two motion

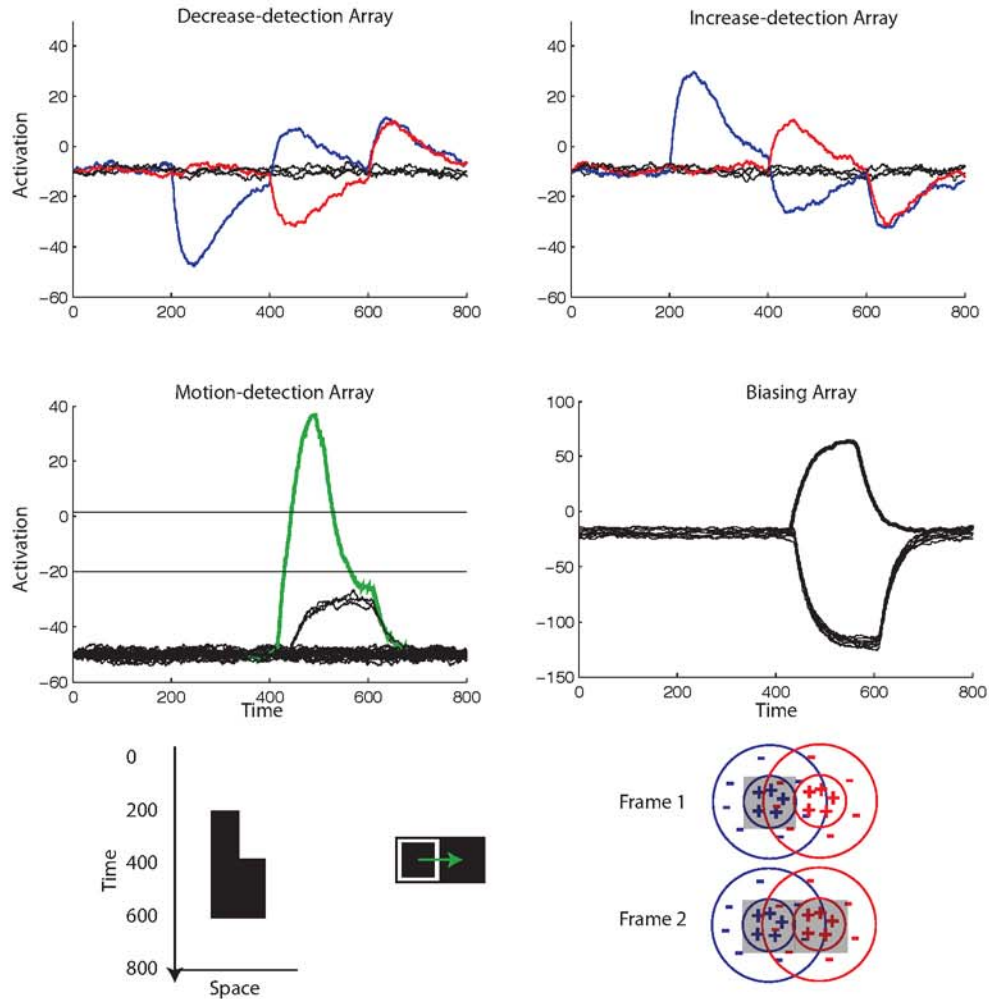


Figure 5.9: Time courses of sub-networks for a single trial one dimensional line motion illusion. The spatial filter at the location of the stationary stimulus is partially inhibited by the newly-presented element on the second frame, leading to a decrease in activation. The spatial filter at the location where the new element is presented undergoes an increase in activation. These two local change-detections combine and counterchange motion signal is formed.

paths are seen to move in opposite directions from a common origin.

In a similar fashion, two elements can be seen to converge in one location during a *fusing motion*. No competition occurs among the motion signals for the same reason as above: the two motions are not *independent*, and so do not cooperatively inhibit any motion signals.

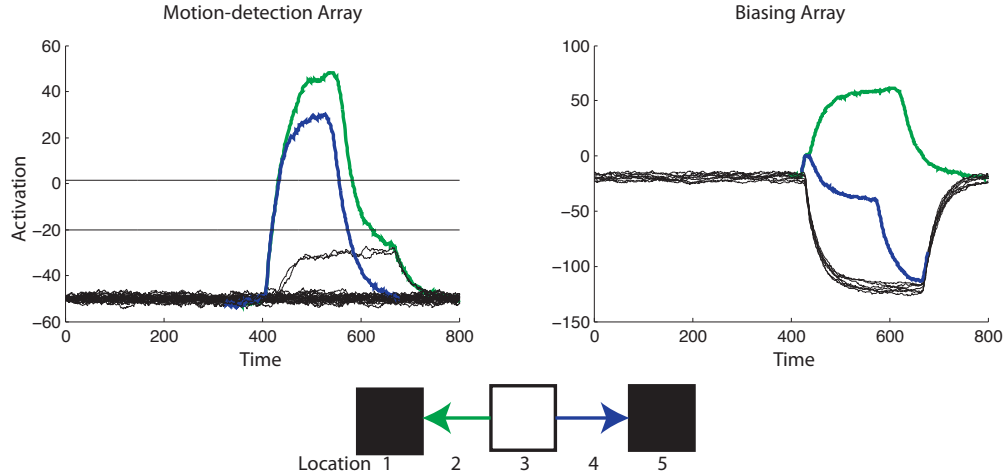


Figure 5.10: Time courses of sub-networks for a single trial of an expansive splitting motion

A simple group motion display

In this two-frame display, both frames contain two visible elements (Figure 5.11). The first frame's elements are at Locations 3 and 7 and the second frame's elements are at locations at Locations 1 and 5. Note that, if the element at Location 7 on the first frame were never displayed, the same splitting solution would be reached as described above. During the initial phase of the trial, sub-threshold counterchange motion signals begin to be generated representing paths from Location 3 to Location 1 and 5, as well as from Location 7 to Location 1 and 5. As the motion-detectors cross the bias threshold, excitation is delivered to the biasing array. The motion detectors $u_{3,1}^{mot}$ and $u_{7,5}^{mot}$ are both relatively short-path motions and are of the same span ($\delta_x = -2$; note that no other two motion paths share the same potential span in this display). Their activity is integrated by the biasing element u_{-2}^{bias} which, by virtue of its relatively high resting-level, and the fact that it is the only biasing element receiving multiple activated motion-detector inputs, tends to win the internal competition on the biasing array. It feeds back excitation to the same two motion

detectors. As they cross the internal interaction threshold on the motion detection network, $u_{3,1}^{mot}$ and $u_{7,5}^{mot}$ cooperatively inhibit motion detectors $u_{3,5}^{mot}$ and $u_{7,1}^{mot}$, which are the other two potential motion signals generated from the stimulus and corresponding change-detection. The biases in this case are strong enough that the same solution emerges for each presentation of this stimulus.

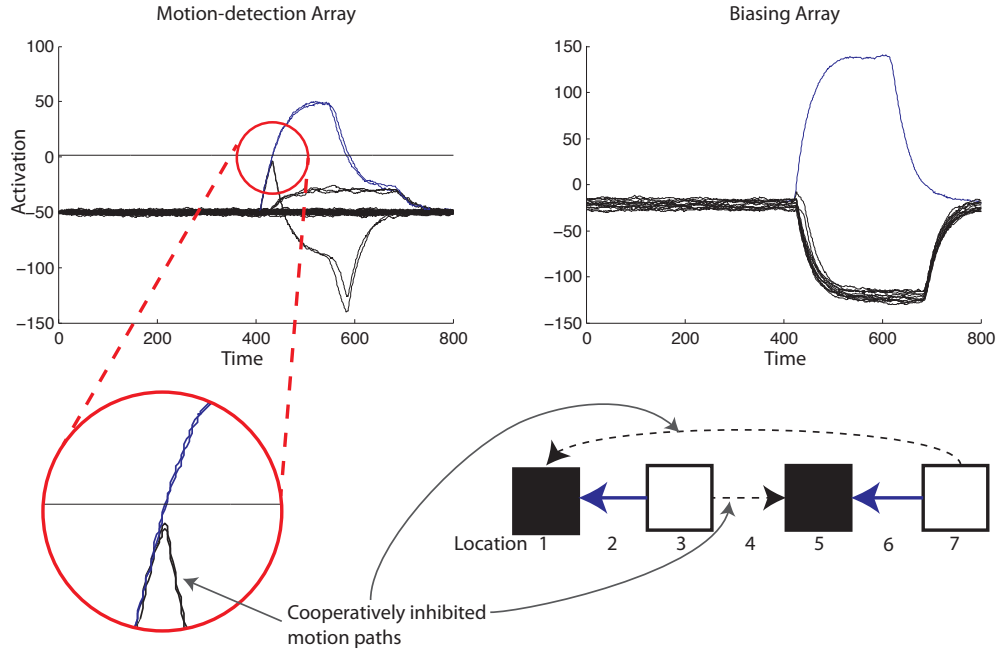


Figure 5.11: Time courses of sub-networks for a single trial one dimensional group-motion with two elements.

As opposed to the previous cases above, this display demonstrates the unique split/fusion principle in action. Combined with the the activity of the biasing array, the two-element group-motion solution emerges over and over. It can be seen that adding any motion path to the solution results in a violation of the principle. If the biasing array is removed, no motion signals are given an inherent advantage before the cooperative interaction takes effect; any differences in activation are due to noise. In this case, the network resolves in the previous solution half of the time, and the

solution entailing the other two motion paths the other half of the time. This is not in agreement with typical human perception, demonstrating the necessity of the biasing array.

5.3.3 Two-dimensional correspondence network

The correspondence network that is capable of handling the more general two-dimensional motion cases is almost exactly the same in form to the network described above for one-dimensional motion. Here, the main differences are described, which mostly consists of extending the dimensionality of each component and does not affect the conceptual scheme of the network.

Stimuli

In the two dimensional version, the dynamic input image $I(x, t)$ is three-dimensional function, with two spatial and one temporal dimension. Spatially, the image consists of $N \times N$ locations. To keep the notation consistent, let x be the duple $x = (x_1, x_2)$. Each frame of the image is convolved with the two-dimensional center-surround spatial filter (again truncating the result to maintain the stimulus size),

$$c(x) = \begin{pmatrix} -1/8 & -1/8 & -1/8 \\ -1/8 & 1 & -1/8 \\ -1/8 & -1/8 & -1/8 \end{pmatrix}$$

producing the spatially filtered stimulus $S(x, t)$.

Change-detection arrays

The change detection arrays also now take a two-dimensional form, corresponding to the two-dimensional stimulus. For the decrease detectors which are indexed by m , let $m = (m_1, m_2)$; likewise for the increase detectors indexed by m , let $n = (n_1, n_2)$. In order to keep the indexing consistent with the row and column format of the arrays,

let m_1 and n_1 stand for the vertical location (distance from the uppermost vertical position, i.e. row number) and m_2 and n_2 stand for the horizontal location (distance from the leftmost position, i.e. column number).

Feedforward counterchange detection

In a similar fashion to the one-dimensional motion network, the two change detection arrays combine to form a kind of neural product space. For each pair of locations to form a corresponding counterchange detector, a four-dimensional motion-detection lattice is entailed. The motion detectors are again indexed by m and n which here represents the quadruple $(m, n) = ((m_1, m_2), (n_1, n_2))$, where a motion detector $u_{m,n}^{mot}$ detects motion from m to n . For example, the detector $u_{2,3,4,4}^{mot}$ represents a motion path beginning at location (2, 3) and ending at location (4, 4).

Competitive selection via cooperative inhibition

In the one-dimensional network, a motion-detector's neighborhood was defined by all the other detectors which share an increase or decrease subunit with the detector, which happened to be all the neurons in its row and its column by virtue of the indexing scheme (with a sub-neighborhood defined as the subset that just share one change-detection subunit or the other, i.e. the detector's row or column only). It was easy to see in this case that two independent detectors' complementary sub-neighborhoods (i.e. the row of one and the column of the other) would intersect at a single location, where cooperative inhibition would occur given sufficient activation of the relevant motion detectors. The other two sub-neighborhoods of the two independent detectors also intersect at a single location. Thus each pair of independent detectors potentially co-inhibit two motion detectors: the one sharing the decrease location with one motion path and the increase location with the other, and vice

versa.

For the two-dimensional motion correspondence network (with a four-dimensional motion-detection lattice), the sub neighborhoods are defined the same way conceptually, but rather than each sub-neighborhood spanning an essentially one-dimensional space (i.e. a row or column), sub-neighborhoods span a two-dimensional subspace of the lattice, defined by all the motion-detectors that share a given subunit. For example, the motion detector $u_{2,3,4,4}^{mot}$ has one sub-neighborhood defined by $u_{2,3,p,q}^{mot} \forall (p, q) \neq (4, 4)$ and the other by $u_{j,k,4,4}^{mot} \forall (j, k) \neq (2, 3)$ (i.e. all the other detectors sharing the decrease subunit $u_{2,3}^{dec}$, and all other detectors sharing the increase subunit $u_{4,4}^{dec}$, respectively).

Although it is not as obvious, each sub-neighborhood of two independent detectors intersect at a single point, as in the one-dimensional correspondence network (and therefore each pair of independent detectors potentially co-inhibit two other motion detectors as they each have two sub-neighborhoods that both intersect at a single location). As an example, if we consider the decrease sub-neighborhood of the detector $u_{2,3,4,4}^{mot}$ and the increase sub-neighborhood of $u_{1,5,2,7}^{mot}$, one can see that their only intersection lies at the location of detector $u_{1,5,4,4}^{mot}$ (with their complementary sub-neighborhoods intersecting at the detector $u_{2,3,2,7}^{mot}$). Thus, the cooperative inhibition scheme that embodies the unique split/fusion principle is conceptually the same as in the one-dimensional case.

Biasing array

Recall that the biasing array is indexed by δ_x . For the two-dimensional motion case, let $\delta_x = (\delta_{x_1}, \delta_{x_2}) = n - m$. Thus, the biasing array in this case is a two-dimensional lattice, where each neuron integrates activity from all same-span motion detectors as in the one-dimensional case, where span implies both the magnitude and direction of

a motion path. Again, by the nature of the formula, δ_x can contain negative values.

5.3.4 Two-dimensional cases

In order to verify the results of the previous one-dimensional cases, each one was run also on the two-dimensional correspondence network. All of the solutions were found to agree with the previous incarnation of the network.

Expanding and contracting

In this stimulus, the first frame has only one visible element, while the second frame contains four elements, positioned above, below, to the left, and to the right of the initial element (Figure 5.12). Therefore, four counterchange motion paths are possible, each originating from the location of the element on the first frame and extending out in the four cardinal directions to the locations of the elements in the second frame. Because the four paths all share the same decrease-detection subunit, they are by definition a set of dependent motion paths. Recall that the mechanism which embodies the unique split/fusion constraint necessitates two independent motion paths to cooperatively inhibit a (mutually dependent) third, thus the conditions are not met.

The contracting stimulus is the same as the expanding stimulus, with the frame order reversed. For the same reasons as above, all four motion paths are seen to converge on a single location.

Another kind of expanding and contracting motion can be seen that is similar to the looming of a physical object as it approaches a perceiver (Figure 5.13). When an object moves closer to the retina the visual angle it subtends increases, resulting in an expansive motion. This can be approximated with a two-frame stimulus in which a central visual element is present on both frames, with all of the surrounding elements becoming visible on the second frame. In a similar fashion to the one-dimensional

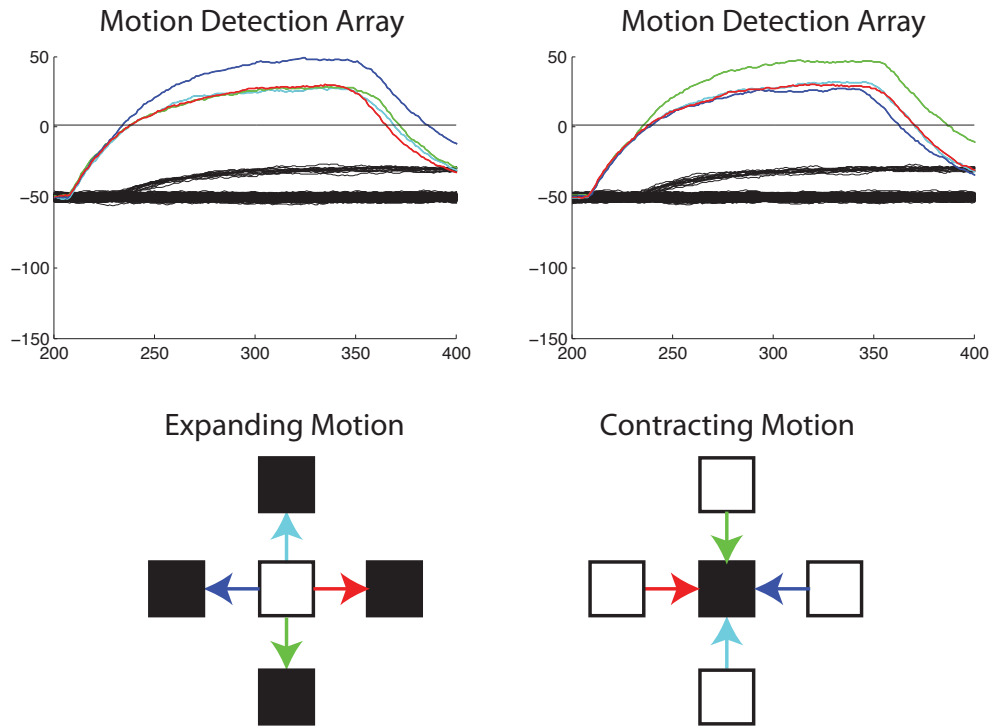


Figure 5.12: Expanding and Contracting Motion

line motion illusion presented above, the central spatial filter undergoes a decrease in activation because of the visual elements occupying its inhibitory surround region during the second frame. The surrounding locations also elicit increases in spatial filter activation at their respective locations, signaling counterchange motions from the central location to all of the surrounding locations. Because all of the signals that are generated share a (decrease) subunit, no cooperative inhibition is entailed.

The inverse contractive stimulus corresponds to a receding rather than approaching object. When all of the elements surrounding the central element are made invisible, decreases in spatial filter activation occur at all of the surround locations, and an increase in activation is elicited by the central element by virtue of the removal of inhibition. Again, all of the motion signals share a (increase) subunit, and therefore there is no competition between them.

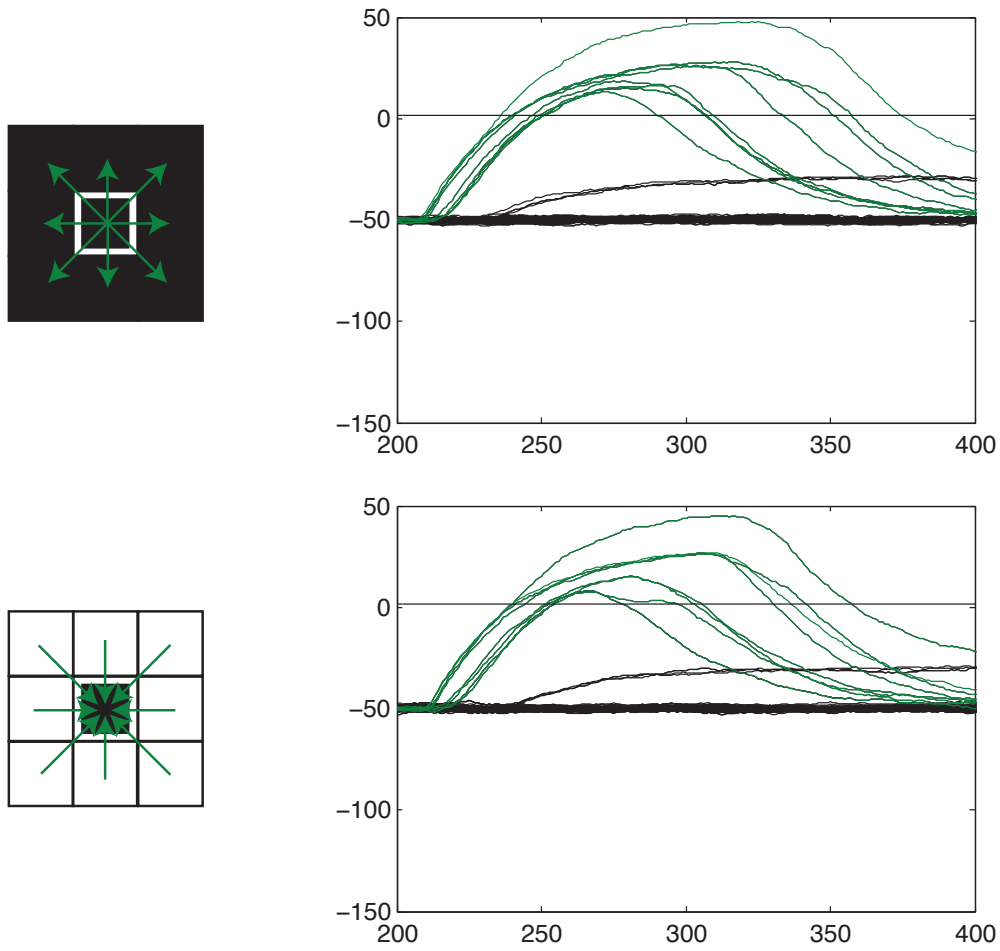


Figure 5.13: Looming and receding motion

Motion Quartet

The motion quartet is a well-studied motion stimulus that exemplifies perceptual bistability. In the standard version (i.e. $BRLC = 2$), each frame has two visible elements that are located in opposite corners of an invisible box (Figure 5.14). When the frame changes, the two visible elements are set to background level (i.e. disappear), and two new elements become visible at the previously unoccupied corners of the (invisible) box. The display may switch between these two frames for an arbitrary number of display cycles. On a given frame-change, human observers report

seeing either the two vertical motion paths (i.e. the ‘upward’ and ‘downward’ motion path) or the two horizontal motion paths (i.e. the ‘leftward’ and ‘rightward’ motion paths), but never any other combination of paths, although they seem to be logically possible. When the aspect ratio of the arrangement is set appropriately, observers will have a 50% chance of seeing vertical or horizontal motion. When the aspect ratio is changed, vertical or horizontal motion may be biased by virtue of the spatial proximity of elements, with closer visual elements being biased to elicit the perception of motion between them; in other words, there is a shorter-path tendency.⁸

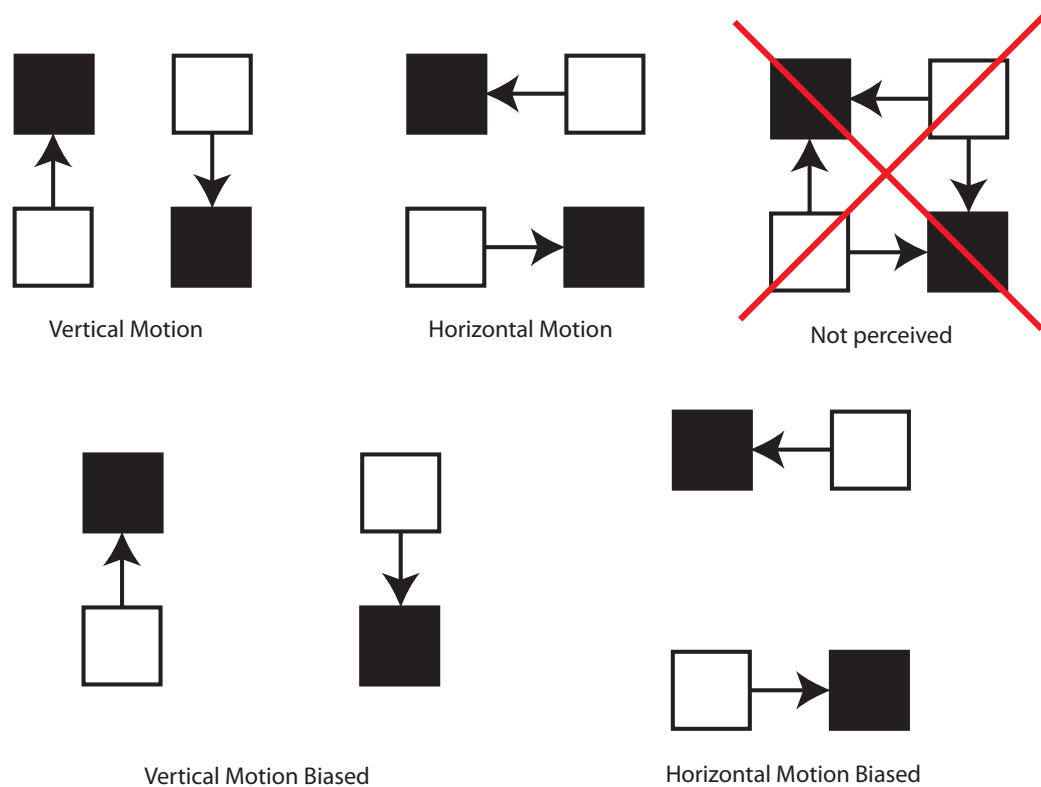


Figure 5.14: Diagram of the motion quartet stimulus

⁸There is evidence that the aspect ratio that elicits perception of vertical and horizontal each 50% of the time is typically not 1:1, with vertical motion biased with such an arrangement. There are also individual differences in the optimal ratio for eliciting 50/50 perception. Here, a simplifying assumption sets the optimal aspect ratio to 1:1.

In addition to each frame-change eliciting perceptual bistability (i.e. either vertical or horizontal motion is perceived in a mutually-exclusive manner), two other dynamic effects are noteworthy in trials that last for several display-cycles: 1) the originally formed percept tends to continue to be perceived on the following frames (e.g. if vertical motion is perceived on the first display-cycle, it will likely also be seen on the next display-cycle), and 2) after a number of display-cycles, perceptual switches from one percept to the other typically occur, and the newly-formed percept also displays some persistence over the following-frames.

While many accounts of perceptual switching appeal to neural adaptation, Hock, Schöner and Voss (1997) have shown empirical evidence that adaptation is neither necessary nor sufficient to account for perceptual switches in the motion quartet paradigm. Instead, they argued that stochastic fluctuations are likely responsible for perceptual switches, and any effects of neural adaptation simply reduce the amplitude of a fluctuation necessary to induce a perceptual switch.

Additionally, the persistence of the vertical and horizontal motion percepts are not accounted for in most proposed explanations of the phenomenon⁹, with the exception of Hock, Schöner, Brownlow, & Taler (2011) who propose a specialized network in which ‘future-shaping interactions’ lead to perceptual persistence by explicitly inhibiting orthogonal motion paths that begin at the location of termination of a perceived motion path.

Here, both persistence of perceived patterns and switching due to stochastic fluctuations are shown to emerge from the network in response to a motion quartet display, although for different reasons than previous models. In order to understand how the network functions as a whole, multiple incarnations will be discussed to show

⁹In fact, motion models that depend on opponency of opposite-direction motions would predict quite the opposite: a reverse motion path should be inhibited by its perceived forward counterpart.

the influence of each aspect of the network.

Bistability in a two-frame motion quartet. If only the feedforward pathway is enabled, the network’s solution entails all possible motion paths. This is clearly insufficient to account for human perception, in which this solution is never perceived. Re-instantiating the competitive dynamics leads to the desired bistability; the so-called vertical and horizontal solutions are perceived on independent trials with an approximately 50% probability (Figure 5.15). This behavior can be intuitively understood by recognizing that including any additional (potential) motion path to either the vertical or horizontal solution results in a violation of the unique split/fusion principle.

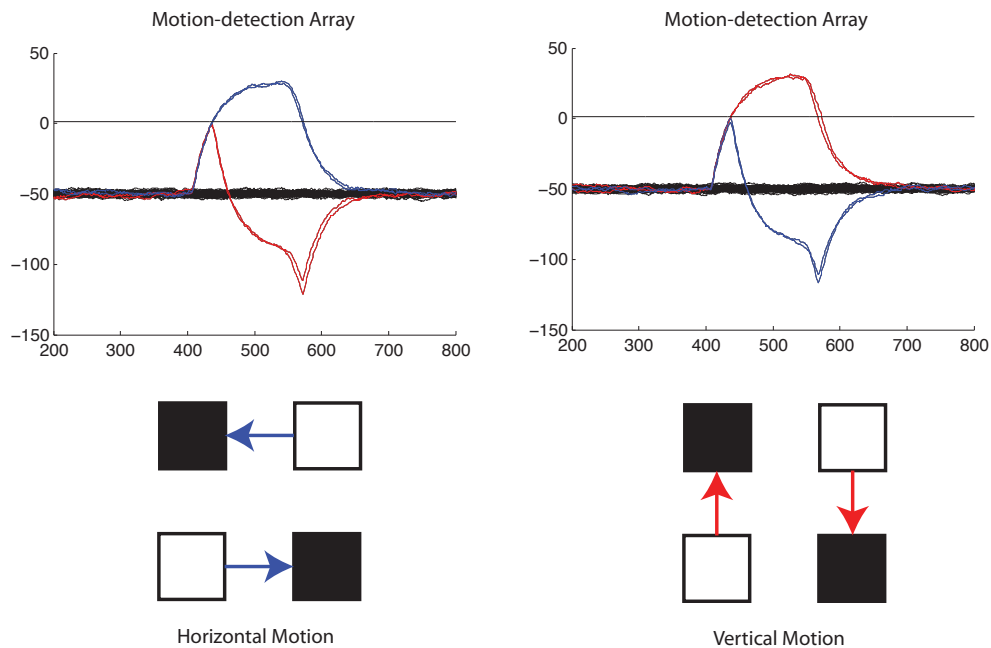


Figure 5.15: Bistability in two-frame motion quartet with no biasing array

However, two shortcomings should be noted: 1) changing the aspect-ratio of the display has no effect on the percept (the longer-path motions are just as likely to be perceived as shorter-path ones), and 2) when presented with additional display cycles,

each frame-change behaves essentially independently such that vertical and horizontal motion percepts do not tend to persist for multiple frame-cycles (i.e. perceptual switching happens 50% of the time, counter to human perception; Figure 5.16).

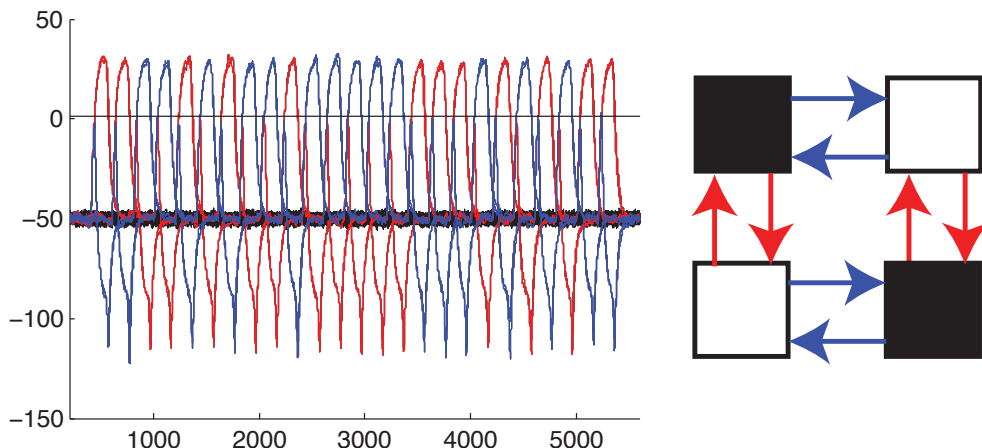


Figure 5.16: Bistability in motion quartet with no biasing array. Notice there is no long-term stability to vertical (red) or horizontal (blue) motion percepts; each frame-change is essentially independent

Biasing of percept via aspect ratio

Changing the aspect ratio by increasing either the vertical or horizontal distance between visual elements biases the likelihood of the network reporting either vertical or horizontal motion. Shorter-path motions have a resting-level advantage on the biasing array, and so in the case of the motion quartet it is one of the biasing elements corresponding to one of the shorter motion paths that tends to become the winner of the competition among biasing elements. The feedback to motion-detectors makes it more likely that the solution entailing the two shorter motion paths will be selected after competition on the motion-detection array.

Persistence and switching of percept for multi-frame display. Including the biasing array into the dynamical interactions leads to longer perceptual ‘runs’ of vertical and horizontal motion on the network (Figure 5.17; comprehensive results for all motion

quartet trials can be found in Appendix D). This occurs as a consequence of the distributed feedback projecting from the biasing array to the motion-detection array. When a percept is initially formed on the first frame-change, either the two vertical ('upward' and 'downward') or two horizontal ('leftward' and 'rightward') motion paths are perceived. Out of the two perceived paths, one of the corresponding biasing elements wins the internal competition on the biasing network which sends excitation to all of the motion-detectors with the same span.

For example, if vertical motion is perceived, and the 'upward' non-local biasing element wins the competition on the biasing array, all 'upward' motion-detectors of the same span receive excitatory input from the biasing element, regardless of their location relative to the stimulus array; motion-detectors whose corresponding biasing element do not win their competition receive no such feedback. Therefore, keeping with the example, on the following frame-change the 'upward' motion-detector on the other side of the motion quartet has a dynamical advantage over the other potential motion paths. In this context, this feedback is termed a *future-shaping interaction*, because what is presently perceived has an effect on future percepts that are yet to be formed.

The situation described above makes it likely that the pair of motion-detectors comprising the stable vertical percept will cross the interaction threshold before the pair of detectors comprising the horizontal percept, leading to cooperative inhibition of the 'leftward' and 'rightward' motion paths. On occasion, the noisy fluctuations of the motion detectors will cause the pair of detectors comprising the horizontal percept to reach the interaction threshold before *both* of the vertical motions do. Because of the nature of the cooperative inhibition, the percept that is ultimately formed is a result of the *pair* of motion detectors which first cross the interaction threshold, thus the advantage conferred to the single detector via biasing feedback is not sufficient to

stifle a perceptual switch.

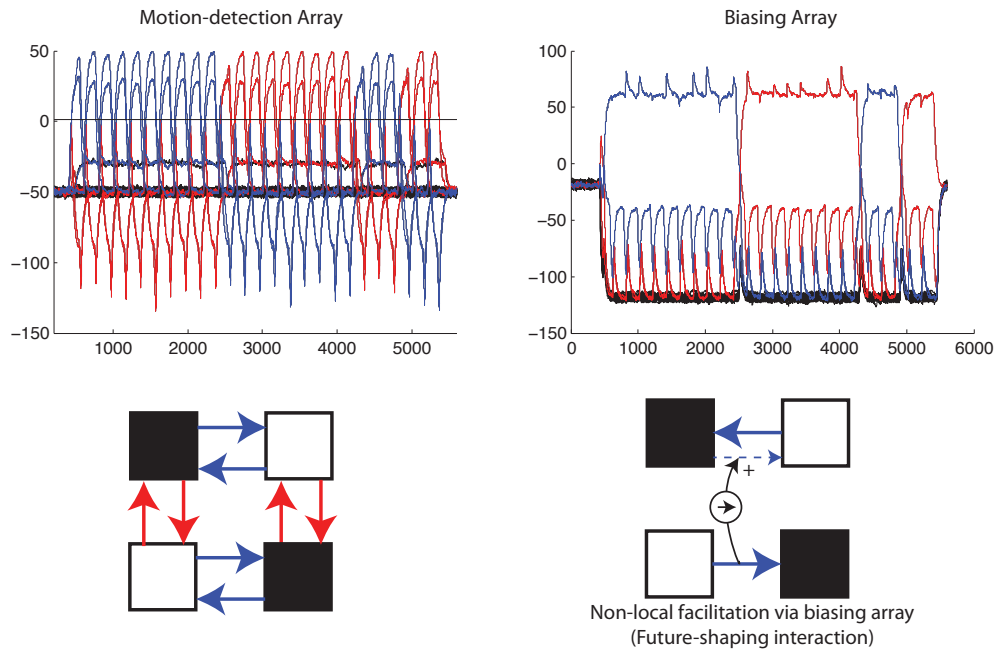


Figure 5.17: Motion quartet with future-shaping interactions and noise-induced perceptual switching

Without noisy fluctuations, the feedback to the single (class of) motion-detector is sufficient to maintain stability of the initially-formed percept indefinitely (Figure 5.18). This is because with all else being equal, the pair of motion-paths that contains a pre-activated motion-detector will always have a competitive advantage over the pair of paths with no such pre-activation.

The motion triplet. The motion triplet is like the motion quartet, except that one of the two frames (in this car, the first) contains only one of the two elements that are normally present in the quartet. Typically, splitting and fusing motion is perceived on alternating frames. Like the expanding and contracting stimuli, the unique split/fusion principle is not invoked in such a display, and therefore there is no effective inhibition between the motion signals. The magnitude of the motion paths in Figure 5.19 are the same, but this is not a necessary condition for the perception

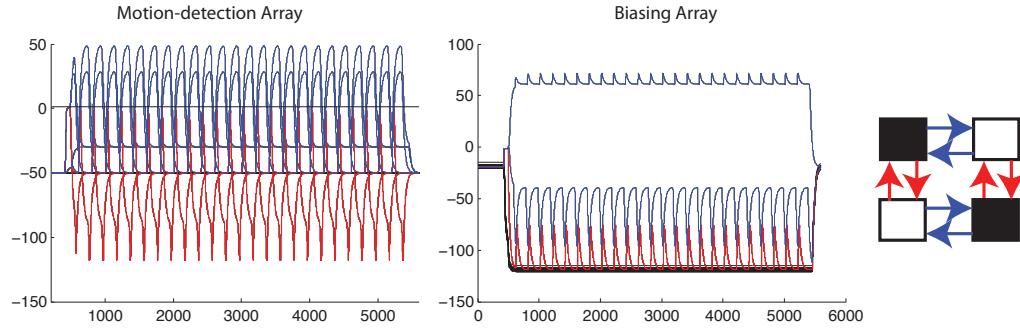


Figure 5.18: Motion quartet with future-shaping interactions without noise

of splitting and fusing in the motion triplet.

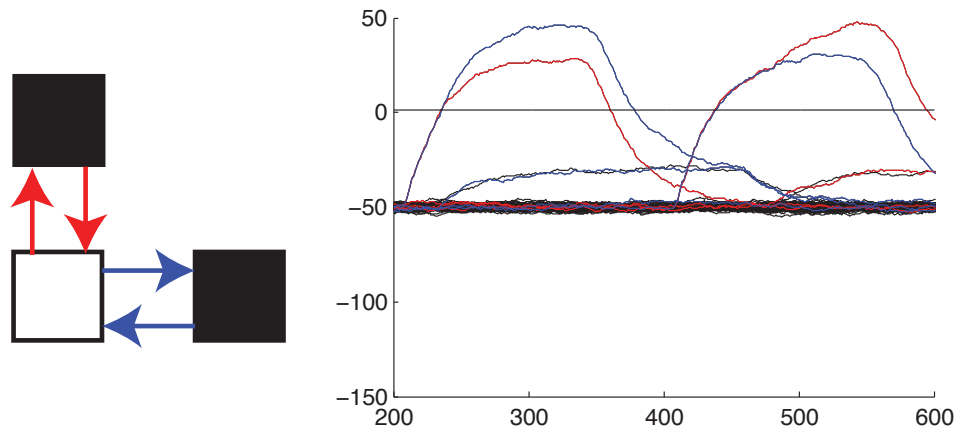


Figure 5.19: Motion triplet simulation

Coupling of two motion quartets. When more than one copy of a multi-stable stimulus is presented to the visual system simultaneously, all of the copies typically conform to the same perceptual interpretation. Static examples of this effect can be seen in e.g. the Necker Cube figure. This effect is also present in dynamic displays such as the motion quartet (Ramachandran & Anstis, 1983).

There are (at least) two problems to solve with respect to the coupling of multiple quartets. The first is obvious: how the multiple quartets ‘transmit’ their perceptual

state to one another in order to become coordinated into a common interpretation. The second is more subtle but equally important: how the quartets themselves are perceived as visual ‘items’ to be coupled. That is, why motion is not perceived to occur between quartets, but only within them. Models of coupling often take the latter for granted, and assume well-insulated stimuli whose states can then be coupled. Here this simplification is not made and, as will be seen, this can have an effect on the formation of the percept.

Presenting a stimulus with two quartets highlights some interesting aspects of the neural network. In the following examples, the distances between elements within a quartet are slightly smaller than the distances between elements between quartets.

In general, in agreement with human perception, the network reaches one of two stable perceptual states (Figure 5.20). Like the single motion quartet either vertical or horizontal motion is perceived, but in this case it is for both quartets. On some trials, the network finds either the vertical or horizontal solution on the first frame-change and then stabilizes around it, on other trials it takes several display-cycles of incoherent percepts to ‘search for’ and find either the vertical or horizontal solution which is then stabilized. When the percept takes several-cycles to stabilize, motion paths between quartets may be observed; this shows that the problem of ‘encapsulating’ each individual quartet is not given a priori, but must be a result of the emergent solution of the network. That perceptual organization may take multiple display-cycles before a stable percept is formed has been noted in the relevant literature (see, e.g., Hock et al. [2011]) and is in line with the author’s personal experience with ambiguous displays.

When the vertical or horizontal solution is found, there is a qualitative shift in the dynamics of the network. In all of the previous examples, the competition within the biasing array had produced one winner that delivered feedback to its correspond-

ing motion-detectors. However, in this case two winners emerge from the biasing array competition; namely, the two biasing elements corresponding to the spans of the two constituent vertical ('upward' and 'downward') or horizontal ('leftward' and 'rightward') motions. The shift from one to two winners on the biasing array can be understood if one considers the inputs the array is receiving. In the case of two motion quartets, on each frame-change there is the potential for *two* motion signals for each span (e.g. two 'upward' motions). This doubles the (potential) effective input to a motion-detector's corresponding biasing element which, in this case, is enough to overcome the competitive inhibition delivered by the other winner on the biasing array. The combined inhibition from both of these biasing-elements is enough to keep all the other biasing elements below threshold. Besides the four spans that correspond to the vertical and horizontal solutions (i.e. 'upward', 'downward', 'leftward', and 'rightward'), no other biasing element in the display can receive more than one motion-detector input during a given frame-change.

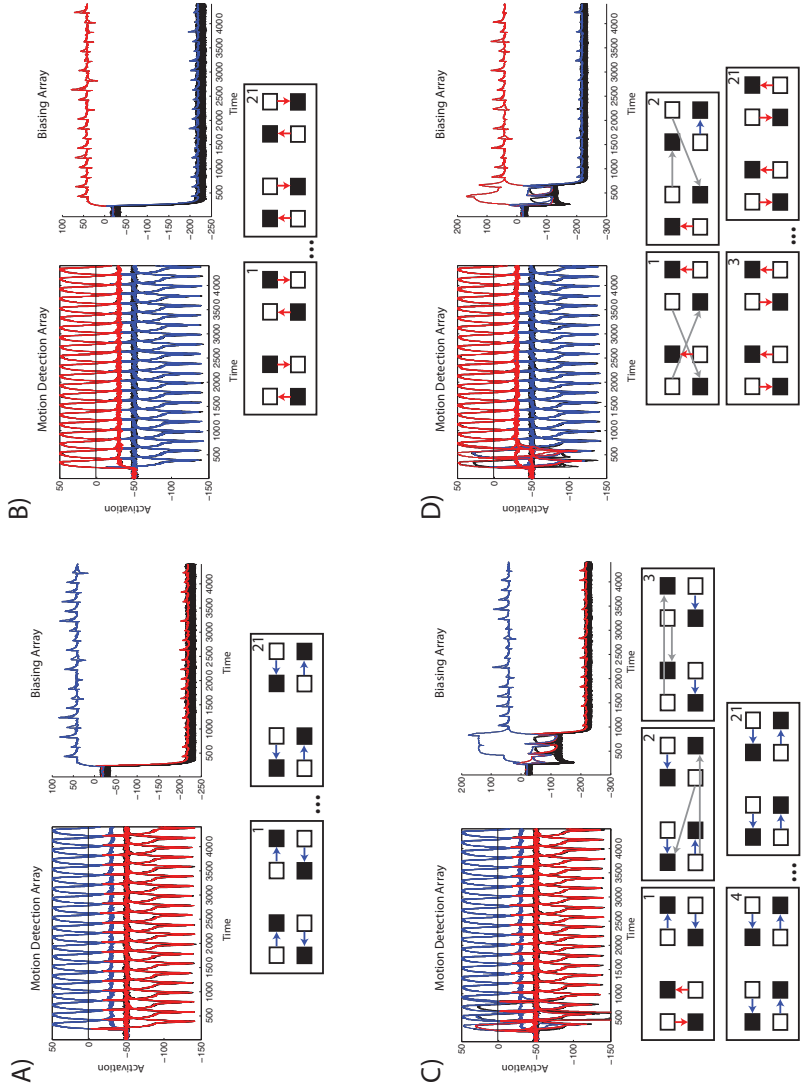


Figure 5.20: Simulations with two motion quartets. On some trials, the percept immediately organizes into the stable ‘vertical’ or ‘horizontal’ solution. On other trials, the network forms a short series of non-stabilizing incoherent percepts before arriving at a stable solution. Below the time series of the example trials, each rectangle represents the transient solution for a given frame-change (i.e. motion event) with the number in the upper standing for the frame change number, with 21 frame-changes total (i.e. 22 frames). Ellipses represent a qualitatively invariant solution between the frame-changes indicated.

When the horizontal or vertical solution is not (globally) reached on the first display-cycle, only one biasing element wins the competition on the biasing array; typically, a span corresponding to one of the constituent motions of the vertical or horizontal solution. Because there is not input from more than one detector to any other biasing element during this time, pre-activation is only delivered to one of the classes of motion-paths that contribute to either the horizontal or vertical percept. Eventually the local activations happen to favor the complementary class of motion-paths to the ones receiving pre-activation from the biasing array (e.g. for the vertical solution, ‘upward’ and ‘downward’ paths complement to form the global percept). When this occurs, there is enough input to the two relevant biasing elements for them to both ‘win’ and remain (transiently) above the interaction threshold. In this situation, the two complementary motion-detector classes (i.e. ‘upward’ and ‘downward’ or ‘leftward and ‘rightward’) are both receive pre-activation from the biasing array and therefor are given a strong advantage in the ensuing competition on the motion-detection array. In fact, the stability conferred to this motion pattern seems to prevent perceptual switching between vertical and horizontal percepts. It would therefore be very useful to look at the relative stability of vertical and horizontal percepts between stimuli containing single or multiple motion quartets to see if this prediction is accurate. As perceptual switches do occur in displays with multiple quartets, it would also be worthwhile to examine if, perhaps, adaptation is necessary for perceptual switches when more than one quartet is present. If not, it may be that the degree of noise in the current formulation is simply not sufficient to induce a switch.

Coupling of four motion quartets. More than two quartets can also become perceptually coupled in the same way (Figure 5.21). For four quartets, the dynamics are essentially the same. When either the vertical or horizontal solution is reached glob-

ally, two biasing elements ‘win’ the competition on the biasing array. Even though on a given frame-change there are potentially more activated motion-detectors converging on a single biasing element (four rather than two), the combined inhibition of the two winners on the other biasing elements is sufficient to keep them below threshold. When the global vertical or horizontal solution is not found on the first frame-change, multiple display-cycles might be needed to discover and stabilize those solutions.

Visual Inertia

Anstis and Ramachandran (1987) presented a variant of the motion quartet in which an apparent motion percept preceding the typical two-frame motion quartet stimulus strongly constrained the interpretation of the bistable motion quartet stimulus. In their display, two apparent motion paths that are collinear and spatially contiguous with either the horizontal or vertical motion paths typically induced by the motion quartet stimulus immediately precede the two-frame quartet display (Figure ??). Human observers tend to see the motion paths that can be perceived as continuation of the initially induced percept.

The network’s response is usually in line with human perception, although some trials show atypical responses. In a typical trial, the first two motions are selected by virtue of the feedback from the biasing array to one or the other of the shortest-path motions. Because of shorter-path motions’ resting-level advantage on the biasing array, one of the shortest-path motions tends to win. Additionally, because there are few concurrent inputs, the the biasing array has only a single winner, and pre-activation is delivered only to one of the relevant short-path motion detectors (and not the other moving in the opposite direction). On the next frame-change, the motion path that that is pre-activated by the biasing array (by virtue of being of the same span as the previous motion path which induced the biasing array winner), tends to

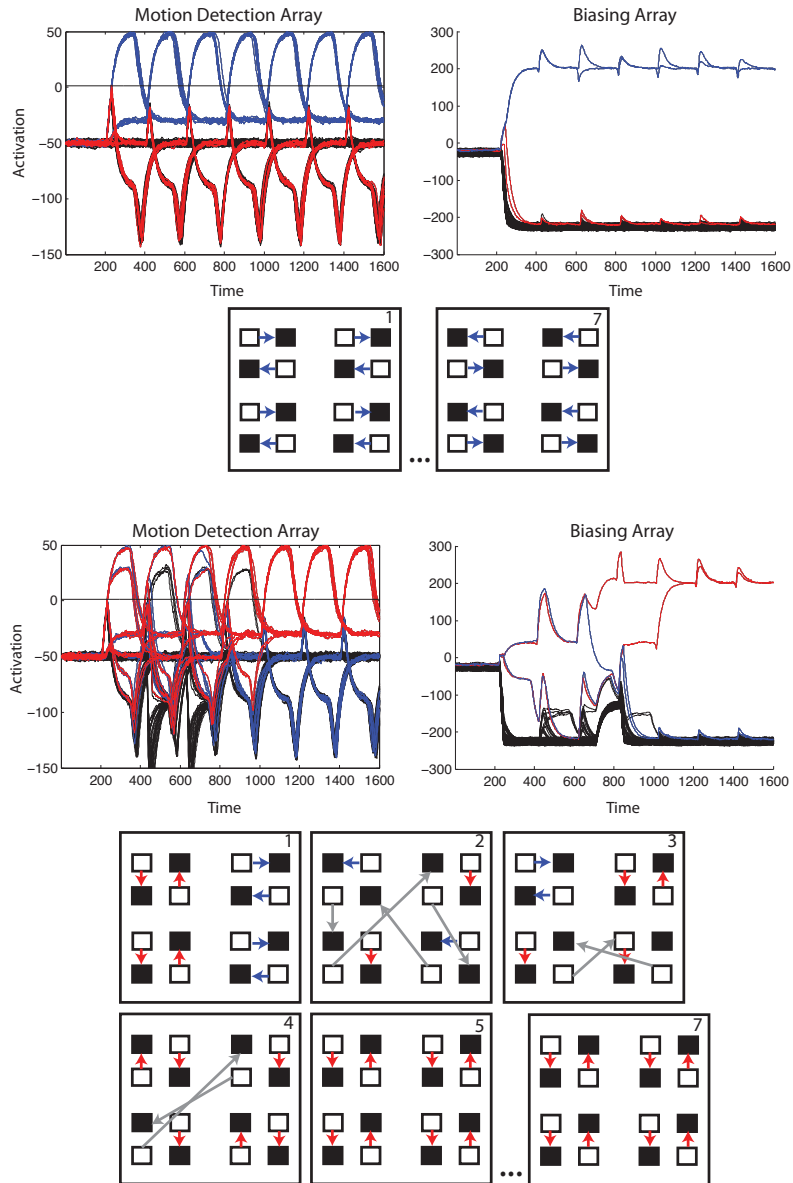


Figure 5.21: Simulations with four motion quartets. On some trials, the percept immediately organizes into the stable ‘vertical’ or ‘horizontal’ solution. On other trials, the network forms a short series of non-stabilizing incoherent percepts before arriving at a stable solution. Below the time series of the example trials, each rectangle represents the transient solution for a given frame-change (i.e. motion event) with the number in the upper standing for the frame change number, with 7 frame-changes total (i.e. 8 frames). Ellipsis represent a qualitatively invariant solution between the frame-changes indicated.

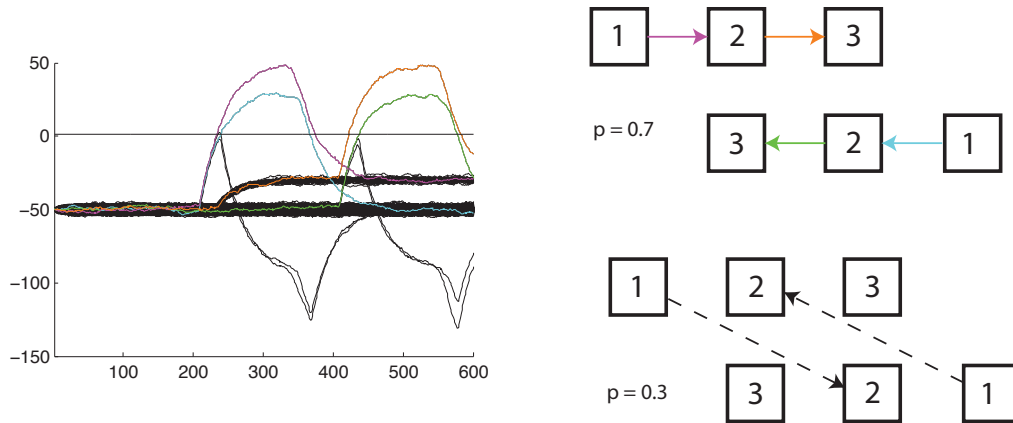


Figure 5.22: Visual Inertia. The number inside each element (box) indicates the frame it is visible on. When the network is primed with the two short-path motions on the first-frame change it constrains the percept of the next frame-change and biases it toward motion paths that ‘continue’ the previous paths; this solute was obtained a majority of the time ($p = 0.7$). On a minority of the trials ($p = 0.3$) the initially perceived motions were the two longer-path motions, and thus no priming effect is seen.

constrain the solution to the motion paths that continue in the same direction as the previous ones.

On atypical trials, the first two motion paths are not the shortest-path ones, but their longer-path alternatives. There are two parameters that might be implicated in what is essentially a wrong answer. The first is the parameter that sets the degree of resting-level falloff on the biasing array. It might be that the longer-path motions, by virtue of a noisy-fluctuation, could win the competition on the biasing array, feeding-back to a longer-path motion. In practice it is difficult to dissociate the causal influence of a winner on the biasing subnetwork on local motion detectors, and motion-signals winning by cooperative competition on the motion-detection network and causing one of the corresponding elements on the biasing array to win, as they will be in agreement when the system (transiently) stabilizes, and before that they are both influencing one another in a reciprocal manner.

Another possibility is that increasing the feedforward synaptic weight from the motion-detection array to the biasing array could allow the biasing subnetwork to function in the two-winner regime displayed when more motion signals are present. This may be important for preventing the typically not-perceived longer-path motions from ultimately being selected for via a noise-induced cooperative (inhibitory) advantage.

Although the atypical trials did not correspond to typical percepts, it is anecdotally reported that with sufficient attentional effort, the atypical perceptual solutions can be stabilized. A rigorous psychophysical study would need to be performed to confirm this report.

Finally, the way the network solves the problem on typical trials suggests a testable prediction. The motion paths on the second frame-change that are seen to be continuous with the motion paths on the first are pre-activated by virtue of a non-local same-span biasing element. This implies that it is not necessary for the motion paths that induce the future-shaping interactions that constrain the solution to the second frame-change to be spatially contiguous with them, they simply have to be of the same span.

Group Motion

When multiple visual elements move in an invariant (i.e. rigid) formation, motion paths are often perceived that do not correspond to the shortest motion paths. This display presents a challenging problem to correspondence solvers because, in addition to the typically perceived group motion, one of the possible (incorrect) solutions also contains group motion. In the display, three vertically-stacked but spatially separated visual elements are visible on the first frame. On the second frame the stack has shifted to the right and down such that the topmost element on the second frame is

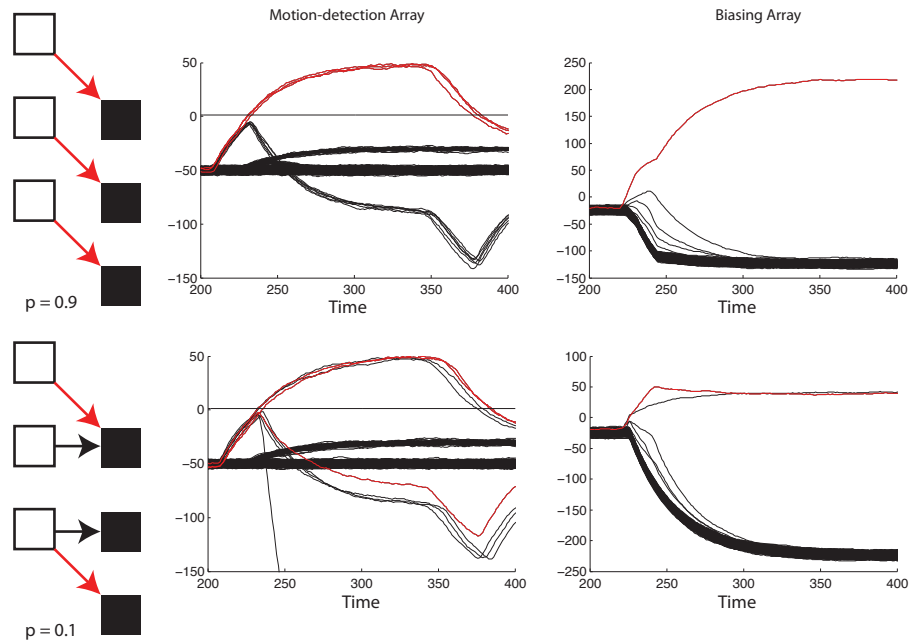


Figure 5.23: Group Motion

in a position directly to the right of where the middle element was on the first frame (Figure 5.23). This implies that two of the elements' shortest path solutions agree with one another (group motion).

The typical percept is reported as the three visual elements moving rigidly down and to the right. That is, three parallel motion paths of the same magnitude, and with their locations of origin and termination vertically aligned. One can see that there is no other potential set of counterchange motions with three (or more) same-span motion paths. The three corresponding motion-detectors converge on a single biasing unit, which typically wins the competition on the biasing array. The feedback to the three motion detectors that correspond to the group motion percept gain a competitive advantage and become the preferred solution.

On one out of the ten total simulations performed on this stimulus, an alternate global solution was obtained. Interestingly, it is the same solution given by Ullman's

minimal mapping theory to this stimulus. In this solution, the two shortest-path motions are elected for, and a fusion as well as a split account for the remainder of the correspondences. This solution is also found consistently in the current formulation if the biasing threshold and the cooperative inhibition threshold are made too similar, not giving enough time for the soft constraints to bias the network before the hard split/fusion constraint is applied.

Sliding and splitting

The final example case consists of a two-frame display in which there are three visible elements on the first frame and four on the second frame (Figure 5.24). The three elements on the first frame are arranged in a vertical stack as in the previous group-motion case. The four elements on the second frame are also vertically stacked, with their vertical locations occupying positions just above and below and to the right of the locations of each element on the first frame. Because the number of elements on frame two is greater than the number of elements on frame one, a splitting motion is likely entailed in the network's solution.

This display presented a challenge for Dawson's (1991) model, in which he had to change the parameters from the ones used for most of his simulations in order to obtain a reasonable response from the network. Interestingly, the solution his network reached using his 'typical' parameter values is not consistent with the unique split/fusion principle as it entails motion paths which take part in both a split and fusion.

Here, three alternative solutions are found by the network in response to the display; one of which matched Dawson's 'correct' solution (which required modification of his parameters), while the other two are novel. Unfortunately, no systematic psychophysical study of this display has been performed to the best of this author's

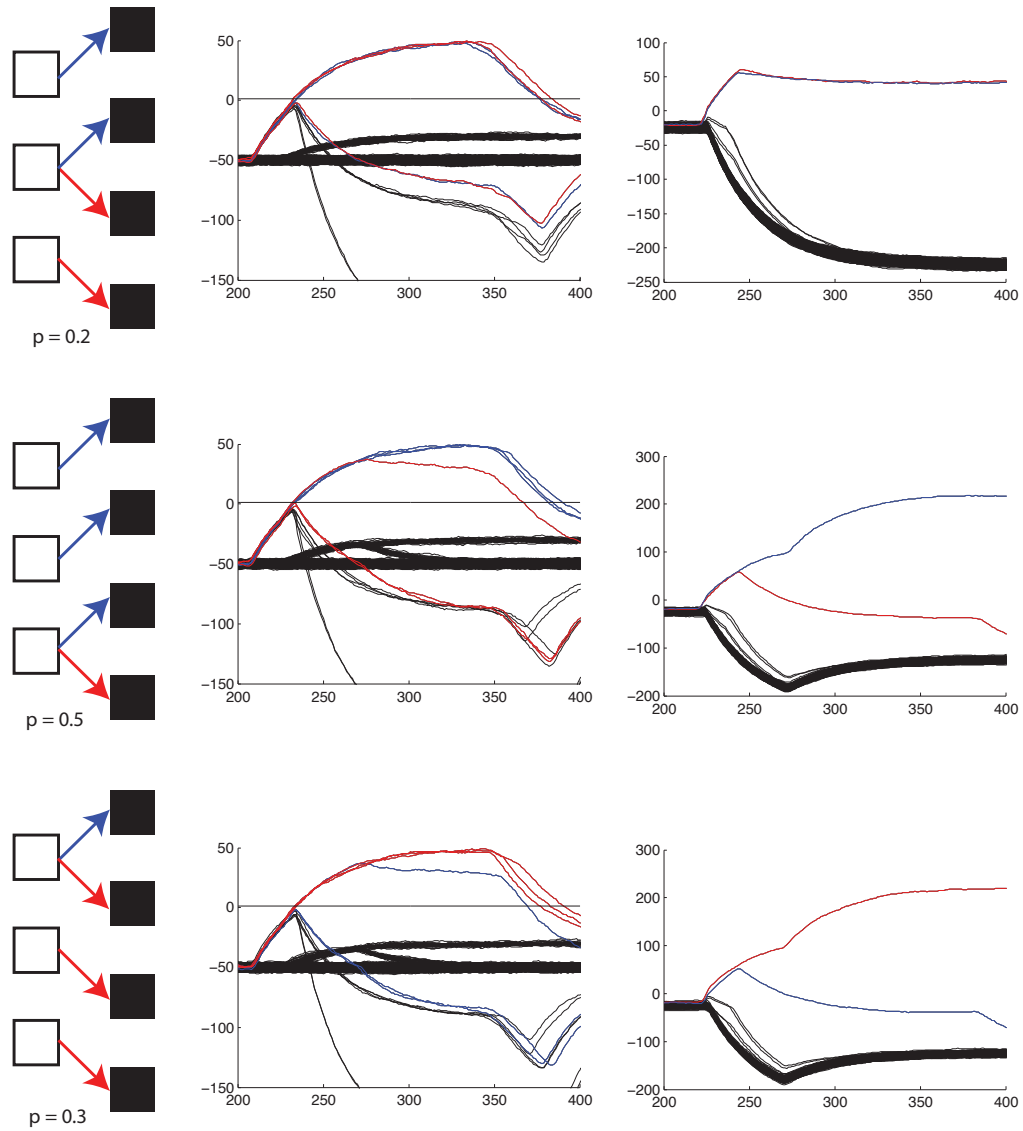


Figure 5.24: Sliding and splitting

knowledge. However, anecdotally, all three solutions of the network are easily perceived (on different trials) by human observers.

The solution that corresponds to Dawson’s ‘correct’ solution entails the topmost element being perceived as moving up and to the right, the bottommost as moving down and to the right, and the middle element splitting and moving to the right and both upward and downward. This solution contains two pairs of same-span motions,

and as such the biasing array reaches a two-winner solution corresponding to those two spans.

The other two solutions are essentially mirror images of one another. In these trials, group motion of three elements is seen (to the right and either up or down), with either the topmost or bottommost element being seen to split, accounting for the additional element on the second frame. The biasing element that corresponds to the span of the three element group motion is the sole winner on the biasing array for these trials.

5.4 GENERAL DISCUSSION

The model presented above is a generalized architecture with homogenous patterns of connectivity out of which solutions to difficult correspondence problems are obtained. The example cases are not exhaustive of the ‘correct’ solutions the model may obtain, but are meant to allow observation of the features of the model under various circumstances. The solutions emerge out of the dynamics on the network, and specific connections between various detectors do not have to be established individually for each case as in previous models.

The stimulus input is given directly to the model with no intervening stages of operator intervention (i.e. motion signals are not pre-labeled or pre-identified, but are detected directly from the dynamic stimulus array). The parameters of the model were held constant for all simulations; specialized parameter values were not used to solve specific problems. It is likely that a more suitable set of parameter values could be implemented and future work should seek to probe the parameter space of the model systematically.

5.4.1 The necessity of cooperative inhibition

An alternative model could be formulated in which competitive interaction between motion detectors that share subunits are additive rather than multiplicative. In other words rather than the cooperative-inhibition entailed by the unique split-fusion principle, dependent motion signals could inhibit one another directly. However, this scheme is insufficient to account for the expanding and contracting motions that are obtained in the simulations above. Here it is shown why this is the case.

In order to develop a standard by which to calibrate a model entailing direct inhibition between dependent motion-detectors, the model is constrained to account of two known perceptual regularities: 1) the bistability of the two-frame motion quartet, and 2) the perception of splitting-motion in the motion triplet. Logically, for these two percepts to be accounted for, 1) the degree of combined inhibition from two motion detectors convergent on a third must be sufficient to drive it below perceptual threshold, and 2) the degree of inhibition from one motion detector to another must be insufficient to drive it below perceptual threshold. This is formalized in the following.

Consider a set of four motion detectors that correspond to the four potential motion paths perceived in a two-frame motion quartet display. For simplicity, assume that neurons receive enough excitatory input to be driven to saturation. Let h be the resting level, α be the perceptual threshold, $\alpha^* = \alpha - h$ be the effective perceptual threshold, w_{exc} be the synaptic strength of input to the motion detectors, w_{inh} be the inhibitory synaptic strength between each pair of motion detectors that share a subunit, and $\beta = w_{exc} - \alpha^*$ be the *excess activation*, which is the amount that a motion detector is driven above the perceptual threshold when it is excited by fully-saturated subunit inputs and not inhibited by any other motion detectors (Figure 5.25). The amount of inhibition required to drive a given motion detector which is receiving

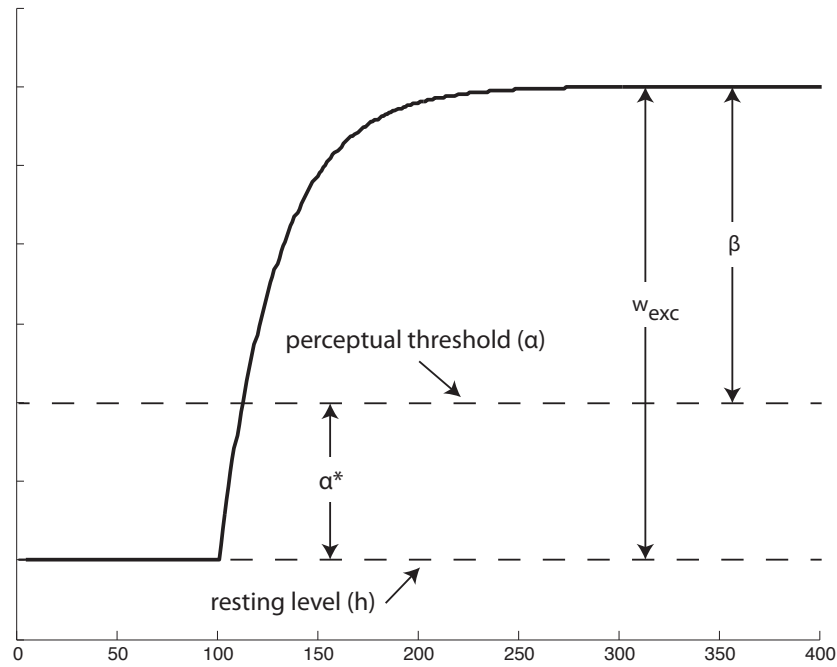


Figure 5.25: Diagram of symbols denoting various degrees of activation for a perceptual neuron

maximum feedforward excitation from its component subunits below the perceptual threshold is equal to β . Therefore, for the model to meet the standard of displaying both bistability of horizontal and vertical percepts in response to the motion quartet and splitting motion in response to the motion triplet, the constraint $-\beta < w_{inh} < -0.5\beta$ must be met. If $w_{inh} > -0.5\beta$, there will not be sufficient inhibition to drive the unperceived motion signals below perceptual threshold in response to the motion quartet, and if $w_{inh} < -\beta$, there will be too much inhibition in response to the motion triplet, preventing splitting motion from being perceived.

If we choose a value for w_{inh} that satisfies these constraints, both the bistable motion quartet and the splitting motion triplet are perceived by the model. However, when additional motion paths are entailed from or to a common location, as in the

expanding and contracting stimuli, the degree of inhibition is too great to allow all motion detectors to reach perceptual threshold and produce the expanding motion percept.

Consider the case of two-frame expanding motion when one central element is visible on frame one and four elements on frame two (to the left, right, above, and below the original central element). In this case, when at least two detectors are driven to saturation, the effective inhibition on the others equals $2w_{inh}$. Because w_{inh} is constrained to be less than half the excess activation ($< -0.5\beta$), a motion detector receiving inhibitory input from two other detectors will be driven below perceptual threshold ($2w_{inh} < -\beta$). This implies that the detectors receiving this inhibition will be pushed below perceptual threshold, and thus expanding motion will not be perceived.

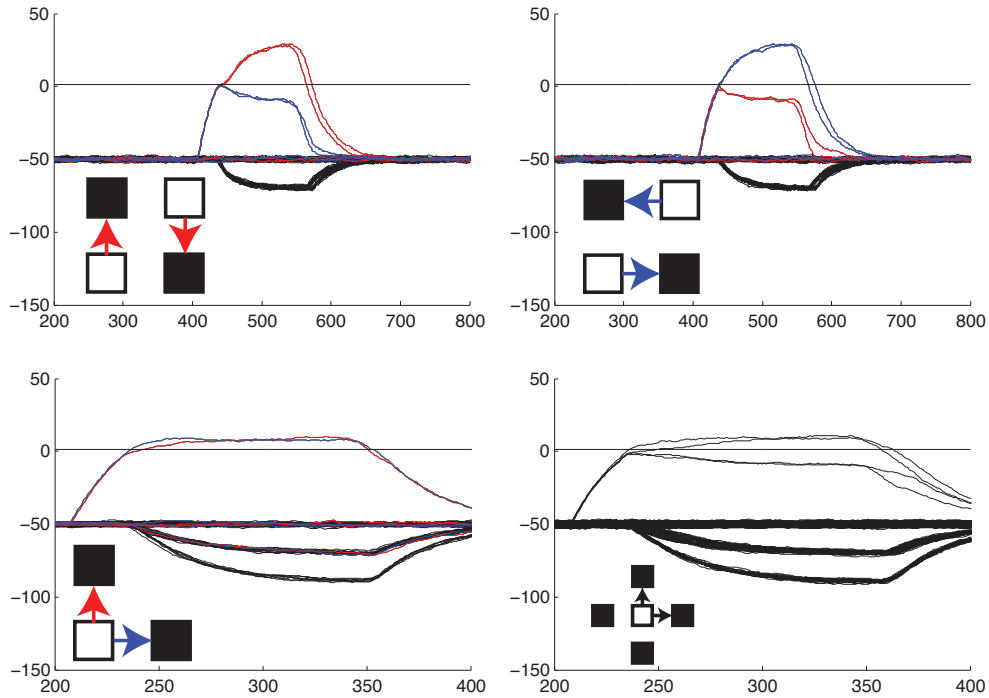


Figure 5.26: Simulations for network with additive inhibition

This has been confirmed in simulations in which the correspondence network was implemented with additive rather than multiplicative inhibitory synapses between motion detectors. When the degree of inhibition is set appropriately to achieve bistability for the motion quartet and splitting for the motion triplet, expanding motion is not found as a solution (Figure 5.26). Additionally, this analysis also pertains to models in which only perpendicularly oriented pairs of motion signals that share a subunit inhibited one another (i.e. cross-direction inhibition) as has been employed in previous models.

5.4.2 The role of the biasing array

The reciprocal interaction between the motion-detection and biasing subnetworks displays an interesting dynamic. In the signal generation phase of a typical trial, (potential) motion signals that are above the bias threshold but below the internal interaction threshold cause activation on the biasing array, which feeds back additional activation to a subset of the motion signals that are being generated. When the motion signals cross the threshold for competitive selection, a subset may be inhibited. When inhibited, the activity of these neural elements is driven below both the selection and biasing thresholds (by virtue of the selected parameters); thus, the activity on the biasing array is being driven solely by the subset of motion signals that are *perceived*. In other words, the activity on the biasing array is driven by both perceived and unperceived motion signals, where the degree of influence of both sources of input varies as a function of the state of the whole network. In other words, at some moments the activity on the biasing array is being driven by mostly unperceived motion signals, at other times mostly by perceived motion signals, and at other times it is a mix of the two.

That the biasing array is mainly driven by perceived motion signals after com-

petitive selection is conceptually critical. For apparent motion percepts that display long-term stability to stimuli with more than two-frames, it is the initially established percept that evidently plays a dominant role on following perceptual judgments, as in the sustained vertical or horizontal motion percepts in the motion quartet. If the biasing array are driven equally by both perceived and unperceived motion signals after competitive selection, there would be no advantage for the initially perceived motion in constraining ensuing percepts.

However, of critical importance is also that the biasing array is effected by unperceived motion signals prior to competitive selection. This is necessary for the biasing array to integrate evidence for *potential* motion signals globally such that collective effects as seen in, for instance, group motion can effectively bias the appropriate subset of motion signals.

The biasing array also has an internal competitive dynamic. The role of this competition is to effectively bias only some motion signals, and not others. If there were no internal competition on the biasing array, all potential motion signals would be pre-activated by essentially the same amount, eliminating the functionality of the biasing array.

The stabilizing role of the biasing array

The stabilizing role of the biasing array is made most clear in the case of multiple motion quartets, especially when a stable solution is not immediately found. When a collection of motion detectors representing paths with the same direction and magnitude are activated simultaneously their convergent input on a common biasing element makes it likely that those detectors will receive excitatory feedback, giving them a competitive advantage. The ‘vertical’ and ‘horizontal’ solutions to displays with multiple motion quartets represent the solutions with the greatest number of

such convergent signals. In other words, they represent the solutions with the most global *coherence*. (Recall the importance of coherent motion patterns for the short-range motion paradigm in Chapter 3.)

Importantly, there is no pre-specified templates or pattern detectors specifically for (for example) the ‘vertical’ or ‘horizontal’ perception of coupled motion quartets. The stable percepts entailed in such solutions is an emergent collective dynamic of the total network. In the cases when one of these two stable solutions is not reached on the first frame-change, an interesting succession of *incoherent* motion patterns may follow. The network will then ‘search’ for a stable solution over the next (few) frame-changes, eventually arriving at either the ‘vertical’ or ‘horizontal’ solution and remaining there for the remainder of the trial. Prior to this, motion is often seen between elements of what are considered different quartets. This highlights the fact that a motion quartet is not an encapsulated entity a priori, but must be actively identified to be treated as an ‘item’ by the visual system.

It can be seen in the comprehensive motion quartet results in Appendix D that when a solution is initially incoherent, coherence of motion signals develops over the following frame-changes. Motion paths that correspond to the most coherent solutions (again, ‘vertical’ and ‘horizontal’) are ‘recruited’ via excitatory feedback, which in turn increases the likelihood of their corresponding biasing element to continue to win in the internal biasing competition and continue to provide an advantage to its motion-detectors. When a stable solution is reached, the maximum number of coherent motion signals are perceived for the display.

Other sources of biasing

While excitatory biasing of the motion-detection network was only specified through the so-called biasing array in the current model, there is no conceptual limitation

that prevents other sources of competitive advantage of motion signals through pre-activation.

An intriguing connection is the potential interplay between stationary and dynamic features in the formation of motion percepts. For example, a simple correspondence display that is composed of two vertically stacked elements on frame 1 and the same stack shifted to the right on frame 2. Coherent, group motion is almost always perceived in such a display. However, if two lines connect the elements that correspond to the typically unperceived (potential) motion paths, the two crossing motions may be perceived rather than the coherent group motion.

It is conceivable that these stationary features bias the motion paths associated with their endpoints, giving them a competitive advantage. The interaction between stationary and dynamic features of a stimulus during visual perception is an extremely important and challenging problem. This suggests one potential functional interplay between the two: the biasing of motion signals based on static cues. Importantly, the ‘crossing’ solution found when the stationary lines are present does not violate the hard constraint of the unique split-fusion principle.

5.4.3 The Dynamic Application of Constraints

The model makes use of two thresholds which are essentially outputs from the motion detection network. The lower threshold (the bias threshold) determines at what activation values motion-detection elements exert causal influence on the biasing array. The higher threshold (the selection threshold), determines the activation at which motion detectors begin to participate in competitive selection. This modeling choice was driven by the need to apply the soft-constraints prior to the hard ones. This is a logical necessity, as by definition soft-constraints are flexible while hard-constraints are non-negotiable. Therefore, if the hard-constraints are applied first,

the soft-constraints have no impact on the eventual solution.

It should be possible to look into neural tissue to look for analogs of this organization. For instance, one could check whether all axons projecting from a given neuron transmit action potentials at the same threshold potential, and if so if it is a function of where the axon projects to. Or, perhaps if not at the level of the single neuron, layers of cortex could have distinctly different thresholds while receiving similar inputs, with one layer's threshold being systematically lower than the others. Again, if this was the case, it would be instructive if the locations of projection were also distinct (e.g. short- vs. long-range).

Additionally, There may be other means of achieving the appropriate application of constraints without necessitating an entity with multiple output thresholds. It has been noted by Hock et al. (2011) that in cortex, axons that make long-range connections between brain areas are typically myelinated, while axons within brain areas are generally non-myelinated. Because myelinated axons transmit action potentials at a much greater speed, transmission between brain areas may in some cases have less latency than transmission within a brain area (especially when the myelinated connections are relatively close, e.g. within the visual cortex). Given this 'vertical' over 'horizontal' speed advantage, one could imagine the interaction between the motion-detection array and the biasing array to occur faster than the lateral competitive interactions on the motion-detection array, achieving the appropriate order of constraint application. Transmission times were not explicitly modeled in the present work (a simplifying assumption), but the sufficiency of this concept could be tested for in future models in place of multiple thresholds.

5.4.4 Relation to neural field models

Neural Fields

The connectivity of the sub-neighborhoods defined in the motion-detection array as well as the connectivity among the neural elements on the biasing array are essentially discrete analogs of connectivity kernels in neural field models (e.g. Amari, 1977; Schöner, 2008). In a neural field, each location integrates neural activity over the continuous neural space as input according to the connectivity kernel which is commonly a mexican-hat-like function entailing local excitation and long-range inhibition.¹⁰

Aside from the discreteness, the lateral connectivity on the motion-detection network differs in two (related) ways from typical field models' connectivity kernels. First, in a field model the connectivity kernel is usually of the same dimension as the neural space it is embedded in. For example, a two-dimensional field typically entails a two-dimensional connectivity kernel. On the motion-detection network, lateral interactions occur in what are essentially subspaces of the neural space. For instance, on the two-dimensional motion detection lattice (for one-dimensional correspondence problems), lateral interaction takes places in rows and columns, each of which can be thought of as one-dimensional subspaces embedded in two-dimensional neural space.

Second, embedding multiple such low-dimensional connectivity patterns in the same space places an important role on their intersection. Having multiple low-dimensional connectivity patterns means that a given neural element can exist, in a sense, in multiple neural spaces simultaneously. In the case of the motion-detection network, one such subspace is defined by its vertical, columnar membership and the other by its horizontal, row-like membership. These subspaces influence each other

¹⁰The current model has no analog for the local self-excitation inherent in many field models. In the appropriate parameter ranges such excitation can lead to self-stabilized peaks of activity in the field, and/or detection instabilities in response to input. Future work should investigate the consequences of introducing self-excitation to the current model.

directly only through their shared elements (at their intersection). This conceptual scheme does not demand the multiplicative interaction employed above, but the intersections of these *synaptic subspaces* are certainly brought to the fore when their effective combination is multiplicative.

Giese's neural field motion detection model

Giese (1998) has developed a neural field model that shares some characteristics with the present work. In his four-dimensional field model, each location represented a motion vector's two-dimensional (retinal) location, radial direction, and magnitude (in contrast to representing a counterchange motion's two-dimensional location of origination and termination). Interactions between motion detectors were mediated by the field's connectivity kernel, causing facilitation between similar motion vectors and competition among dissimilar ones; no mediating sub-networks for biasing were used as in the model developed here. Giese (1998) also derived his elementary motion vectors from a field of Reichardt detectors, all with the same optimal displacement (i.e. magnitude of the span). Here the motion-detection scheme entails detectors that can detect motion optimally and displacements even octaves apart.

Berger's counterchange neural field model

A counterchange-based neural field model for motion detection has also been developed by Berger et al. (2012). This model tested the sufficiency of the counterchange concept in the context of continuous motion. However, it was limited in the same way as the Giese (1998) model by only having one displacement over which all motion detectors operate. This presents a challenge in accounting for apparent motion where displacements sufficient to induce a motion percept may vary substantially. Additionally, no interaction among motion detectors is entailed, and thus it cannot

account for correspondence problems.

5.4.5 Natural Constraints

Approaching problems of visual perception in terms of constraints raises a question: where do the constraints come from? Constraints are related to regularities in the environment that can be leveraged to reduce uncertainty. When things move they often *move together as one*, implying that group motion might be more viable than an alternative percept entailing shorter-paths. Presumably, the correspondence between the constraints imposed by the nervous system and the ones inherent in the Newtonian ecological scale we inhabit have been gained through evolutionary, developmental, and learning processes. They are the constraints that have worked, and that's why they are our constraints. Constraints are as much about what what is not as what is; but in constraining functionality is gained.

Expansion and contraction are ecologically meaningful patterns. In addition to providing information about direction and heading in optic flow patterns, they can signal meaningful transformations in local objects e.g. 'looming' when an object approaches (withdraws) and its retinal projection expands (contracts). In most contexts, expansion and contraction are mutually exclusive; they signify opposite directions of both locomotion and object motion. This is perhaps the ecological 'logic' of the unique split/fusion constraint. It is not likely to encounter a meaningful pattern in an ecological context that combines the both expansion and contraction as sub-patterns into a single object or event, while both expansion and contraction by themselves may carry valuable information.

5.4.6 Limitations

Spatial filtering

Group motion has been defined conceptually here as elements that are perceived to have parallel counterchange motion paths of the same magnitude. For many problems this is sensible, and perhaps necessary. However, the well-known Ternus display presents a challenge. The Ternus display is composed of two frames each with three visible elements aligned in a row. Two elements are stationary across the two frames, while an element is visible at one end of the row in the first frame and the other end in the second frame. For some parameter ranges, motion of a single element tends to be seen from the location where the element disappears on frame 1 to where the new element appears on frame 2 while the two stationary elements are not seen to move. This single-element motion case is accounted for in a straightforward manner by the counterchange network here. However, under other conditions, observers tend to report seeing the whole group of three elements moving together; thus the ‘stationary’ elements are seen to shift over and occupy their neighbor’s former location. In this case, there is no counterchange motion at the scale of the individual elements as the stationary elements can not cause a decrease nor increase in their respective detectors. However, a larger spatial filter could detect counterchange motion of the group *as a whole*.

Only one scale of spatial filtering is used in the model above. This is a simplifying-assumption, as it is well known that multiple parallel channels tuned to different spatial frequencies are evident in the visual system of humans and other animals. The question then is, would incorporating multi-scale spatial sampling into the model be able to account for all cases of group motion?

The dense random-dot cinematogram studied in Chapter 3 would seem to suggest

not. As the moving figure in those displays is composed over eighteen-hundred individual elements, it is difficult to imagine that a spatial filter at the scale of the figure would be sufficient to detect the motion while being embedded in noise. In other words, the spatial filters that correspond to the scale of the moving figure would give very little response by virtue of the fact that the visual (random) elements are so well-mixed at that scale. Therefore, it is likely that the visual system does indeed leverage multiple parallel motion paths as evidence for collective motion.

While it may not be sufficient to account for all cases of group motion, multi-scale sampling will likely be necessary for a full account of human visual perception. Future work should look to dissociate ‘true’ collective effects of elementary motion pathways from low-level grouping as when multiple distinct visual elements excite a single spatial filter.

Higher-level patterns

The highest level of the perceptual hierarchy in the model is the biasing array, in which each neural element represents a given class of motion-path, invariant with respect to the location of the path. This represents a simple perceptual pattern that could be termed *global coherence*. Regardless of their locations, motion detectors representing paths of the same direction and magnitude facilitate one another and tend to form patterns of group motion. While this is sufficient in accounting for the typical coupling of multiple motion quartets, for example, it is unable to account for more nuanced motion patterns composed of multiple elementary motion signals. For example, Hock et al. (2011) have developed a display referred to as the diamond quartet. The diamond quartet is composed of four motion quartets arranged in a diamond pattern whose motion can be perceived as either the globally coherent ‘vertical’ or ‘horizontal’ solutions found here or as a global ‘rocking’ rotation. The rocking percept entails the

‘horizontal’ solution for the top and bottom quartet and a ‘vertical’ solution for the left and right quartet. Thus, this higher-level pattern is capable of stabilizing a solution that implies a more specific and nuanced structure among a collection of elementary motion detectors than simply their coherence with respect to direction and magnitude. The set and source of these higher-level patterns remains an open question, and future work should address the generation and maintenance of such patterns (of patterns).

Spatial discreteness

Additional limitations stem from the discreteness of the model. For example, only motions with precisely the same span facilitate one another via a common biasing element. If two motion-detectors have similar-but-different spans, they do not facilitate one another (and their collective biasing elements actively compete). In reality, there is likely a range over which motion detectors that are ‘similar enough’ facilitate one another. A field formulation of the model would allow for specification of such a nuance.

Resource limitations

Resource and time limitations also made it infeasible to study random-dot displays in sufficient detail. Future work should evaluate the model’s response in the context of dense random-dot displays like those used in Chapter 3 as well as sparser variants as commonly used in, for instance, structure-from-motion displays.

5.4.7 Viability as a real-time computer vision system

Because of its front-end specification and formulation as a continuous-time dynamical system able to cope with a stream of dynamic input, the model presents the possibility

of being developed into a real-time computer vision system. However, some additional development would be necessary for this to be realized.

Testing of the model was limited by computational resources. The implementation of the model is inherently processing intensive. At each time-step every neuron must integrate all of its inputs and calculate its future state; including neurons that are essentially inactive (i.e. below the interaction threshold). A more lightweight alternative would be to adopt an event-based approach as advocated in, e.g., the neuromorphic approach (e.g., Benosman et al., 2014). In an event-based approach, explicit calculations take place only when there is interaction among elements. That is, a computation is triggered by an *event*. In the nervous system, action potentials serve as events, and downstream computations only need to be realized when these events trigger them.

In addition to minimizing its inherent processing demands through implementing an event-based algorithm, the processing that will still be necessary can be enhanced through optimizing hardware-software interface. For example, there has recently been a surge in using computers' graphical processors (GPU) to accomplish computationally expensive processes more quickly by parallelizing distributed computations (see Sanders & Kandrot, 2010).

5.4.8 Hierarchical pattern formation

Since the Gestalt movement of the early twentieth century, there has been considerable interest in understanding the relationship between parts and wholes in perception. For instance, is a face merely a collection of face parts (e.g. eyes, nose, mouth, etc.), or is it something over and above such a collection? Can what is otherwise perceived simply as a circle be perceived as an eye within the context of a face? More generally, do parts determine wholes or do wholes determine parts (or is it both or

neither)?

Here perceptual patterns occur at several levels of description. Local changes in spatial filter activation may combine into counterchange patterns. Counterchange patterns carry perceptual meaning that is not partially present in either a decrease or increase in local spatial filter activation, but emerges out of their combination. One could say that a pair of increase and decrease detections *entail* a counterchange motion, while a counterchange motion *accounts for* the pair of oppositely-signed change-detections.

Perception is arguably a process of sense-making. That is, a functional perceptual system transforms dynamic sensory events that underdetermine the state of the environment into meaningful signals that provide the perceiving organism with an understanding of ongoing environmental events and opportunities. In this sense, the role of ‘high-level’ perceptual modules would be to provide concise signals that are able to account for a large number of ‘low-level’ neural events. In other words, a small number of high-level patterns can potentially account for larger number of low-level patterns. In the case of counterchange motion, a motion signal is able to account for exactly two events at the preceding level of patterns; namely an increase and decrease in local spatial filter activation.

When a stimulus induces a number of decrease- and increase-detection events, each local change-detection can potentially be accounted for by multiple counterchange motion signals (this is simply a reframing of the correspondence problem). However, for the perceiver to make sense of the deluge of sensory events, not all of the higher-level patterns that are potentially entailed by the combinations of low-level patterns are necessary. When low-level events are sufficiently accounted for, additional high-level patterns are unnecessary for sense-making. The unique split/fusion constraint ensures that no motion path accounts for both an increase and decrease event that

are already accounted for by other motion patterns. This does not imply that a low-level event is never multiply accounted for (i.e. take part in multiple higher-level patterns, as in splitting and fusing), but that if it is so it is only to account for a separate low-level event that is not otherwise accounted for. Therefore, the unique split/fusion constraint tends to minimize the number of high-level patterns that are able to account for low-level patterns (although a global minimum is not guaranteed).

Not only do local change-detections combine into counterchange motion patterns, but multiple counterchange motion patterns also combine into higher-level patterns such as the spatially invariant biasing elements that represent a class of motion paths rather than a specific instance of that class. As can be seen most clearly in the simulations of coupled motion quartets, the reciprocal excitation between local motions and collective biasing elements results in global stability when there is agreement between high- and low-level patterns. The greater the number of counterchange motion signals a biasing element is able to account for, the more likely it is for it to remain stabilized against alternative percepts (by virtue of its competitive advantage in the biasing array competition). When a biasing element only accounts for one or two lower-level signals, it does not tend to stabilize the system to the same degree, and the network tends to display a somewhat chaotic dynamic. Thus the system tends to stabilize solutions that maximize the number of low-level patterns accounted for by the minimum number of high-level patterns.

A general organizational principle may be gleaned by approaching the problem in this manner. In the perceptual hierarchy entailed by the model, it may be observed that *vertical* connections (connections between subnetwork components) that exist between congruent low- and high-level patterns interact cooperatively, while *horizontal* connections (connections within a sub network component) interact competitively. Here, congruent is meant to imply a lack of contradiction between two levels of de-

scription. For instance, a counterchange motion pattern is *congruent* with a pair of local increase- and decrease-detections, and vice versa. This scheme allows each hierarchical level to self-organize according to its own internal logic of congruency (e.g. the unique split/fusion constraint for counterchange motion signals) while being biased to achieve congruence with patterns of activation at both higher and lower levels. Such a principle is speculative at present, but is potentially testable.

5.4.9 Neural Correlates

While assigning specific brain areas to the components proposed in the model is speculative, the scheme does fit in well with known anatomy and functional physiology of the primate visual system. The visual cortex is understood to be arranged in a more or less hierarchical fashion. Areas closer to the sensory surface have receptive fields responsive to spatially localized and highly general image features (e.g. oriented edges), with ‘higher’ areas responding to more complex patterns, and often displaying some degree of invariance over spatial translation (i.e. larger receptive fields; Fuster, 2003).

The components in the model are arranged in a clear hierarchical fashion, beginning with localized spatiotemporal filtering combining to form the more complex counterchange motion pattern. Spatial and temporal filtering begin at the retinal surface, and receptive fields remain retinotopically localized through the lateral geniculate nucleus up to at least the primary visual cortex, so all of these areas could fulfill the functional role implicated by the front end of the model.

The extrastriate area referred to as MT is believed to play a crucial role in visual motion processing. This is a likely candidate for the motion pattern detectors, as well as a medium for their interaction. The local motion signals converge in the biasing array, on which neural elements respond invariantly to their preferred class of motion

pattern (i.e. span), and could correspond to an area in parietal cortex, for instance. Elements on the motion-detection and biasing arrays are reciprocally connected to one another, a well-documented feature of primate cortex. The model then gives a hypothetical role to these observed patterns of connectivity; one of mutual biasing across hierarchical levels.

While the local change detectors and motion detectors clearly have a perceptual flavor, the biasing array straddles the line between perception and cognition. Its elements don't represent a *specific* motion path, but a 'class' or 'category' of motion paths. Because of the competitive dynamics, in response to stimuli that generate many individual motion signals, still only one or two biasing units will remain active, acting as global decision units.

When a person participates in a psychophysical experiment and must respond with a keystroke or two to a complex spatiotemporal transformation of the optic array across the retina, they must make a very low-dimensional decision from a high number of incoming signals. In other words, one must *abstract* from the sensory stimulation *invariances* that can be used for meaningful action. The biasing array embodies such a low-dimensional dynamic, and could be functional as a decision-making process.

Closing Remarks and Future Work

It has been argued here that a theory of the perception of object-motion is necessary as distinct from other motion perception processes. Chapter 3 presented strong evidence that commonly used spatiotemporal correlational models of motion detection do not account for human perception in a dense random-dot display, and showed that instead, an account based on counterchange detection showed many of the hallmarks of human perception, for both direction and shape discrimination. Chapter 5 extended this work by proposing a continuous-time dynamical system capable of both generating and selecting counterchange motion signals from a dynamic stimulus.

This work lays down a basic theoretical framework, based on the counterchange motion detection principle, for the continued development of a theory of object-motion perception. The long-term viability of the theory is an open scientific question, and further work will have to be done to support or refute it. Both where it succeeds and fails will deepen the scientific understanding of the problems at hand. Some future directions and implications of the current work are discussed briefly below, followed by a few short closing remarks.

6.1 CHANGE AND STABILITY

It was said in Section 1.3 that perception had to play a dual role of providing stability while remaining flexible to change. The continuous-time dynamical network in Chapter 5 showed an example of how such a process can be achieved. The patterns stabilized by the network in many of the cases showed lifetimes much longer than the change of the stimulus itself. This is perhaps most evident in the motion quartet examples where stable vertical and horizontal solutions lasted for many display-cycles. However these stable regimes are entirely dependent on continuous stimulus input. Without dynamic input, all nodes on the network return to resting level and no interesting dynamics take place. This allows the network to remain very responsive to the changing input even while promoting stability. Future work should address the interplay of such a stimulus-driven network with self-stabilizing networks that don't necessitate input to maintain non-trivial dynamics.

6.2 PERCEPTION AND INTENTIONALITY

This dissertation has presented visual perception as an essentially passive process. That is, stimulation comes from the optic array, feeds into the neural networks where motion is detected and selection takes place, and ultimately a percept is formed. No role is given to the perceiving agent in which such a perceptual system is presumably embedded. This is clearly a simplifying assumption. Perceptual systems we find in the natural world are *for* something; organisms *use* perception to accomplish meaningful behavior. Self-interested agents move about the world meaningfully, they have *intentions* that govern the selection of behavior. How the intention of an agent results in causal efficacy in the physical world is a philosophical problem at the heart of perceptual science.

Despite the nature of this philosophical mystery, there is evidence that intentions not only select action, but perception to some degree as well. Experiments have shown that having intentions of either promoting perceptual stability or, conversely, perceptual switching of multi-stable stimuli leads to significant effects in switching rates as compared to conditions of ‘passive’ viewing, where an observer attempts to simply observe (see, e.g. Kohler, Haddad, Singer, & Muckli, 2008). That intentions can affect percepts suggests a complex interplay between the ‘highest’ levels of cognition and ‘low’ level perception. The degree and limits of influence that intention can have on perception should be an area of continued investigation.

Anecdotal evidence as well as (internal) unpublished pilot data suggest that within an appropriate parameter range, perceptual switching can be intentionally induced on every frame-change of the bistable motion quartet stimulus. If this finding is confirmed, it goes beyond giving intentionality the role of merely affecting the rates of switching of a bistable stimulus. Rather, having such ‘continuous control’ over the interpretation of a visual stimulus suggests the existence of *perceptual know-how* that bears a strong resemblance to the kind of sensorimotor know-how employed when a skilled agent performs coordinated, functional motor behaviors. Also of note is the anecdotal evidence that there are individual differences evident in the degree to which a multi-stable percept can be intentionally controlled. Likely, spending considerable time with a particular class of stimulus improves the ability to control the interpretation of that stimulus. This acquisition process of perceptual know-how is also of extreme scientific interest, and studying this could provide clues about the nature of learning to control oneself in general.

The role of know-how in cognition has been emphasized most explicitly in so-called *embodied* theories of cognition. Without going into the details and differences between various incarnations of embodied approaches to cognition, in general they argue that

ongoing motor engagement is not only a product of perception and cognition, but an integral part of their essential character. Visual displays like the ones used in this manuscript present a challenge to such a framing. The apparent motion stimuli described here induce the most compelling motion percepts when the eye remains fixated during frame-changes. While this does not rule out a role for micro-saccades, it does make it difficult to assign a role to overt motor behavior in the formation of these motion percepts. However, if we take a subtler stance on embodiment, such as that of Sandamirskaya, Zibner, Schneegans, & Schöner (2013), we can see that the neural dynamics themselves have the potential to *embody action*, where action can be understood in the broad sense of functional biological behavior (such as the firing of action potentials).

6.3 INTERPLAY BETWEEN RETINAL MOTION AND EYE-MOVEMENTS

This dissertation addressed motion percepts relative to the retina as signified by counterchange motion. Of course, in addition to fixations, eye movements play an integral part in active perception, especially in the context of perceiving moving objects. Future work should address the interplay of these two complementary aspects of visual perception in the context of dynamic stimuli.

There are already some hints about a possible connection between counterchange motion perception and saccades. Schütz (2013) in a free-viewing experiment (i.e. no task) showed a strong tendency for gaze to be repelled by decreases in the contrast of a visual element and attracted by increases in contrast. While counterchange motion perception is typically studied under conditions of fixation, this presents the intriguing possibility that the counterchange pattern may also play a role in motor behavior.

Furthermore, it is conceivable that the perception of counterchange under conditions of fixation could be related to an active inhibition of motor behavior.

6.4 CONCLUSION

Studying perception means studying life. Perception is how we make sense of our world. Even when we act, it is through perception that we feel our actions. Thus, solving the problem of perception is, in a sense, to solve all problems. Even the entirety of the scientific enterprise is established and maintained solely through the perception of purposeful agents acting in the world. Not only embedded in the mystery, but composed of it.

There are, no doubt, mechanisms that enable our perceptual capacities, such as the ones theorized about in this dissertation. However, it remains an open question as to whether *mechanism* is the source of perception or merely the medium of it. How do we come to consciously know our world and our selves? What is our relationship to the world? And why are we aware at all? These are the seemingly bottomless questions, always lurking under and behind our abstracted description of *the real thing*. A science of life and a science mind may indeed be inseparable, and both must come to understand the nature of agency and autonomy. Perception will almost certainly be at the heart of such an understanding, if such an understanding is possible.

APPENDIX A

COMPUTATIONAL SIMULATIONS FOR CHAPTER 3

A.1 STIMULUS

The stimuli each consisting spatially of 240 bars and temporally of two frames, are defined as $S(x, t)$ at all locations along the stimulus array $x = [1960]$. Each random bar is composed of 4 pixels, with each location taking on the value 0 (representing black) or 1 (representing white) and $t = 1, 2$ (representing frames 1 and 2). A central figure region (either 60 or 120 bars long) is translated to the right by 2, 4, 6, 8, 10, 12, 14, or 16 bar-widths from frame 1 to frame 2, while the background regions are independently and randomly generated for each frame.

A.2 EDGE FILTERS

One-dimensional real-valued Gabor functions (a Gaussian window modulated by a sine function) are used for all spatial filtering in both the ERD and counterchange detectors. The function is centered around zero, and uses a 0-phase sine-wave modulator so that it serves as a balanced receptive field by virtue of it being anti-symmetrical around zero. The filter is described by the equation $g(x)$ below. Parameter σ sets the standard deviation of the Gaussian window, and parameter p sets the period of the sine wave modulator (in dot units). The ratio between the two parameters is the same in all edge filter instantiations, regardless of scale ($p/\sigma = 5$). Finally, edge-filters are normalized such that their positive lobes always are integrated to 1, and

their negative lobes to -1.

$$g(x) = w(x, \sigma)c(x, p)$$

$$w(x) = e^{-\frac{x^2}{2\sigma^2}}$$

$$c(x) = \sin(2\pi \cdot \frac{1}{p} \cdot x)$$

A.3 IMPLEMENTATION OF THE ERD (FIGURE 3.4, PANEL A)

Four scales of edge-filters are used for the ERD simulation. This is necessary in order to approximate quadrature for motion detectors with differing spans (i.e. distance between the center of the receptive fields that serve as inputs to a motion detector). Parameter values are listed below, with subscripts indicating layer numbers with layers 1, 2, 3, and 4 corresponding to motion detectors with spans of 2, 4, 6, and 8 bars, respectively.

$$p_1 = 8; p_2 = 16; p_3 = 24; p_4 = 32$$

$$\sigma_1 = 1.6; \sigma_2 = 3.2; \sigma_3 = 4.8; \sigma_4 = 6.4$$

For each motion detector layer, the entire one-dimensional stimulus is convolved with the corresponding edge-filter kernel (convolution being notated by $*$) for both frames (here notated with the index i). The result is truncated at both ends to maintain the original stimulus size.

$$r_i(x) = g(x) * S(x, i)$$

For each location along the detector array a motion signal $m(x)$ is calculated by

$$m(x) = r_1(x)r_2(x + x') - r_1(x + x')r_2(x)$$

where $r_i(x)$ is the edge-filter response at location x for frame i and x' is the magnitude of the detector span corresponding to a given motion detection layer. The resulting

array is padded with zeros in order to maintain the one-to-one correspondence between the motion detector array and the stimulus.

A.4 IMPLEMENTATION OF THE COUNTERCHANGE DETECTOR (FIGURE 3.4, PANEL B)

Only one scale of edge-filter was used for the counterchange detector, regardless of the span. The parameters were

$$p = 2; \sigma = 0.4.$$

Both frames of the stimulus are convolved with the edge filter. In a computational shortcut, two polarity channels are derived from the filter response by half-wave rectifying the filter responses to form channel 1 and inverting and half-wave rectifying the responses to form channel 2.

Half-wave rectification:

$$h(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Channel 1 responses for frames $i = [1, 2]$:

$$r_{1i}(x) = h(r_i(x))$$

Channel 2 responses for frames $i = [1, 2]$:

$$r_{2i}(x) = h(-r_i(x))$$

The frame-to-frame change in filter response is calculated as:

$$c_c(x) = r_{c2}(x) - r_{c1}(x)$$

Where subscript c stands for channel $c = [1, 2]$ and the second subscript refers to the frame index.

Decrease and increase responses for each channel are calculated by taking the half-wave rectified change-values for the increase response, and inverting and half-wave rectifying for the decrease responses.

$$i_c(x) = h(c_c(x))$$

$$d_c(x) = h(-c_c(x))$$

For each location along the detector array of a given span, motion is computed for both channels and summed into a single motion vector.

$$m(x) = \sum_{c=1,2} d_c(x) \cdot i_c(x + x') - d_c(x + x') \cdot i_c(x)$$

x' equals the span of a given layer. Motion response arrays are padded with zeros in order to maintain correspondence with the stimulus. Finally, motion responses over a threshold (.2 in the reported simulations) inhibit longer range motions originating from the same decrease locations (i.e. the inhibited motions are set to 0). If two motions sharing a decrease location are of the same span but opposite directions, the stronger response is taken and the other set to 0, or else for equal strength motions one is selected with a 0.5 probability and the other set to 0. Thresholding prevents near-zero responses from contributing to inhibitory interactions, but the exact size of this threshold did not have much effect as individual counterchange responses tended to be very vigorous or very weak (i.e. well above or well below threshold). These interactions serve as a weak shortest-path assumption in the counterchange model (weak because splitting motions are prevented, but converging motions are not).

A.5 DIRECTION DECISIONS

After each trial, for each of the two detector arrays (ERD and counterchange), the motion responses are summed across all locations and spans. Rightward motion was signified by positive values, and leftward motion by negative values.

A.6 SHAPE DECISIONS

Two templates that corresponded to the two one-dimensional figures are used to make a shape decision, long vs. short line, after each trial. The templates consist of an interior positive region that corresponds to the figure sizes (60 and 120 bars), and flanking negative regions that extended to the edge of the stimulus. The value is homogenous across the interior region (i.e. the same at all locations) and likewise is homogenous across the flanking regions. The interior regions were normalized such that they integrated to a value of 1, and the flanking regions are normalized to integrate to a value of -1. After a trial, rightward and leftward motions for each detection layer are separated in order to assess their template response independently (leftward motions were made positive so that only positive responses were considered as template matches). Each span-layer (separated by direction) is correlated with both templates. The corresponding figure of the maximum template response is taken as the shape decision for a given trial.

APPENDIX B

SYMMETRY OF ELABORATED REICHARDT DETECTOR TO

TWO-FRAME SAME- AND INVERTED-POLARITY STIMULI FOR

CHAPTER 3

Consider a 2-frame stimulus J in which frame 1 is some spatial function $f(x)$ and frame 2 is some other spatial function $g(x)$. A 'point-delay' Reichardt detector exposed to stimulus I can then be described as

$$R_{x_1, x_2, \delta_t}[I](t) = I(x_1, t - \delta_t)I(x_2, t) - I(x_1, t)I(x_2, t - \delta_t)$$

where x_1 and x_2 are the two points in space which the detector is sensitive to and δ_t is the delay used to detect motion across the two points. In response to the 2-frame stimulus J , $R_{x_1, x_2, \delta_t}[J](t)$ will be zero except for those t during frame 2 for which frame 1 was on at time $t - \delta_t$. For all such t ,

$$R_{x_1, x_2, \delta_t}[J](t) = f(x_1)g(x_2) - f(x_2)g(x_1).$$

Now consider the inverted-polarity version of the same stimulus K with the same first frame $f(x)$ but in which spatial function of the second frame $h(x)$ is the opposite of $g(x)$.

$$h(x) = -g(x)$$

$$\begin{aligned} R_{x_1, x_2, \delta_t}[K](t) &= f(x_1)h(x_2) - f(x_2)h(x_1) \\ &= f(x_1)(-g(x_2)) - f(x_2)(-g(x_1)) \\ &= -(f(x_1)g(x_2) - f(x_2)g(x_1)) \end{aligned}$$

Thus,

$$R_{x_1, x_2, \delta_t}[K](t) = -R_{x_1, x_2, \delta_t}[J](t).$$

Further, Chubb and Sperling (1988) proved that the response of any Reichardt detector with arbitrary spatial and temporal filters can be expressed as a linear combination of point-delay Reichardt detector responses. This implies that the response of any given Reichardt detector, regardless of its spatial or temporal sampling characteristics, is the negative of the detector's response to an otherwise identical 2-frame stimulus in which the luminance polarity of the second frame is inverted. This is true regardless of whether or not the two frames represent a spatiotemporally shifted pattern (i.e. motion) or not (i.e. noise).

APPENDIX C
IMPLEMENTATION, PARAMETERS, AND VARIABLES FOR
CHAPTER 5

C.1 EQUATIONS

Decrease detector array:

$$\begin{aligned}\tau \dot{u}_m^{dec} &= -u_m^{dec} + h^{dec} - v_m^{dec} - S(m, t) + \xi^{dec}(m, t) \\ \tau^{slow} \dot{v}_m^{dec} &= -v_m^{dec} - S(m, t)\end{aligned}\tag{C.1}$$

Increase detector array:

$$\begin{aligned}\tau \dot{u}_n^{inc} &= -u_n^{inc} + h^{inc} - v_n^{inc} + S(n, t) + \xi^{inc}(n, t) \\ \tau^{slow} \dot{v}_n^{inc} &= -v_n^{inc} + S(n, t)\end{aligned}\tag{C.2}$$

Motion detector array:

$$\begin{aligned}\tau \dot{u}_{m,n}^{mot} &= -u_{m,n}^{mot} + h^{mot} \\ &+ A \cdot f(u_m^{dec}) \cdot f(u_n^{inc}) \\ &+ B \cdot \sum_{p \neq m} f(u_{p,n}^{mot}) \cdot \sum_{q \neq n} f(u_{m,q}^{mot}) \\ &+ C \cdot f(u_{(n-m)}^{bias}) \\ &+ \xi^{mot}(m, n, t)\end{aligned}\tag{C.3}$$

Biasing array:

$$\begin{aligned}
\tau \dot{u}_{\delta x}^{bias} &= -u_{\delta x}^{bias} + h^{bias}(\delta x) \\
&+ D \cdot \sum_{n-m=\delta x} g(u_{m,n}^{mot}) \\
&+ E \cdot \sum_{r \neq \delta x} f(u_r^{bias}) \\
&+ \zeta^{bias}(\delta x, t)
\end{aligned} \tag{C.4}$$

C.2 PARAMETER AND VARIABLE DEFINITIONS AND VALUES

C.2.1 Time constants

$\tau = 30$: time constant of dynamics

$\tau_{slow} = 60$: slower time-constant of antagonistic dynamic in change-detection neurons

C.2.2 Neuron state variables

$u_{m,n}^{mot}$: activation variable of a neural element

u_m^{dec} : activation variable of a decrease detector

v_m^{dec} : antagonistic component of decrease detection neuron

u_n^{inc} : activation variable of an increase detector

v_n^{inc} : antagonistic component of increase detection neuron

$u_{\delta x}^{bias}$: activation variable of a biasing element

C.2.3 Indices

m : Indexes the decrease-detector array and motion-detector array with respect to decrease-detection input. For the two-dimensional motion network, $m = (m_1, m_2)$.

n : Indexes the increase-detector array and motion-detector array with respect to increase-detection input. For the two-dimensional motion network, $n = (n_1, n_2)$.

$\delta_x = n - m$: Indexes the biasing array. For the two-dimensional motion network,
 $\delta_x = (\delta_{x_1}, \delta_{x_2})$.

C.2.4 Resting levels

$h^{dec} = -10$: resting level of decrease detectors

$h^{inc} = -10$: resting level of increase detectors

$h^{mot} = -50$: resting level of motion detectors

$h_{\delta_x}^{bias} = \beta + \gamma \|\delta_x\|$

$\beta = -15$: global offset of biasing array resting levels

$\gamma = 2$: rate of biasing array resting level drop-off

C.2.5 Noise terms

ξ : normally distributed with mean 0 standard deviation 10. Each noise term is independent with superscripted index referring to each sub-network array and is a function of the neural element (referred to by its corresponding index) and time.

C.2.6 Interaction functions

$f(u)$: sigmoidal interaction function of neural element's state variable u

$g(u)$: sigmoidal interaction function of neural element's state variable u

$$f(u) = \frac{1}{1+e^{-u}}$$

$$g(u) = \frac{1}{1+e^{(-u+20)}}$$

C.2.7 Stimulus input

$I(x, t)$: dynamic motion stimulus where $x = (x_1, x_2)$ for the two-dimensional correspondence network.

$c(x) = [-\frac{1}{2}, 1, -\frac{1}{2}]$: 1-dimensional spatial filter

$$c(x) = \begin{bmatrix} -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \\ -\frac{1}{8} & 1 & -\frac{1}{8} \\ -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \end{bmatrix} : \text{2-dimensional spatial filter}$$

$\epsilon = 80$: scaling factor of filtered stimulus

$S(x, t) = \epsilon(I(x, t) * c(x)) \forall t$: spatially filtered and scaled motion signal

C.2.8 Synaptic weights

$A = 80$: synaptic weight for excitatory feedforward motion detection

$B = -120$: synaptic weight for inhibitory competitive dynamics

$C = 20$: synaptic weight for feedback from biasing array array

$D = 80$: synaptic weight from the motion-detection network to the biasing array

$E = -100$: synaptic weight of winner take all inhibition on collective biasing array

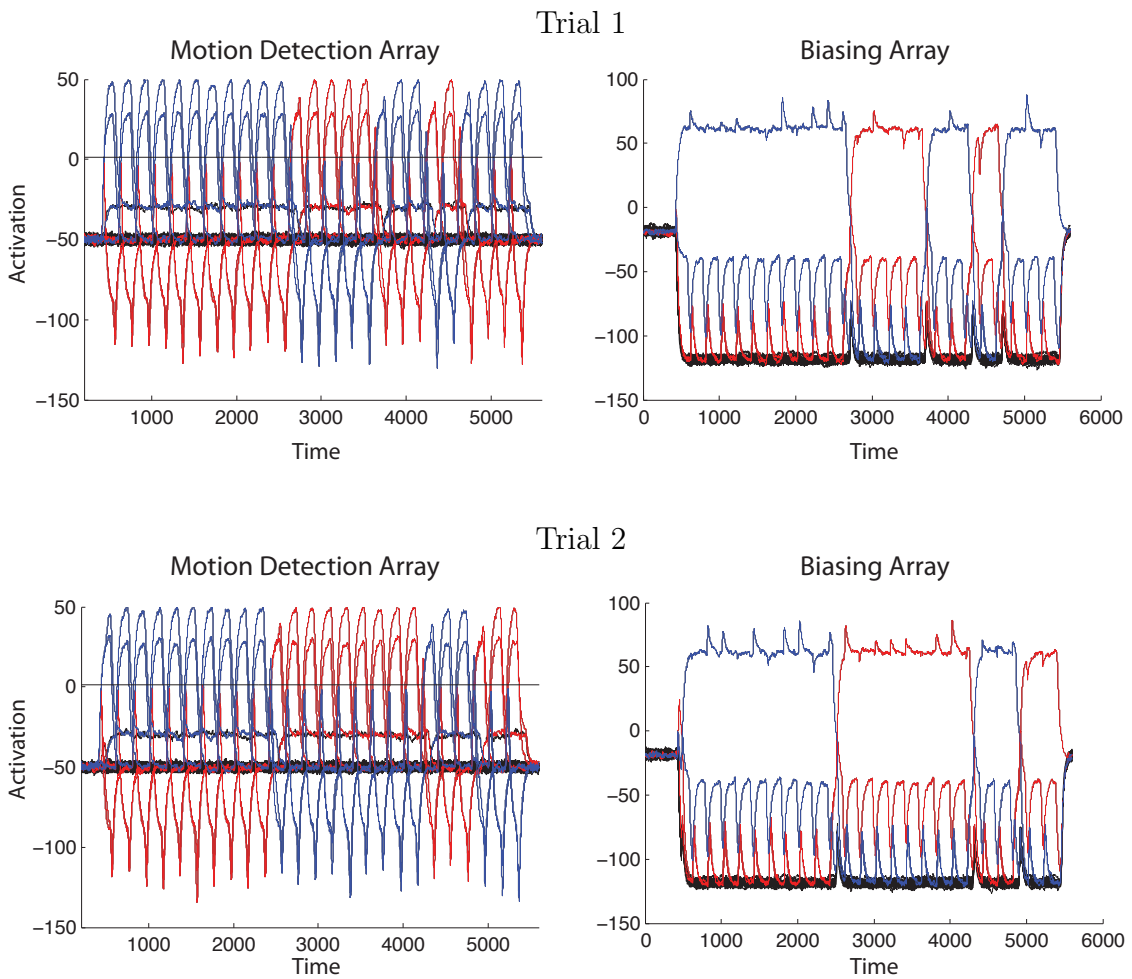
C.3 IMPLEMENTATION OF ADDITIVE INHIBITION NETWORK FOR SECTION 5.4.1

$$\begin{aligned} \tau \dot{u}_{m,n}^{mot} &= -u_{m,n}^{mot} + h^{mot} \\ &+ A \cdot f(u_m^{dec}) \cdot f(u_n^{inc}) \\ &+ B \cdot \sum_{p \neq m} f(u_{p,n}^{mot}) + B \cdot \sum_{q \neq n} f(u_{m,q}^{mot}) \\ &+ \xi^{mot}(m, n, t) \end{aligned} \tag{C.5}$$

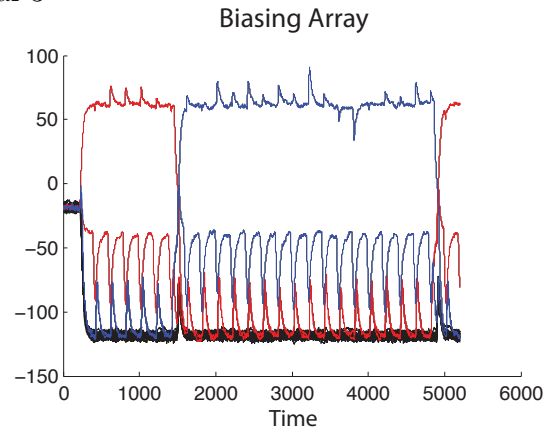
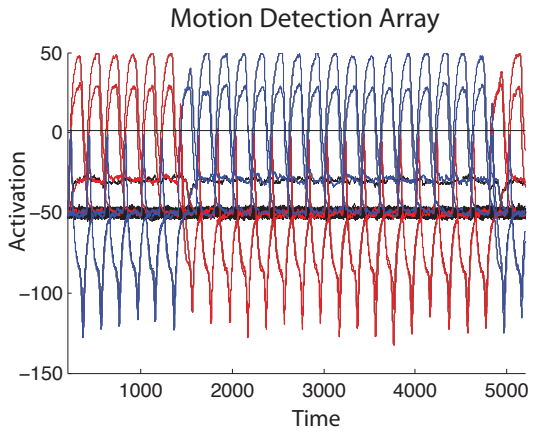
Where $B = -20$, and all other parameters are the same as above. Note, in this formulation there is no biasing array. This has no effect on the qualitative results of the corresponding simulations.

APPENDIX D
COMPREHENSIVE RESULTS FOR MOTION QUARTET
SIMULATIONS IN CHAPTER 5

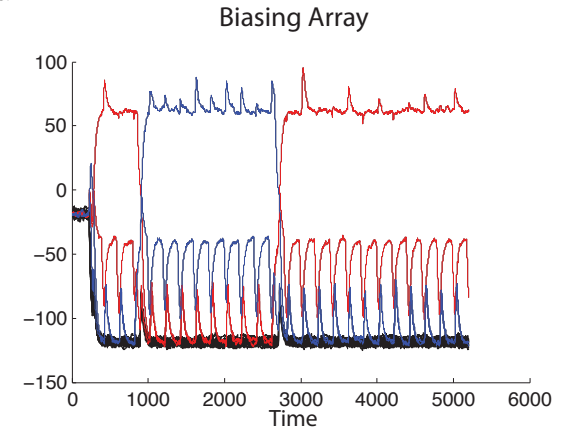
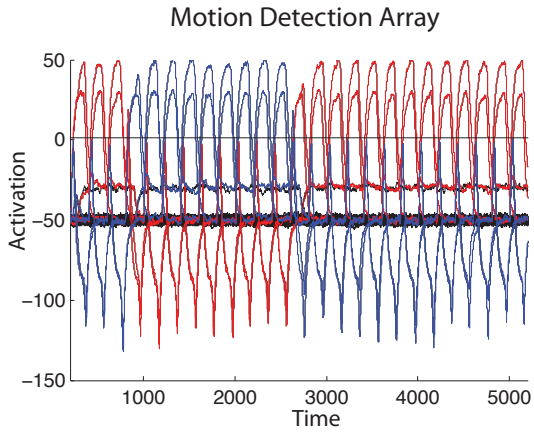
D.1 SINGLE MOTION QUARTET



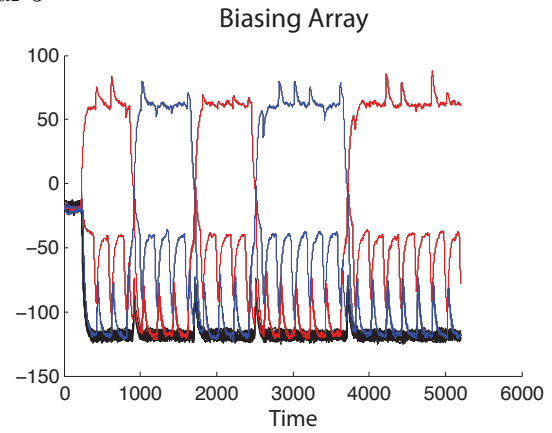
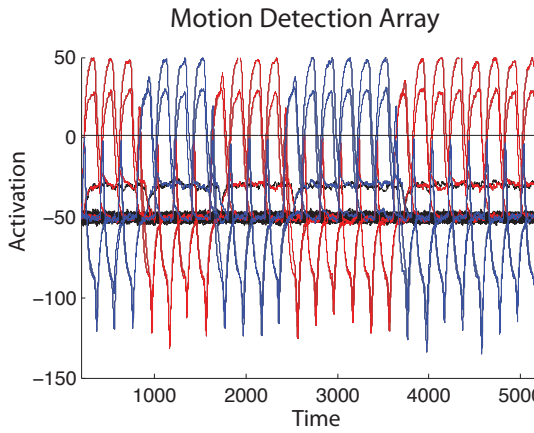
Trial 3



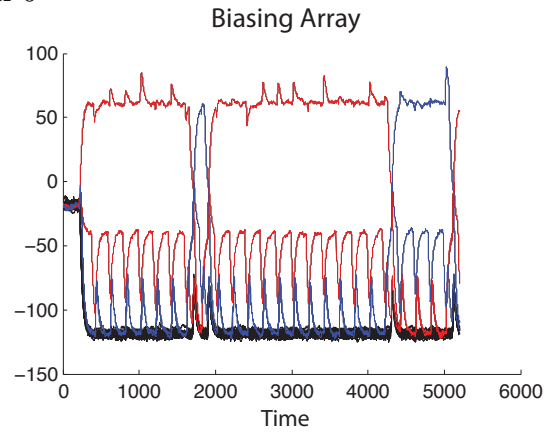
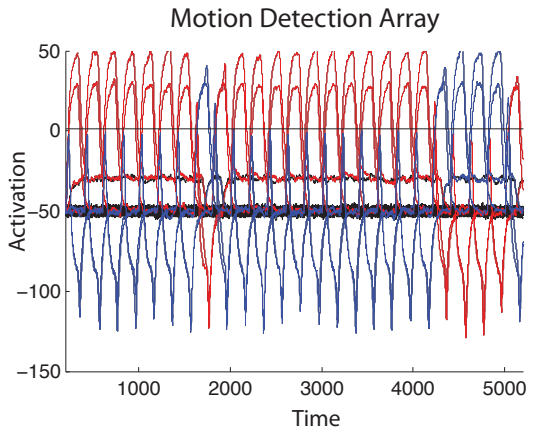
Trial 4



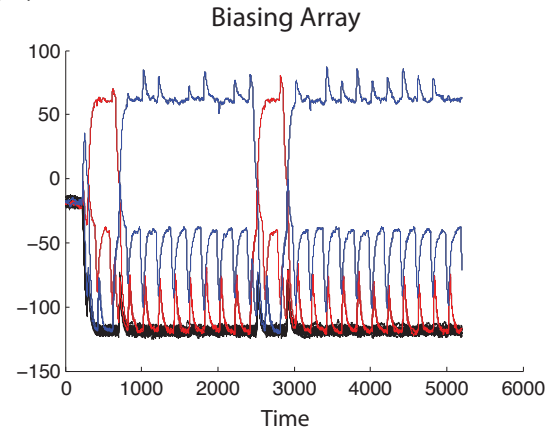
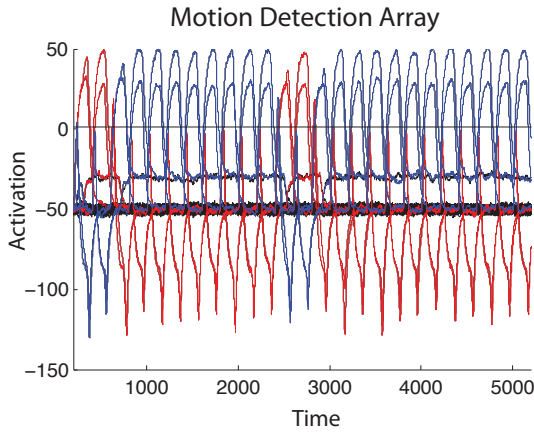
Trial 5



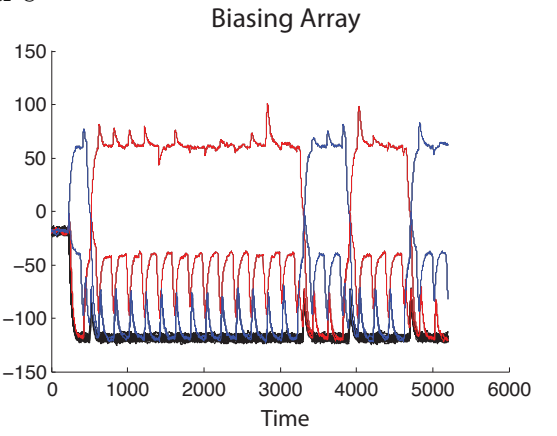
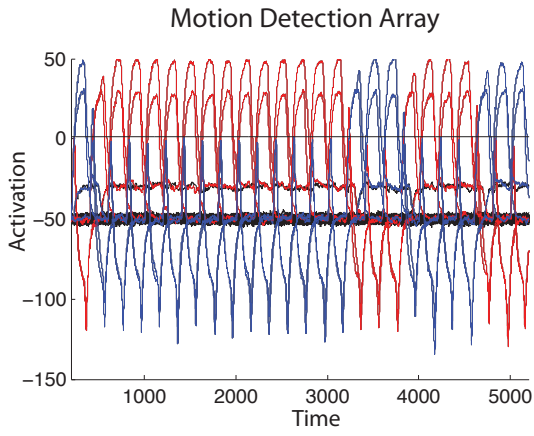
Trial 6



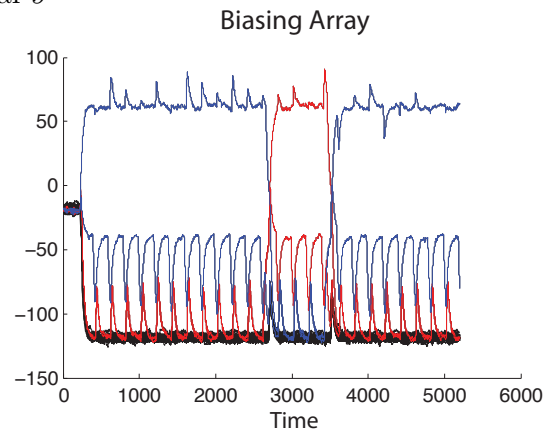
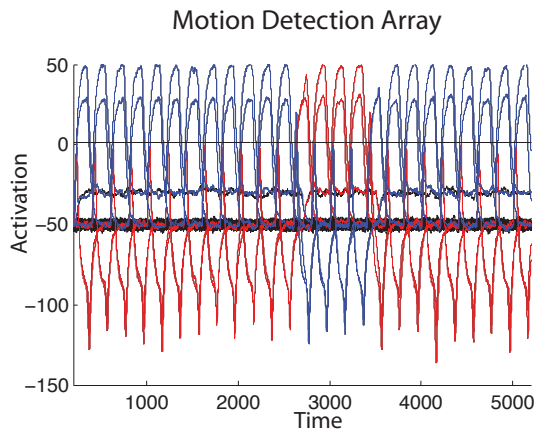
Trial 7



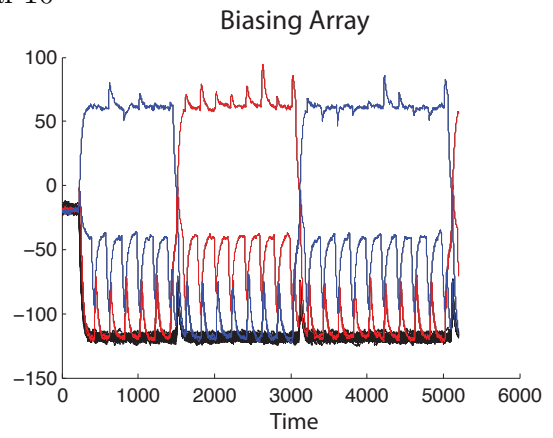
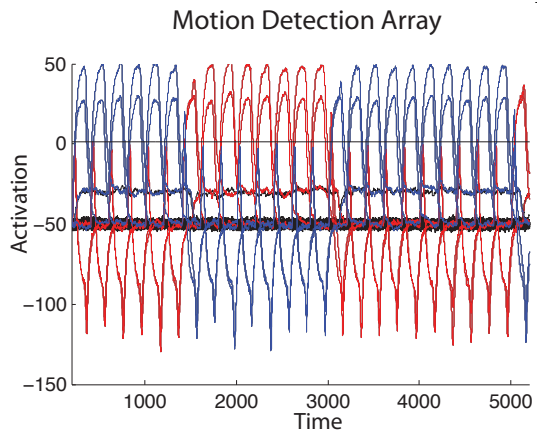
Trial 8



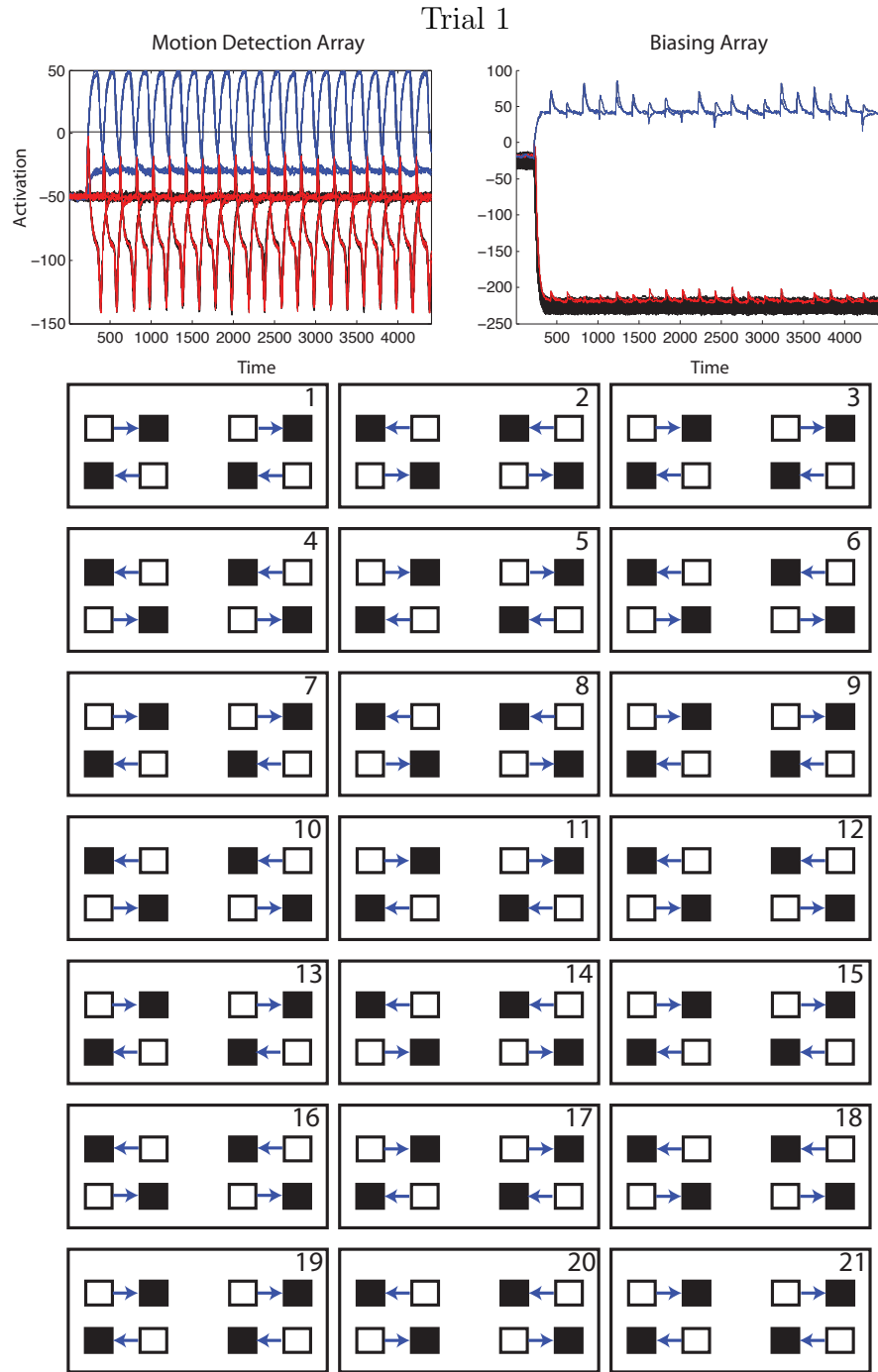
Trial 9



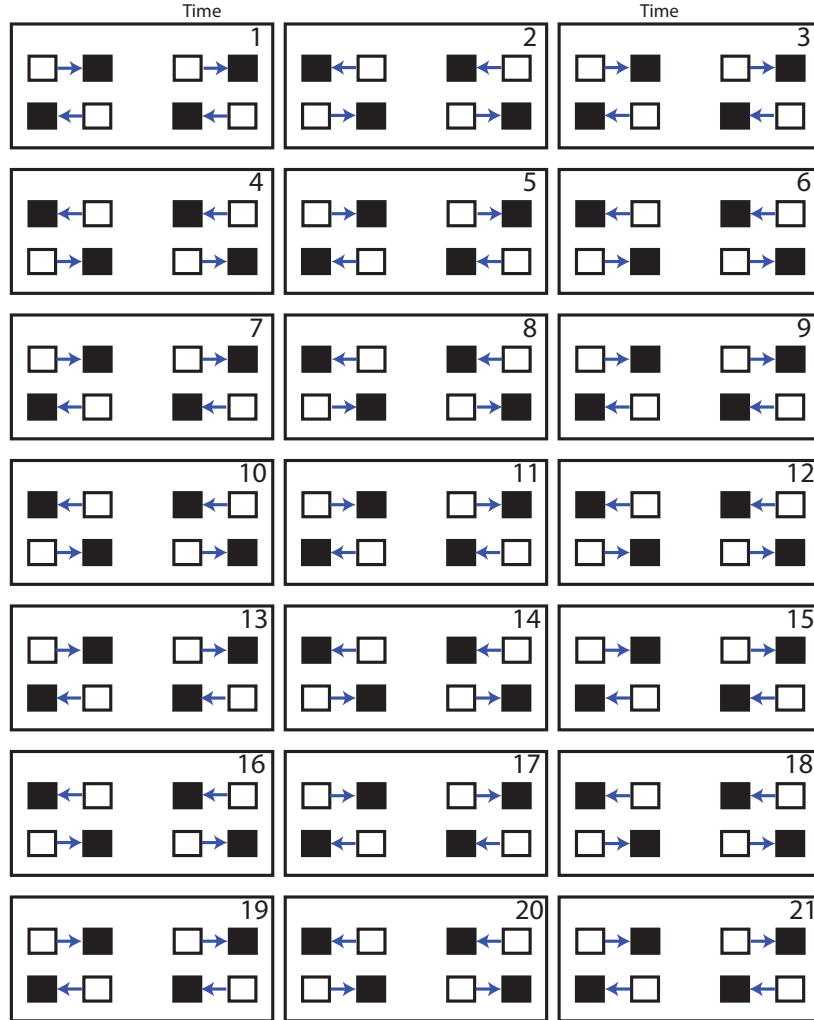
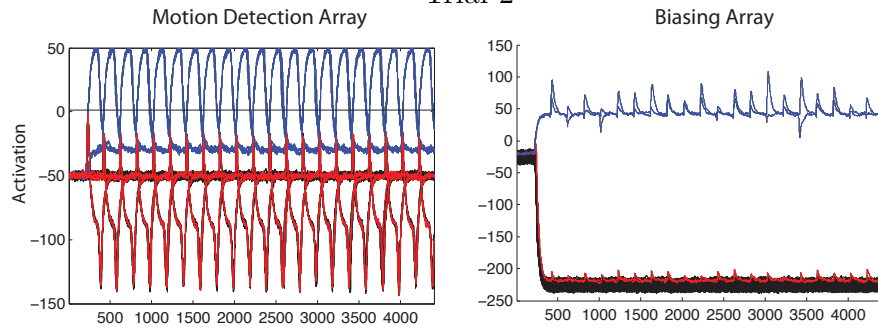
Trial 10



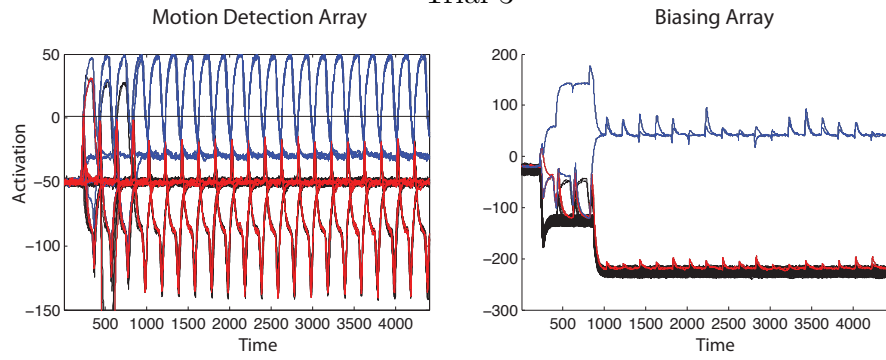
D.2 TWO QUARTETS



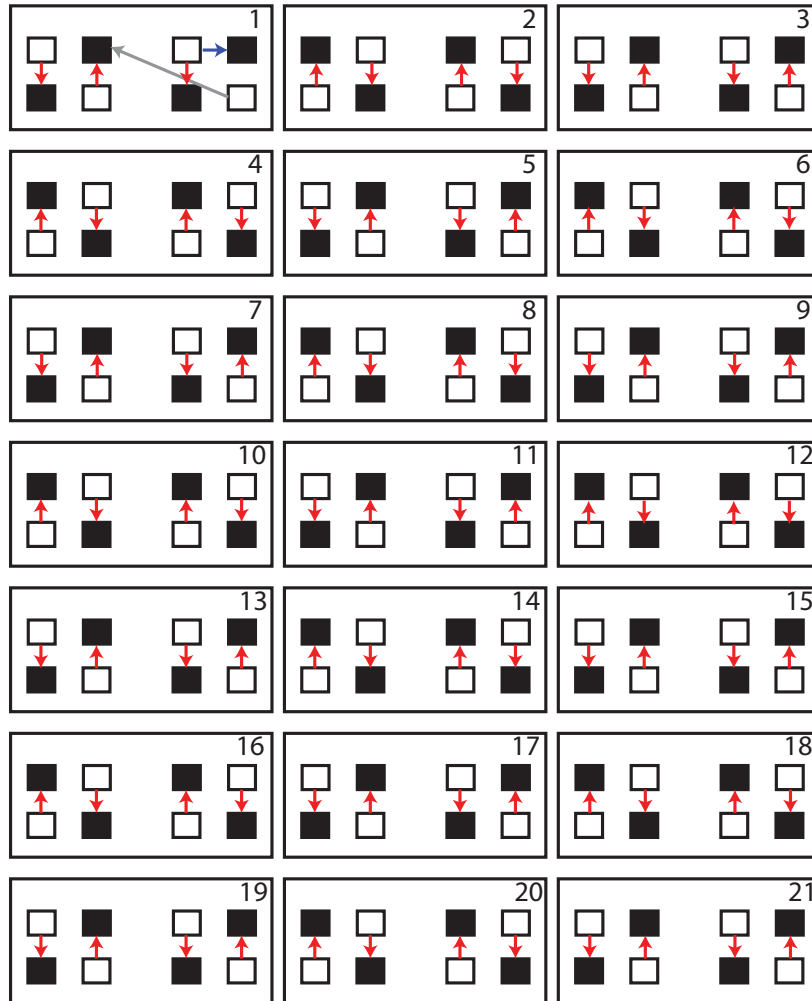
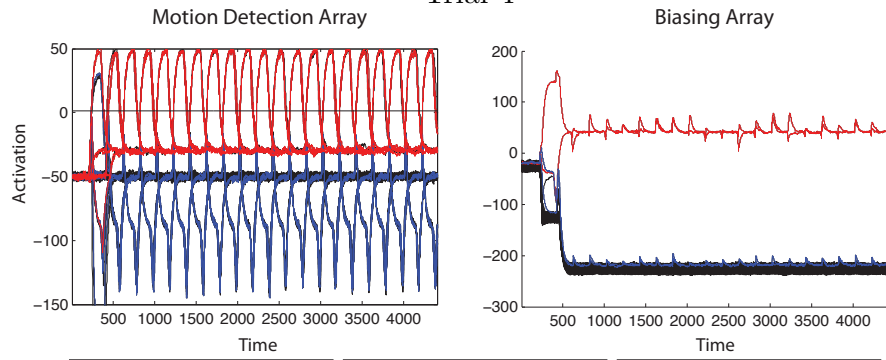
Trial 2



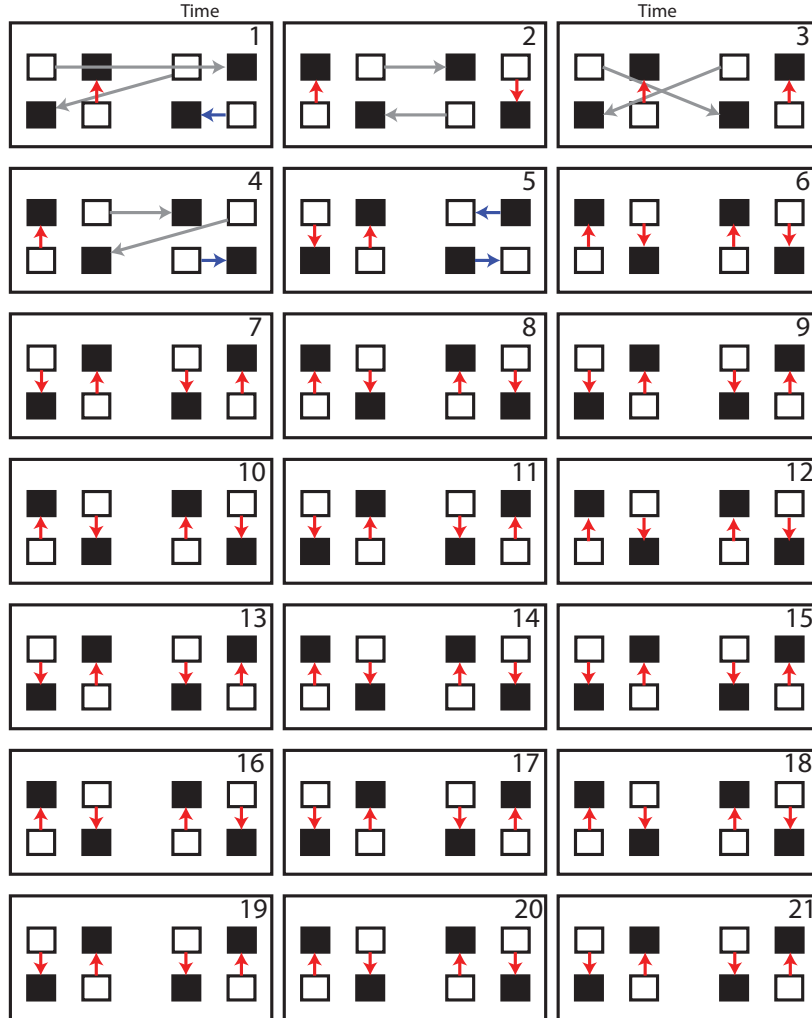
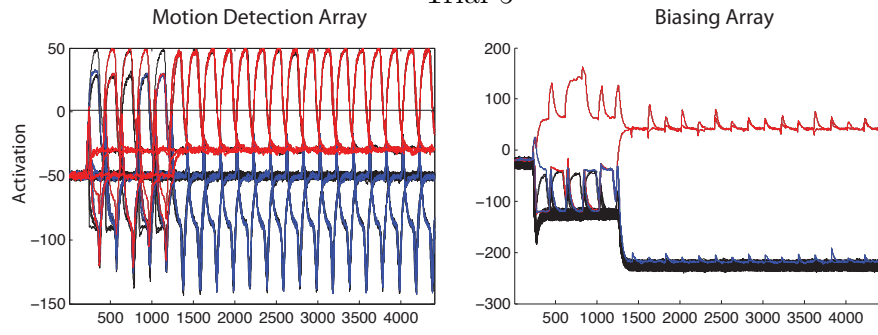
Trial 3



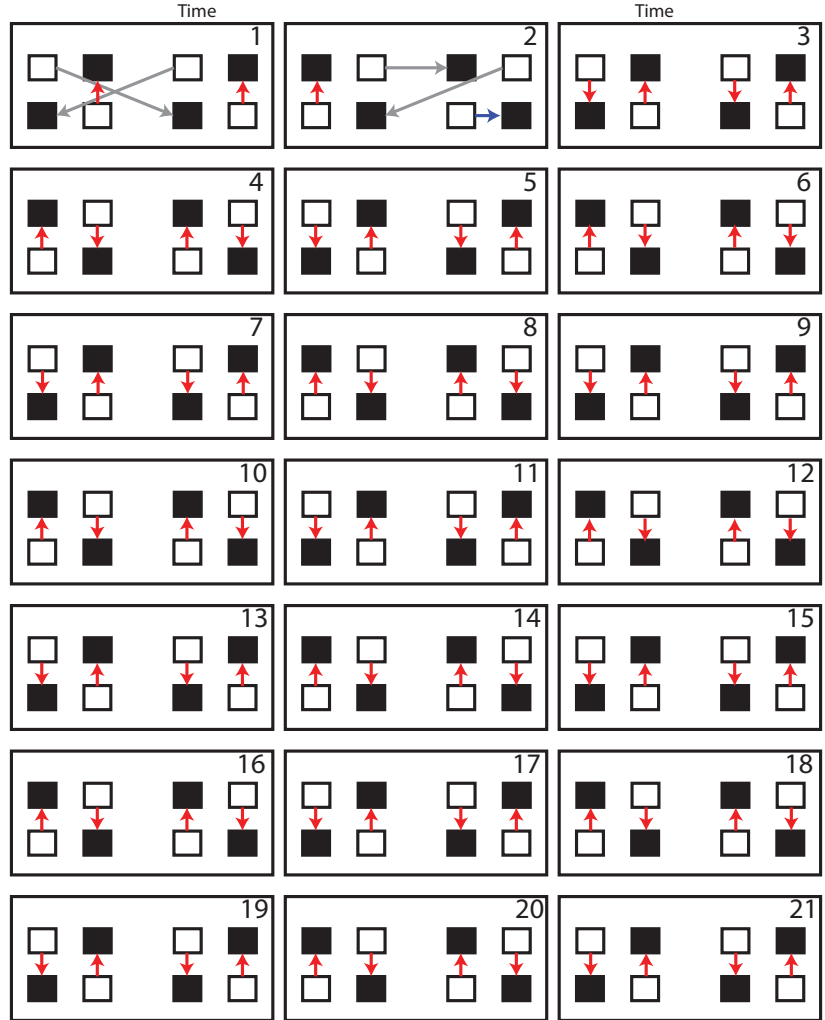
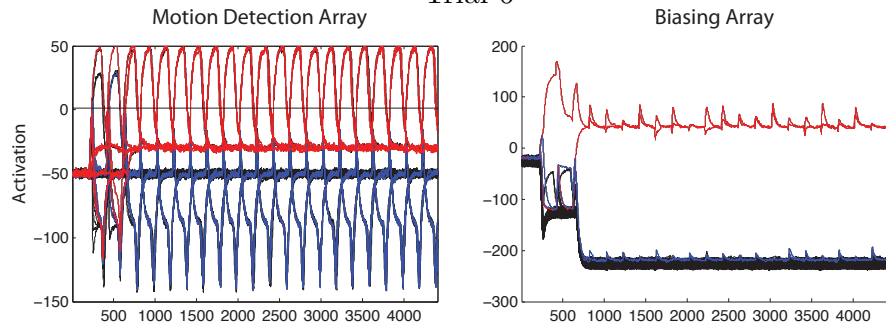
Trial 4



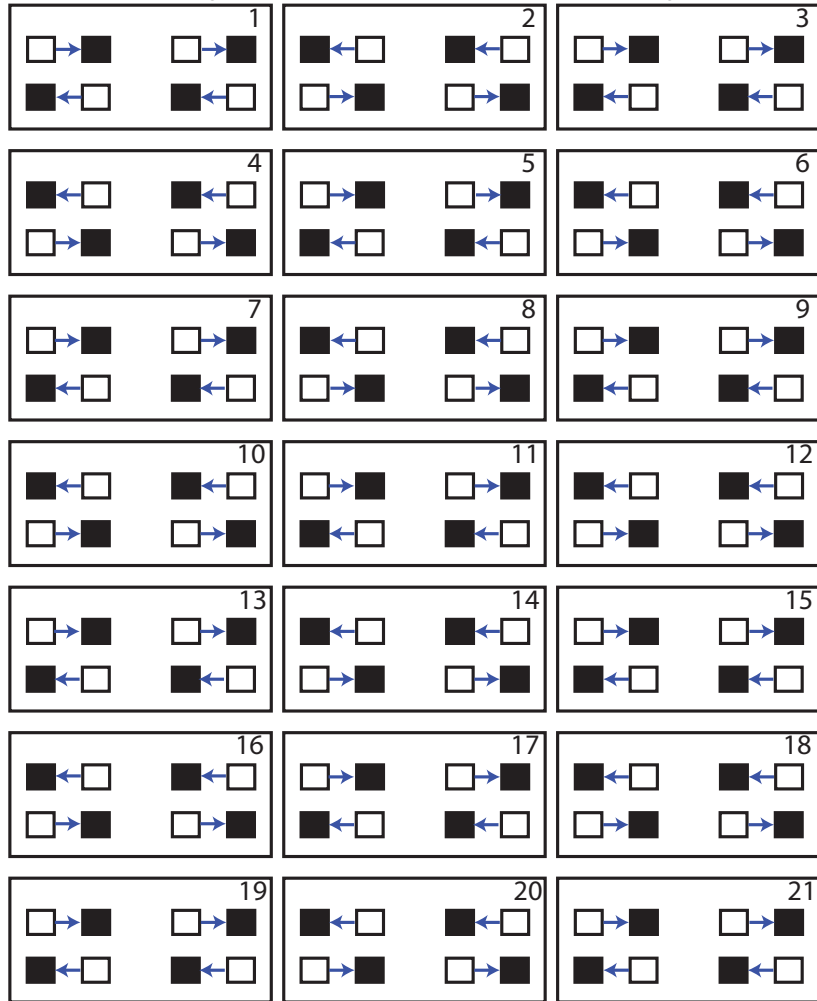
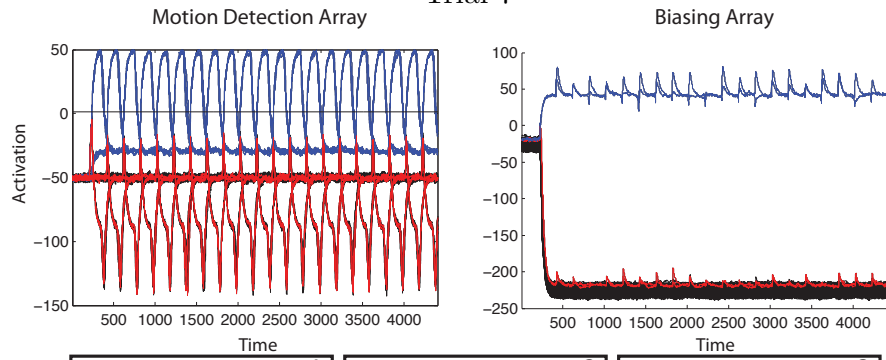
Trial 5



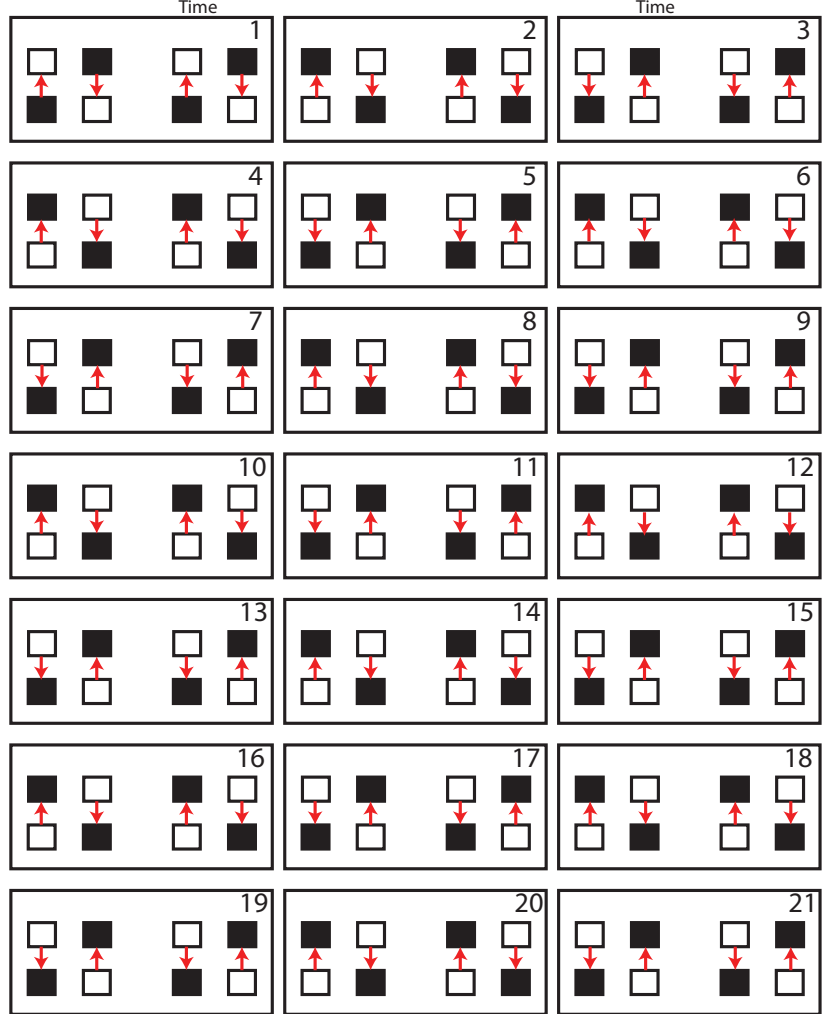
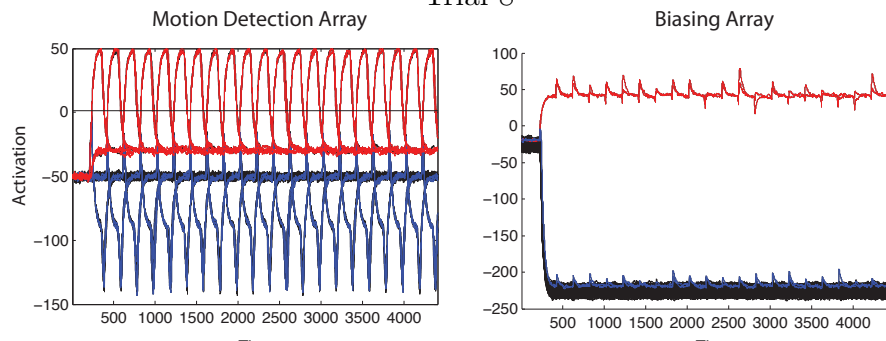
Trial 6



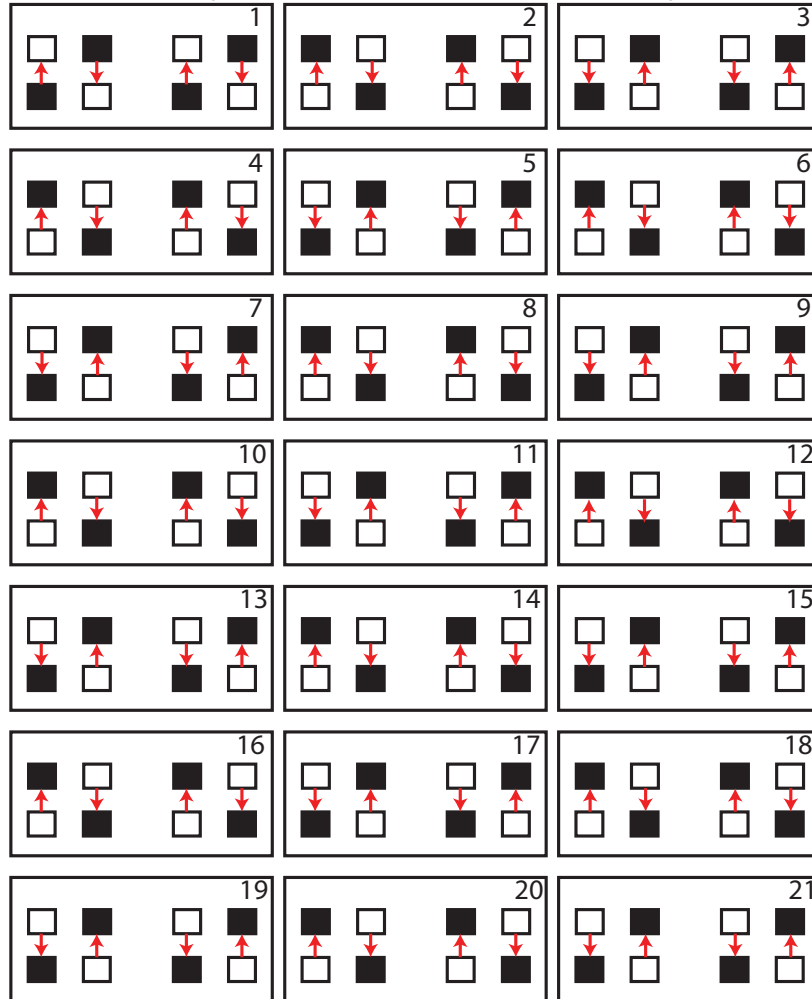
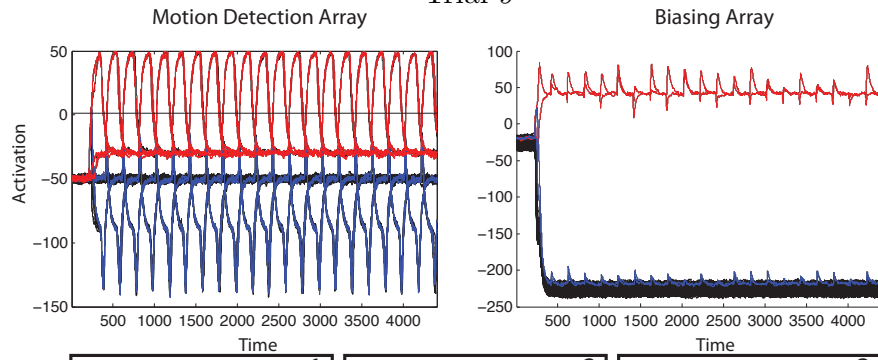
Trial 7



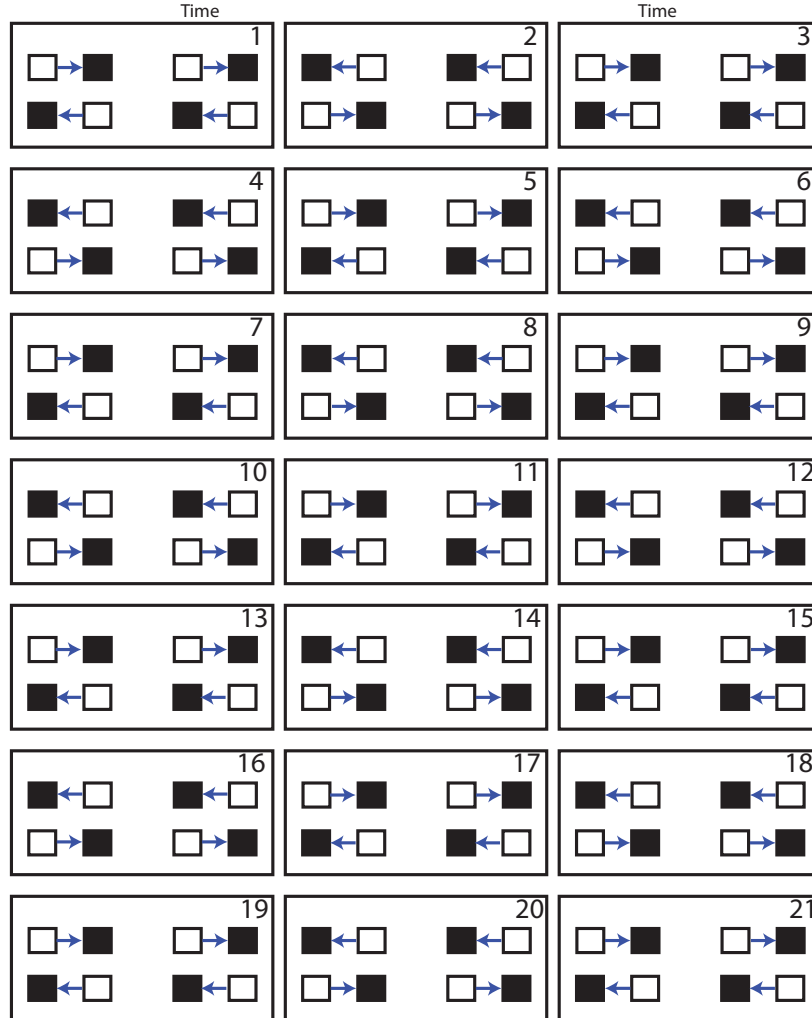
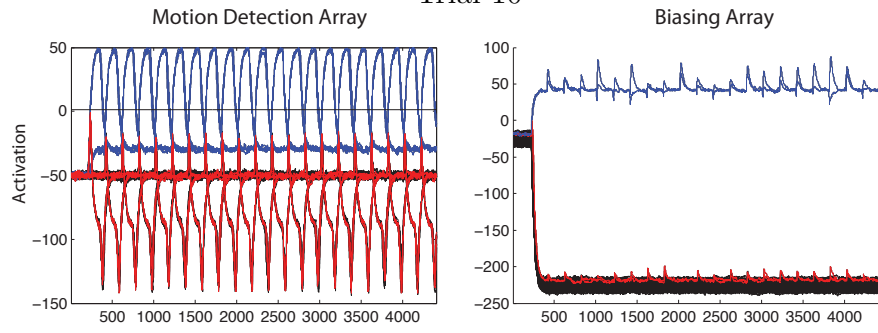
Trial 8



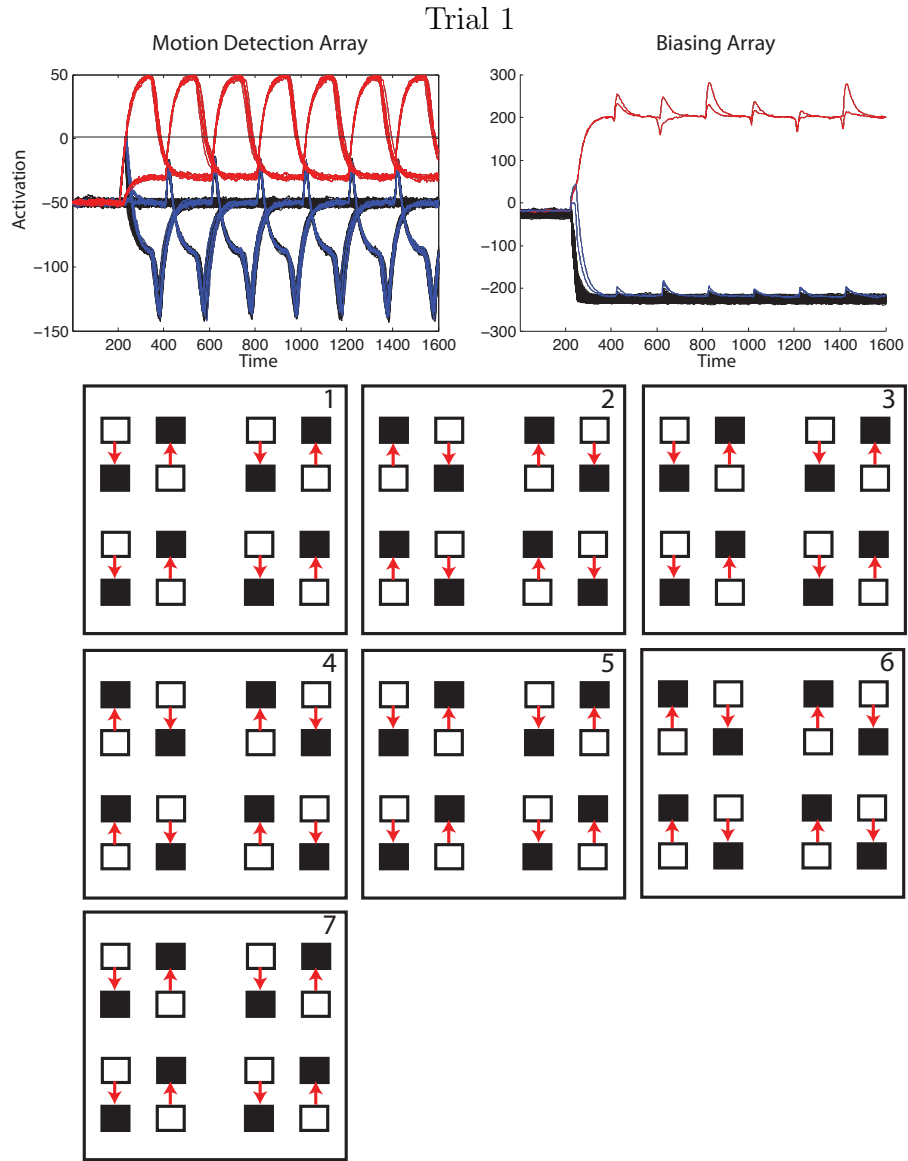
Trial 9



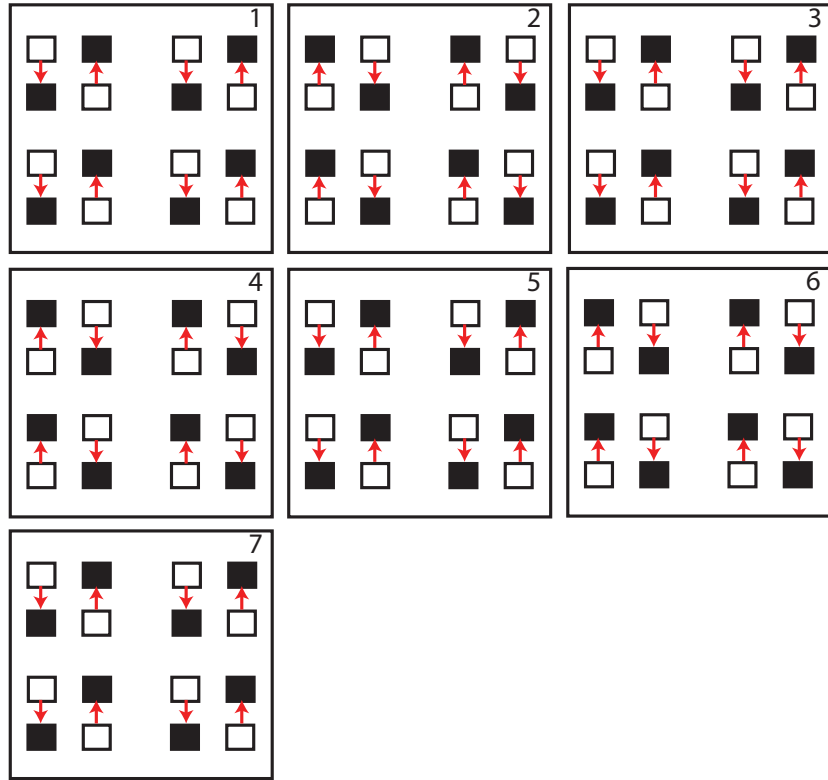
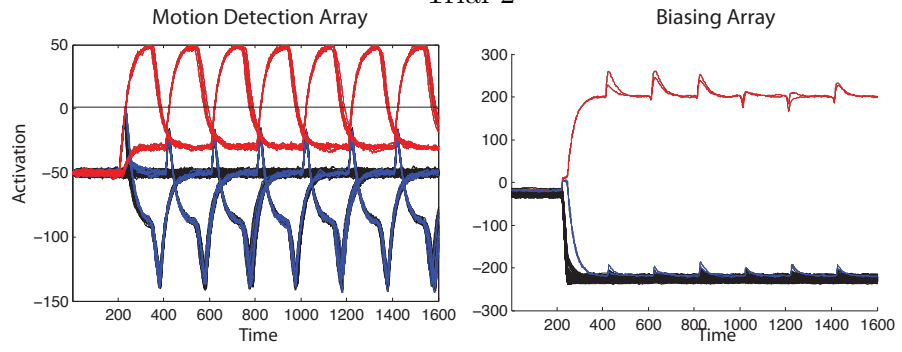
Trial 10



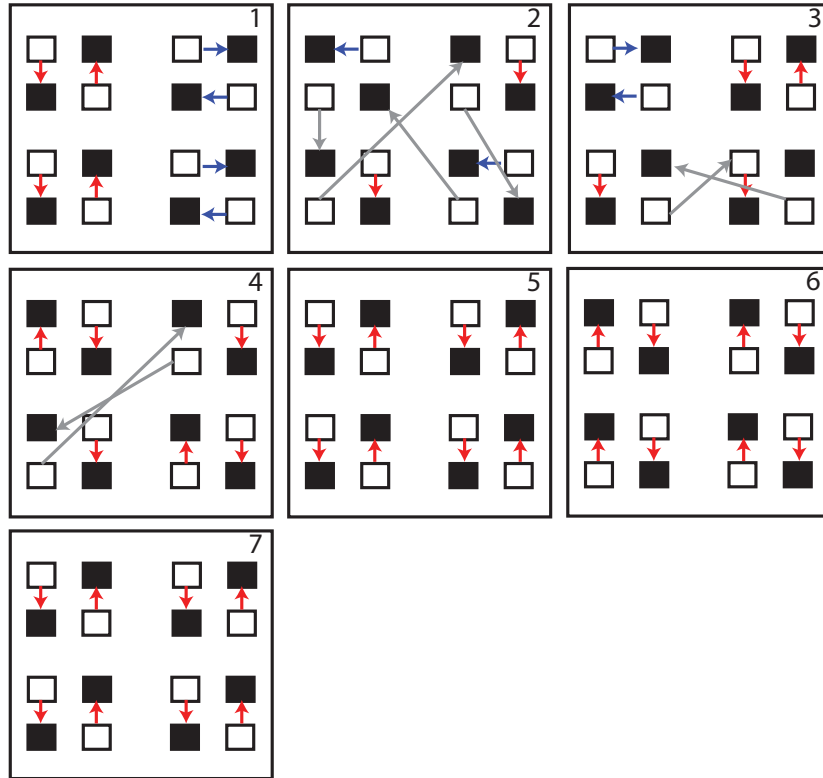
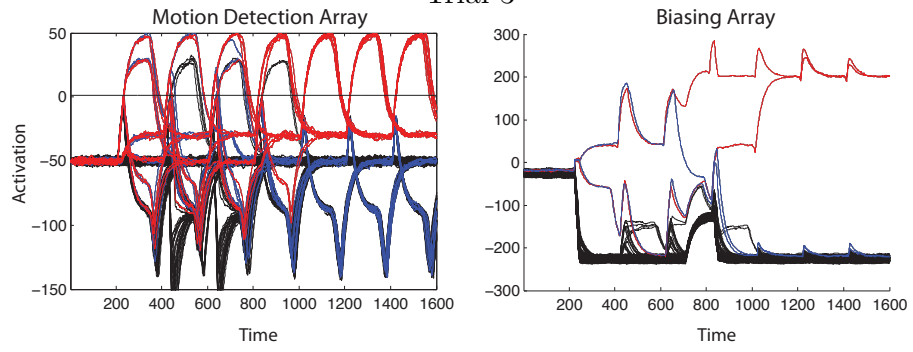
D.3 FOUR QUARTETS



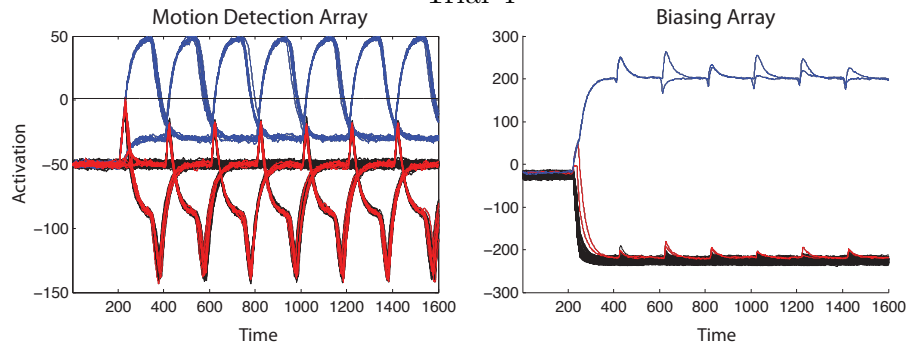
Trial 2



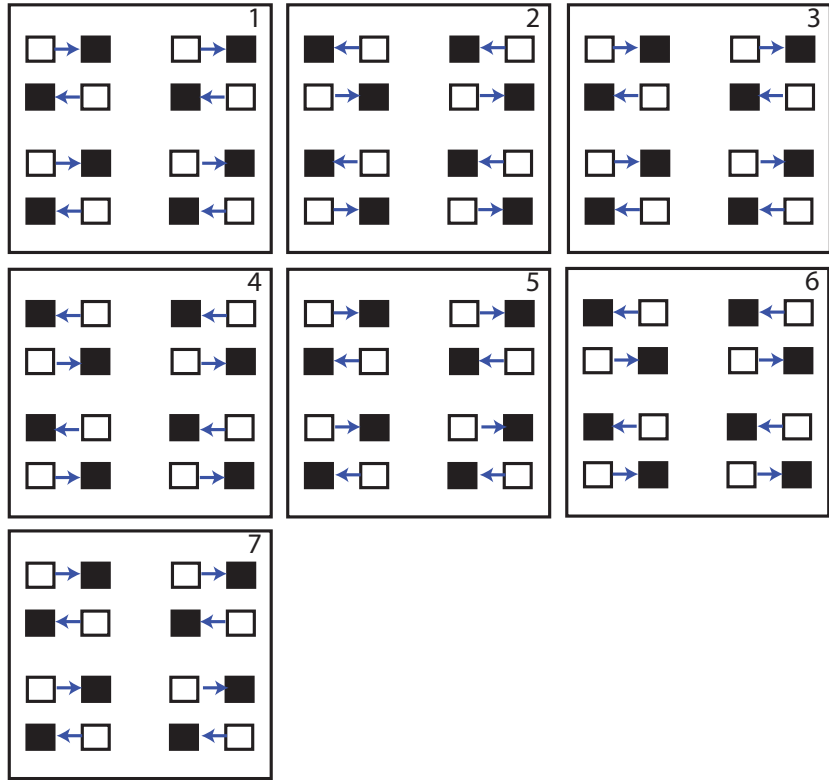
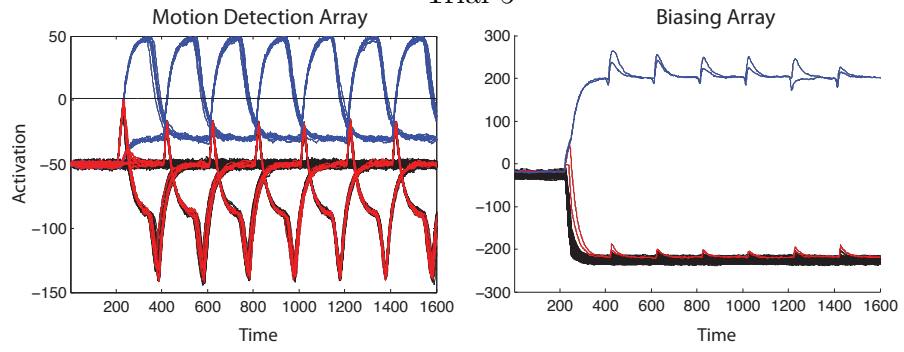
Trial 3



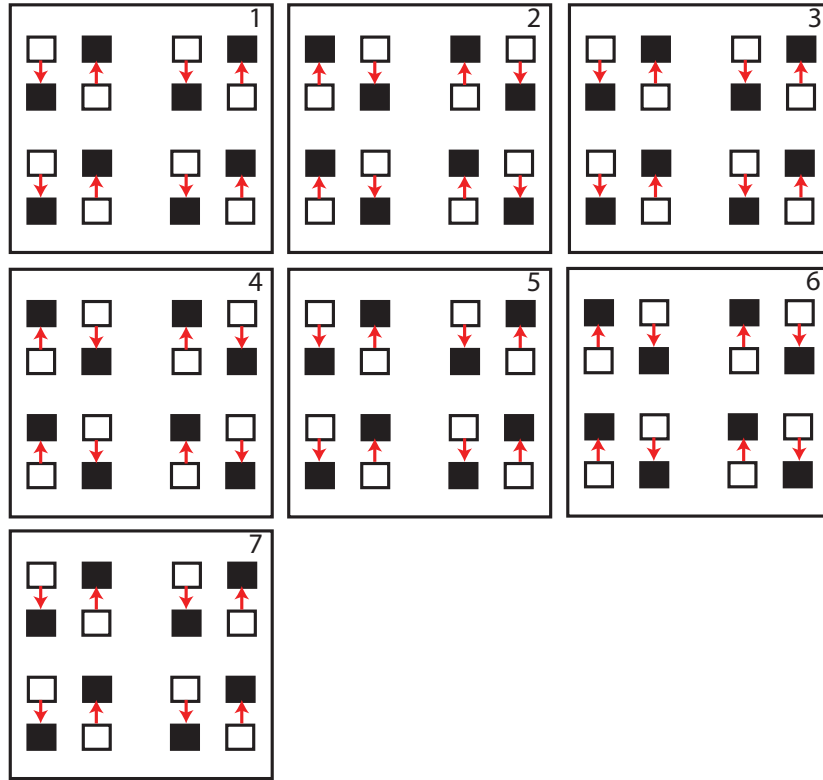
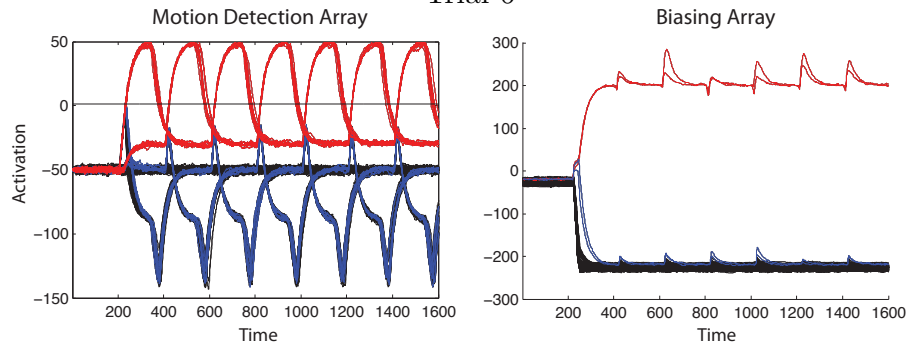
Trial 4



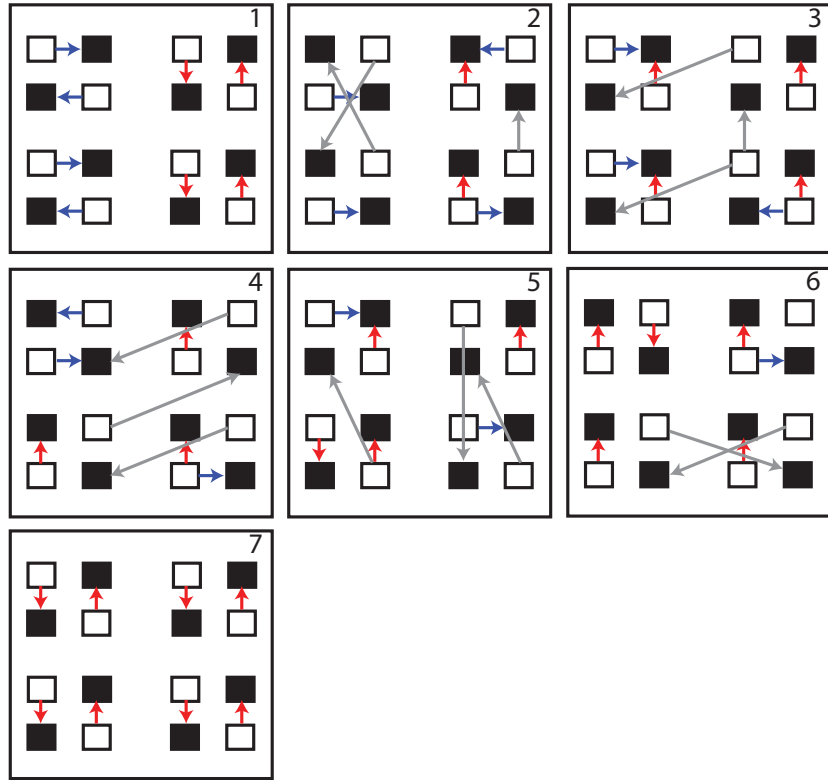
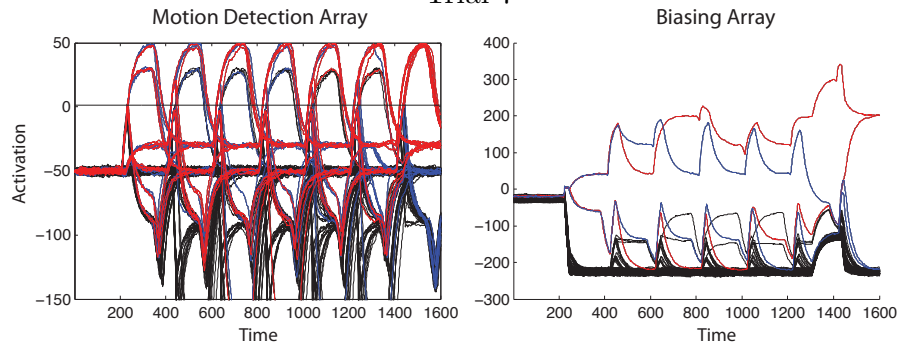
Trial 5



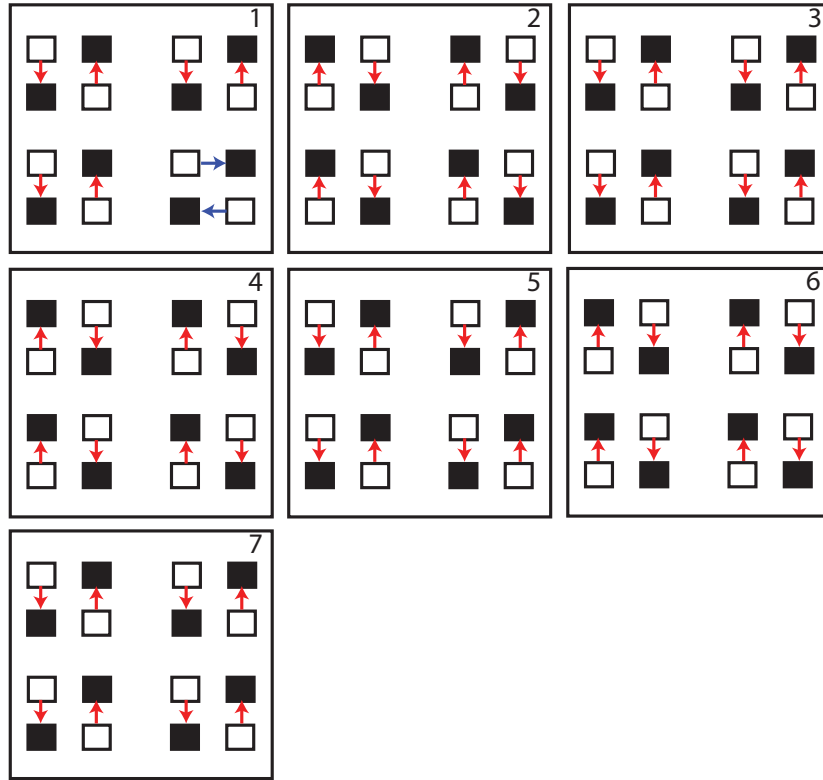
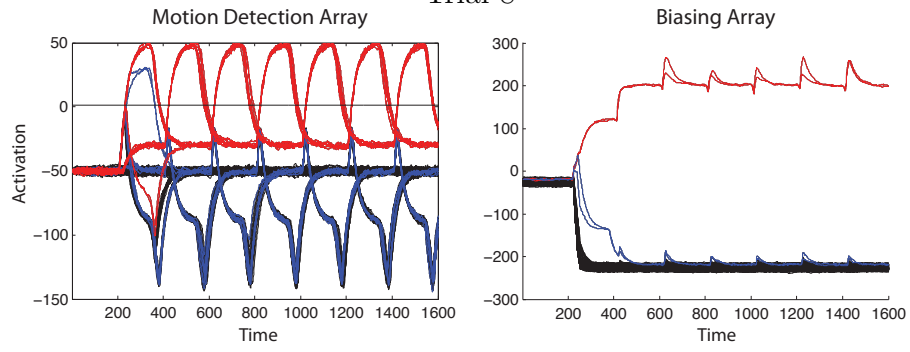
Trial 6



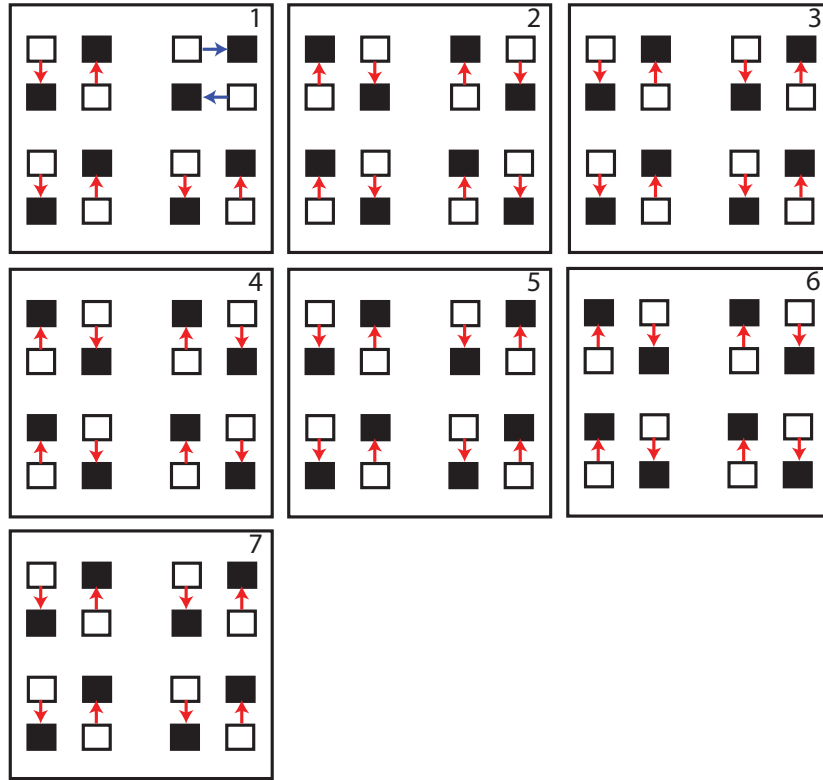
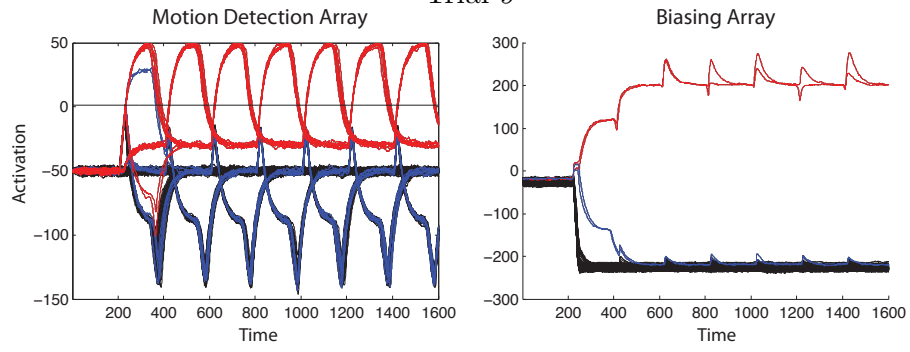
Trial 7



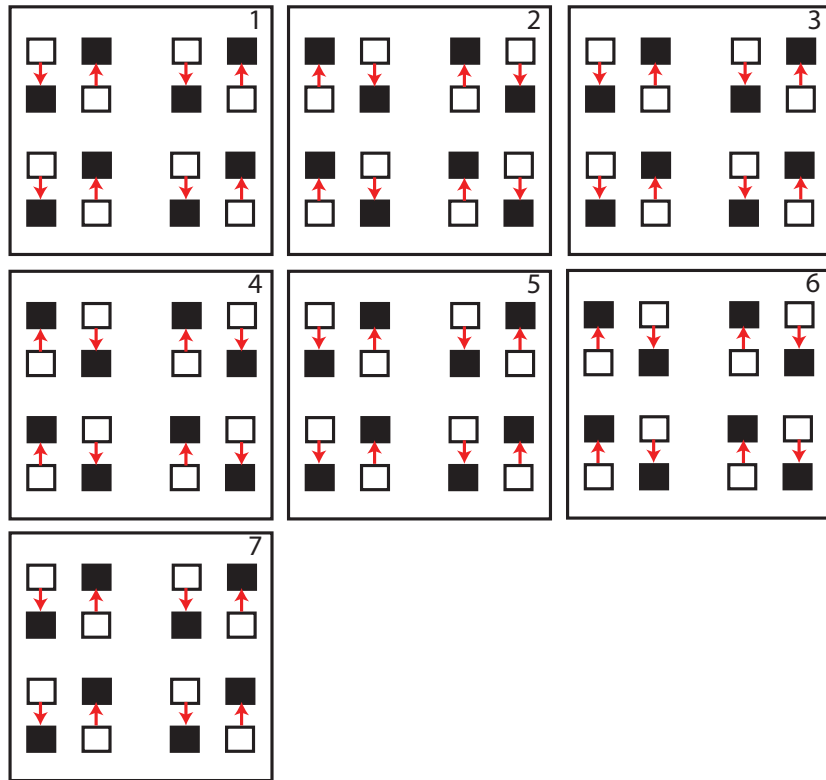
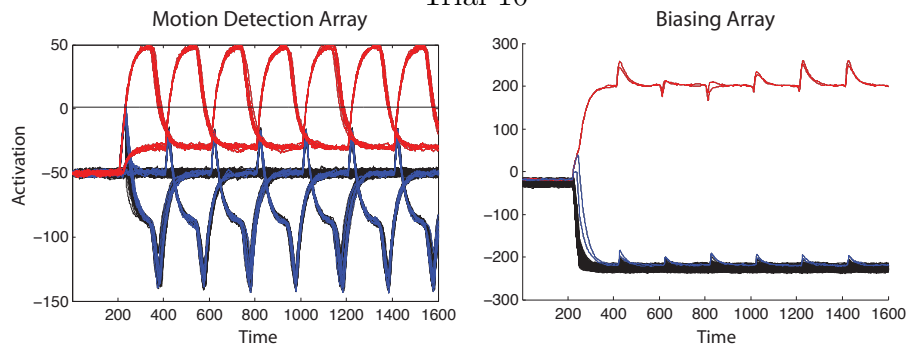
Trial 8



Trial 9



Trial 10



APPENDIX E

COPYRIGHT NOTICE FOR CHAPTER 3

3/18/2014

Rightslink Printable License

SPRINGER LICENSE TERMS AND CONDITIONS

Mar 18, 2014

This is a License Agreement between Joseph Norman ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3352030860285
License date	Mar 18, 2014
Licensed content publisher	Springer
Licensed content publication	Attention, Perception, & Psychophysics
Licensed content title	Contrasting accounts of direction and shape perception in short-range motion: Counterchange compared with motion energy detection
Licensed content author	Joseph Norman
Licensed content date	Jan 1, 2014
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	
Title of your thesis / dissertation	A Theory of The Perception of Object Motion
Expected completion date	May 2014
Estimated size(pages)	170
Total	0.00 USD

Terms and Conditions

Introduction

The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

Limited License

With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use

<https://s100.copyright.com/AppDispatchServlet>

1/4

indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: <http://www.sherpa.ac.uk/romeo/>). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis, and with a maximum of 100 extra copies in paper.

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted

You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com))

Reservation of Rights

Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:Disclaimer

You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity

You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License

This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing

This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms

Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).**OR:**

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other terms and conditions:**v1.3**

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501254350.

Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

**Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006**

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

BIBLIOGRAPHY

- [1] Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2), 284–299.
- [2] Amari, S. I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics*, 27(2), 77–87.
- [3] Anstis, S. M. (1970). Phi movement as a subtraction process. *Vision Research*, 10(12), 1411–1430.
- [4] Anstis, S., & Ramachandran, V. S. (1987). Visual inertia in apparent motion. *Vision research*, 27(5), 755–764.
- [5] Arutyunyan-Kozak, B. A., & kimyan, A. A. (1985). Organization of receptive fields of tonic and phasic neurons of the pulvinar in the cat. *Neuroscience and behavioral physiology*, 15(4), 290–295.
- [6] Azzopardi, P., & Hock, H.S. (2011). Illusory motion perception in blindsight. *Proceedings of the National Academy of Sciences*, 108, 876–881.
- [7] Bar-Yam, Y. (2013) The Limits of Phenomenology: From Behaviorism to Drug Testing and Engineering Design. *New England Complex Systems Institute (NECSI) Report*, arXiv:1308.3094 [physics.soc-ph]
- [8] Barlow, H. B., & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit’s retina. *The Journal of physiology*, 178(3), 477.
- [9] Benosman, R., Clercq, C., Lagorce, X., Ieng, S. H., & Bartolozzi, C. (2014) Event-Based Visual Flow. *Neural Networks and Learning Systems, IEEE Transactions on*, 25, 407–417. doi: 10.1109/TNNLS.2013.2273537
- [10] Berger, M., Faubel, C., Norman, J., Hock, H., & Schöner, G. (2012). The counter-change model of motion perception: an account based on dynamic field theory. In *Artificial Neural Networks and Machine Learning–ICANN 2012* (pp. 579–586). Springer Berlin Heidelberg.
- [11] Bours, R., Kroes, M., & Lankheet, M. J. (2009). Sensitivity for reverse- phi motion. *Vision Research*, 49(1).

- [12] Braddick, O. (1974). A short-range process in apparent motion. *Vision Research*, 14, 519–527.
- [13] Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., and Wiener, M. C. (2002). What geometric visual hallucinations tell us about the visual cortex. *Neural Computation*, 14(3), 473–491.
- [14] Cavanagh, P., & Mather, G. (1989). Motion: The long and short of it. *Spatial vision*, 4(2–3), 2–3.
- [15] Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli- A general basis for studying non-Fourier motion perception. *Optical Society of America, Journal, A: Optics and Image Science*, 5, 1986–2007.
- [16] Ciardiello, F., Caputo, R., Bianco, R., Damiano, V., Pomatico, G., Pepe, S., ... & Tortora, G. (1998). Cooperative inhibition of renal cancer growth by anti-epidermal growth factor receptor antibody and protein kinase A antisense oligonucleotide. *Journal of the National Cancer Institute*, 90(14), 1087–1998.
- [17] Cloutier, J. F., & Veillette, A. (1999). Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. *The Journal of experimental medicine*, 189(1), 111–121.
- [18] Dawson, M. R. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological review*, 98(4), 569.
- [19] Doshier, B. A., Landy, M. S., & Sperling, G. (1989). Kinetic depth effect and optic flow—I. 3D shape from Fourier motion. *Vision Research*, 29(12), 1789–1813.
- [20] Doursat, R., Sayama, H., & Michel, O. (2012). *Morphogenetic Engineering: Toward Programmable Complex Systems*. Springer.
- [21] Edwards, M., & Badcock, D. R. (1994). Global motion perception: Interaction of the ON and OFF pathways. *Vision Research*.
- [22] Eichner, H., Joesch, M., Schnell, B., Reiff, D. F., & Borst, A. (2011). Internal structure of the fly elementary motion detector. *Neuron*, 70(6), 1155–1164. doi:10.1016/j.neuron.2011.03.028
- [23] Fuster, J. M. (2003). *Cortex and mind: Unifying cognition*. Oxford university press.
- [24] Gabbiani, F., Krapp, H. G., Koch, C., & Laurent, G. (2002). Multiplicative computation in a visual neuron sensitive to looming. *Nature*, 420(6913), 320–324.

- [25] Gibson, J. J. (1986). *The ecological approach to visual perception*. Psychology Press.
- [26] Giese, M. A. (1998). Dynamic neural field theory for motion perception. Kluwer Academic Publishers.
- [27] Gilroy, L. A., & Hock, H. S. (2004). Multiplicative nonlinearity in the perception of apparent motion. *Vision research*, 44(17), 2001-2007.
- [28] Gilroy, L. A., & Hock, H. S. (2009). Simultaneity and sequence in the perception of apparent motion. *Attention, Perception & Psychophysics*, 71(7), 1563–1575. doi:10.3758/APP.71.7.1563
- [29] Heeger, D. J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology*, 70(5), 1885–1898.
- [30] Hock, H. S., Gilroy, L., & Harnett, G. (2002). Counter-changing luminance: A non-Fourier, nonattentional basis for the perception of single-element apparent motion. *Journal of Experimental Psychology Human Perception and Performance*, 28(1), 93.
- [31] Hock, H. S., Kelso, J. S., & Schöner, G. (1993). Bistability and hysteresis in the organization of apparent motion patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 19(1), 63.
- [32] Hock, H. S., & Nichols, D. F. (2010). The line motion illusion: The detection of counterchanging edge and surface contrast. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 781.
- [33] Hock, H. S., & Nichols, D. F. (2013). The perception of object versus objectless motion. *Attention, Perception, & Psychophysics*, 75(4), 726–737.
- [34] Hock, H. S., Schöner, G., Brownlow, S., & Taler, D. (2011). The temporal dynamics of global-to-local feedback in the formation of hierarchical motion patterns: psychophysics and computational simulations. *Attention, Perception, & Psychophysics*, 73(4), 1171–1194.
- [35] Hock, H. S., Schöner, G., & Giese, M. (2003). The dynamical foundations of motion pattern formation: Stability, selective adaptation, and perceptual continuity. *Perception & psychophysics*, 65(3), 429-457.
- [36] Hock, H., Schöner, G., & Gilroy, L. (2009). A counterchange mechanism for the perception of motion. *Acta Psychologica*, 132(1), 1–21.
- [37] Hock, H. S., Schöner, G., & Voss, A. (1997). The influence of adaptation and stochastic fluctuations on spontaneous perceptual changes for bistable stimuli. *Perception & Psychophysics*, 59(4), 509-522.

- [38] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8), 2554–2558.
- [39] Hopfield, J. J., & Tank, D. W. (1985). Neural computation of decisions in optimization problems. *Biological cybernetics*, 52(3), 141–152.
- [40] James, W. (1890). *The principles of psychology*. Digireads. com Publishing.
- [41] Johansson, G. (1950). *Configurations in Event Perception*. Almqvist and Wiksell, Boktryckeri AB, Uppsala, Sweden.
- [42] Kim, N., Growney, R., & Turvey, M. T. (1996). Optical flow not retinal flow is the basis of wayfinding by foot. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1279–1288. doi:10.1037/0096-1523.22.5.1279
- [43] Kohler, A., Haddad, L., Singer, W., & Muckli, L. (2008). Deciding what to see: The role of intention and attention in the perception of apparent motion. *Vision Research*, 48(8), 1096–1106.
- [44] Lu, Z. L., & Sperling, G. (2001). Three-systems theory of human visual motion perception: Review and update. *Journal of the Optical Society of America A, Optics, image science, and vision*, 18(9), 2331–2370.
- [45] Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society B: Biological Sciences*, 211(1183), 151–180. doi:10.1098/rspb.1981.0001
- [46] Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, 355(6358), 344–346. doi:10.1038/355344a0
- [47] Murakami, G., Watabe, T., Takaoka, K., Miyazono, K., & Imamura, T. (2003). Cooperative inhibition of bone morphogenetic protein signaling by Smurf1 and inhibitory Smads. *Molecular biology of the cell*, 14(7), 2809–2817.
- [48] Norman, D. O., and Kuras, M. L. (2006). Engineering complex systems. In *Complex Engineered Systems* (pp. 206–245). Springer Berlin Heidelberg.
- [49] Norman, J., Hock, H., & Schöner, G. (2014). Contrasting accounts of direction and shape perception in short-range motion: Counterchange compared with motion energy detection. *Attention, Perception, & Psychophysics*, 1–21. doi:10.3758/s13414-014-0650-2
- [50] Pelah, A., Barbur, J., Thurrell, A., & Hock, H.S. (2014). The Coupling of Vision with Locomotion in Cortical Blindness. *Vision Research*, (in revision).

- [51] Read, J. (2002). A Bayesian model of stereopsis depth and motion direction discrimination. *Biological Cybernetics*.
- [52] Ramachandran, V. S., & Anstis, S. M. (1983). Perceptual organization in moving patterns. *Nature*, 304(5926), 529–531. doi: 10.1038/304529a0
- [53] Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), *Sensory Communication* (pp. 303–317). Cambridge: MIT Press.
- [54] Sandamirskaya, Y., Zibner, S. K., Schneegans, S., & Schöner, G. (2013). Using dynamic field theory to extend the embodiment stance toward higher cognition. *New Ideas in Psychology*, 31(3), 322–339.
- [55] Sanders, J., & Kandrot, E. (2010). CUDA by Example. *An Introduction to General-Purpose GPU Programming*. Addison-Wesley Professional.
- [56] Sato, T. (1989). Reversed apparent motion with random dot patterns. *Vision Research*, 29(12), 1749–1758.
- [57] Schöner, G., (2008) Dynamical Systems Approaches to Cognition. In Ron Sun, (ed.), *The Cambridge Handbook of Computational Psychology* (pp. 101–126). Cambridge University Press.
- [58] Schütz, A. C., Lossin, F., & Kerzel, D. (2013). Temporal stimulus properties that attract gaze to the periphery and repel gaze from fixation. *Journal of vision*, 13(5), 6.
- [59] Seifert, M.S., & Hock, H.S. (2014). The Independent Detection of Motion Energy and Counterchange: Flexibility in Motion Detection. *Vision Research*, (in press).
- [60] Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research*, 38(5), 743–761.
- [61] Sperling, G., & Lu, Z.-L. (1998). A systems analysis of visual motion perception. *High-level motion*, 153–183.
- [62] Stevens, M., & Merilaita, S. (2009). Animal camouflage: Current issues and new perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1516), 423–427. doi:10.1098/ rstb.2008.0217
- [63] Taleb, N. N. (2012). *Antifragile: things that gain from disorder*. Random House LLC.
- [64] Ullman, S. (1979). *The interpretation of visual motion*. Massachusetts Inst of Technology Pr.

- [65] van Santen, J. P., & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A, Optics and image science*, 1(5), 451–473.
- [66] van Santen, J. P., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, Optics and image science*, 2(2), 300–321.
- [67] Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems*, 5(4), 187–196.
- [68] Vukicevic, M. & Fitzmaurice, K. (2008). Butterflies and black lacy patterns: the prevalence and characteristics of Charles Bonnet hallucinations in an Australian population. *Clinical and Experimental Ophthalmology*. 36, 659–65.
- [69] Watson, A. B., & Ahumada Jr, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, Optics and image science*, 2(2), 322–341.
- [70] Wehrhahn, C., & Rapf, D. (1992). ON- and OFF-pathways form separate neural substrates for motion perception: Psychophysical evidence. *The Journal of Neuroscience*, 1–4.
- [71] Wertheimer, M. (1912). Experimental studies of the perception of movement (Experimentelle Studien ber das Sehen von Bewegung). *Zeitschrift fr Psychologie under Physiologie der Sinnesorgane*, 61, 161–265.
- [72] Yuille, A. L., & Grzywacz, N. M. (1998). A theoretical framework for visual motion. In T. Watanabe (Ed.), *High-level motion processing* (pp. 187–211). Cambridge: The MIT Press.