COMPUTATIONAL ROLE OF DISINHIBITION IN BRAIN FUNCTION

A Dissertation

by

YINGWEI YU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2006

Major Subject: Computer Science

COMPUTATIONAL ROLE OF DISINHIBITION IN BRAIN FUNCTION

A Dissertation

by

YINGWEI YU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Yoonsuck Choe
Committee Members,	Ricardo Gutierrez-Osuna
	Thomas Ioerger
	Takashi Yamauchi
Head of Department,	Valerie E. Taylor

August 2006

Major Subject: Computer Science

ABSTRACT

Computational Role of Disinhibition
in Brain Function. (August 2006)
Yingwei Yu, B.E., Beihang University, China
Chair of Advisory Committee: Dr. Yoonsuck Choe

Neurons are connected to form functional networks in the brain. When neurons are combined in sequence, nontrivial effects arise. One example is disinhibition; that is, inhibition to another inhibitory factor. Disinhibition may be serving an important purpose because a large number of local circuits in the brain contain disinhibitory connections. However, their exact functional role is not well understood.

The objective of this dissertation is to analyze the computational role of disinhibition in brain function, especially in visual perception and attentional control. My approach is to propose computational models of disinhibition and then map the model to the local circuits in the brain to explain psychological phenomena. Several computational models are proposed in this dissertation to account for disinhibition. (1) A static inverse difference of Gaussian filter (IDoG) is derived to account explicitly for the spatial effects of disinhibition. IDoG can explain a number of complex brightness-contrast illusions, such as the periphery problem in the Hermann grid and the White's effect. The IDoG model can also be used to explain orientation perception of multiple lines as in the modified version of Poggendorff illusion. (2) A spatio-temporal model (IDoGS) in early vision is derived and it successfully explains the scintillating grid illusion, which is a stationary display giving rise to a striking, dynamic, scintillating effect. (3) An interconnected Cohen-Grossberg neural network model (iCGNN) is proposed to address the dynamics of disinhibitory neural networks

with a layered structure. I derive a set of sufficient conditions for such an interconnected system to reach asymptotic stability. (4) A computational model combining recurrent and feed-forward disinhibition is designed to account for input-modulation in temporal selective attention.

The main contribution of this research is that it developed a unified framework of disinhibition to model several different kinds of neural circuits to account for various perceptual and attentional phenomena. Investigating the role of disinhibition in the brain can provide us with a deeper understanding of how the brain can give rise to intelligent and complex functions.

To my grandmother *Herling Chen* (1923-1996).

ACKNOWLEDGMENTS

First of all, I would like to sincerely thank my advisor, Dr. Yoonsuck Choe, for introducing me to the research of neural networks and computational neuroscience, and providing me support and guidance. He spent a tremendous amount of time and effort in helping me write and present my dissertation and all of my papers throughout these years. Without his help and advise, I think I could not have finished my Ph.D. research. I would also like to thank my committee members, Dr. Ricardo Gutierrez-Osuna, Dr. Takashi Yamauchi, and Dr. Thomas Ioerger, for their guidance, and Dr. Van Rullen for his valuable comments on my research.

I would like to thank my NIL lab member, Heejin Lim, for her feedback and comments on my presentations and papers. I would also like to thank other NIL lab members and friends, Sejong Oh, S. Kumar Bhamidipati, Ji Ryang Chung, Jae Rock Kwon, Feng Liang, Nan Zhang, and Xiafeng Li, who have made these years of research a pleasant experience.

Finally, I would like to give my special thanks to my beloved wife, Rui Zhang, for her love and support. I would also like to thank my grandfather, Darui Yu, my dad, Pei Yu, my mom, Qingyun Yang, and my brother, Yingnan Yu, for their absolute confidence in me and their care and support.

This work was supported in part by Texas Higher Education Coordinating Board ATP program grant #000512-0217-2001.

TABLE OF CONTENTS

CHAPTER Page

I INTRODUCTION 1

A. Motivation 1
1. The periphery problem in the Hermann grid 2
2. The scintillating grid illusion 3
B. The main research question: Role of disinhibition in the brain 6
C. Approach 9
D. Outline 9

II BACKGROUND 11

A. Various disinhibition phenomena at different levels of brain organization 11
B. Hartline-Ratliff equation 13
C. Fourier model 13
D. Summary 16

III COMPUTATIONAL MODELS OF DISINHIBITION 17

A. Spatial disinhibitory filter: The IDoG model 17
B. Spatio-temporal disinhibitory filter: The IDoGS model . . 19
C. Summary 24

IV ROLE OF SPATIAL DISINHIBITION IN STATIC BRIGHTNESS-CONTRAST PERCEPTION 27

A. The Hermann grid illusion 28
B. The White's effect 31
C. The Mach band 32
D. Summary 32

V ROLE OF SPATIO-TEMPORAL DISINHIBITION IN DYNAMIC BRIGHTNESS-CONTRAST PERCEPTION 36

A. Scintillating grid illusion 36
B. Methods 39
C. Experiments and results 40

CHAPTER Page

1. Experiment 1: Perceived brightness as a function of receptive field size . . . 40
2. Experiment 2: Perceived brightness as a function of time . . . 41
3. Experiment 3: Strength of scintillation as a function of luminance . . . 41
4. Experiment 4: Strength of scintillation as a function of motion speed and presentation duration . . . 44
D. Discussion . . . 48
E. Summary . . . 51

VI ROLE OF DISINHIBITION IN ORIENTATION PERCEPTION 52

A. Poggendorff illusion . . . 52
B. Methods . . . 54
C. Model . . . 55
1. Activation profile of orientation columns . . . 55
2. Column level inhibition and disinhibition . . . 58
3. Applying disinhibition to orientation cells . . . 58
D. Results . . . 60
1. Experiment 1: Angle expansion without additional context . . . 60
2. Experiment 2: modified Poggendorff illusion . . . 61
E. Discussion . . . 67
F. Summary . . . 69

VII ROLE OF DISINHIBITION IN ATTENTIONAL CONTROL . 71

A. Disinhibition in the thalamocortical circuit . . . 71
1. Role of disinhibition in the thalamus . . . 72
2. Biologically accurate model of the thalamocortical circuit . . . 73
B. Selective attention over time . . . 74
C. Model . . . 80
1. Model architecture . . . 80
2. Temporal control profile . . . 83
D. Experiments and results . . . 86
1. Experiment 1: Using irrelevant control profile for distractor . . . 86
2. Experiment 2: Using relevant control profile for distractor 88

CHAPTER Page

E. Discussion 91
F. Summary 93

VIII ROLE OF DISINHIBITION FROM A SYSTEM PERSPECTIVE 94

A. Stability analysis of thalamocortical circuit 94
1. Interconnected Cohen-Grossberg neural networks . . . 95
2. Dynamics of interconnected CGNN 97
3. Assumptions 98
4. Existence of equilibrium point 99
5. Lyapunov function for interconnected systems in general 100
6. Asymptotic stability of interconnected CGNN 101
7. Examples and computer simulations 106
B. Controllability: Control resolution 107
C. Computability: The logic of control 112
D. Discussion 114
E. Summary 116

IX DISCUSSION AND CONCLUSION 118

A. Summary 118
B. Discussion 119
1. Limitations of the approach 120
2. Predictions 121
3. Contributions 124
C. Future directions 125
D. Conclusion 126

REFERENCES 127

VITA 139

LIST OF TABLES

TABLE Page

I Parameters 106

II Computational Roles of Disinhibition in Brain Function . 120

LIST OF FIGURES

FIGURE		Page
1	The Hermann grid illusion	2
2	The Hermann grid and lateral inhibition	3
3	The periphery problem in the Hermann grid	4
4	The scintillating grid illusion	5
5	Type I disinhibition	6
6	Type II disinhibition	7
7	Type III disinhibition	8
8	Outline	10
9	A possible configuration of lateral inhibition between orientation detectors	12
10	Lateral inhibition in Limulus optical cells (Redrawn from [1])	14
11	Sinusoidal input	15
12	An inverse DoG filter (IDOG)	19
13	The dynamics of self-inhibition	20
14	The impulse response functions $f(t)$ and $g(t)$	23
15	Self-inhibition rate	24
16	Inversed DoG with self-inhibition filter (IDoGS) at various time points	25
17	The Hermann grid illusion under DoG filter	29
18	The Hermann grid illusion under IDoG filter	30

FIGURE Page

19 The White's effect under DoG filter 33

20 The White's effect and prediction under IDoG filter 34

21 The Mach band under DoG and IDoG 35

22 The scintillating grid illusion and its polar variation 37

23 A variation without the scintillating effect 38

24 Response under various receptive field sizes 42

25 Response at various time points 43

26 Strength of scintillation under various luminance conditions 45

27 Strength of scintillation under varying speed and presentation duration 47

28 The Poggendorff Illusion 53

29 Activation profile 57

30 A possible configuration of lateral inhibition between orientation detectors 58

31 The variations of perceived angle between two intersecting lines . . . 60

32 Initial orientation column activations (green) and final responses of orientation columns (red) after disinhibition 63

33 Perceived angle in a modified Poggendorff illusion 64

34 Perceived angles under various values of η and σ 65

35 The endpoint effect in the Poggendorff illusion 70

36 A schematic drawing of thalamus connectivity 72

37 Input vs. no-input condition 75

38 Strong vs. weak input condition 75

39 A schematic drawing of attentional control 77

FIGURE Page

40 Human data and model result of SOA experiment 79

41 An overview of the input-modulation model 81

42 Module II: The stimulus competition model 82

43 Module I: A solution for temporal learning 83

44 Temporal control profiles 85

45 Result of experiment 1: Human data and model result of SOA experiment 87

46 Result of experiment 2: Using relevant control profile for distractor . 90

47 An illustration of multiple CGNNs with column-wise connection . . . 96

48 Simulation of thalamocortical circuits (two loops) showing convergence behavior 107

49 Disinhibition in basal ganglia 109

50 The abstract model of basal ganglia 111

51 Control accuracy of the basal ganglia model with feedforward disinhibition 112

52 Signal routing by disinhibition 114

53 Experiment 1: Signal is allowed from the sensory input to the cortex cells 115

54 Experiment 2: Signal is blocked from the sensory input to the cortex cells 116

55 Main results in studying the computational role of disinhibition . . . 119

CHAPTER I

INTRODUCTION

Neurons are connected to form functional networks, and networks are connected to form a system, the brain. The interaction between two connected neurons can be categorized as either excitatory or inhibitory. Through those interactions in space and time, neurons are tightly coupled into a whole system, and exhibit complex behavior. A repeatedly observed pattern of local circuits is disinhibition. Disinhibition is basically inhibition of another inhibitory factor. There is a large number of disinhibitory circuits in the brain, such as in the retina, the thalamus, the cortex, the basal ganglia, and the cerebellum. What is the computational role of disinhibition? What is the function of disinhibition in visual perception? What is the function of disinhibition in attention? These questions are the main motivations for my dissertation research.

A. Motivation

Visual perception is an important function of the brain. Human visual perception may not always reflect the real world, and in such a case of misinterpretation, we say we have visual illusion. Visual illusions are important phenomena because of their potential to shed light on the underlying functional organization of the visual system. My research is motivated by two interesting brightness-contrast (B-C) visual illusions. One is the Hermann grid, and the other is a variation of the Hermann grid – the scintillating grid. I will introduce the two illusions in the following.

The journal model is *IEEE Transactions on Neural Networks*.

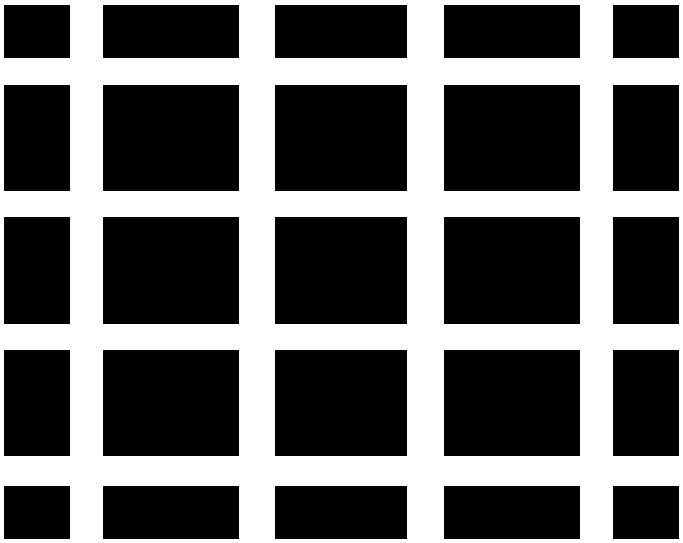


Fig. 1. **The Hermann grid illusion.** Hermann grid contains some tiled black blocks and white streets. The intersections look darker than the streets.

1. The periphery problem in the Hermann grid

Hermann grid consists of tiled black blocks and white streets as shown in figure 1. Illusory dark spots can be perceived at the intersection of the white street. The illusory dark spots are due to lateral inhibition in the retina and in the lateral geniculate nucleus (LGN) [2]. Lateral inhibition is the effect observed in the receptive field where the surrounding area inhibits the central area. The lateral inhibition process in the Hermann grid is demonstrated in figure 2. Due to the fact that the neuron at the intersection receives more inhibition than those in the streets, the intersection appears much darker than the streets.

However, lateral inhibition alone cannot account for all subtleties in the visual illusion. For example, in the Hermann grid illusion, although the illusory spots are explained pretty well by feedforward lateral inhibition, it cannot explain why the periphery appears brighter than the illusory dark spots. The purely lateral inhibitory

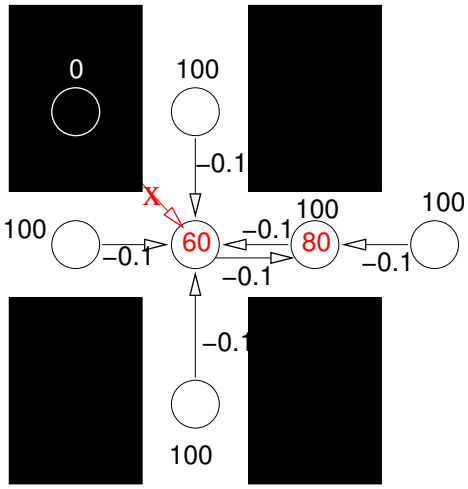


Fig. 2. **The Hermann grid and lateral inhibition.** Let us assume the neurons in the white area receive 100 unit as the initial input, and neurons in the dark block receive 0 input. Let the feedforward inhibition rate be 0.1. As a result, the response of the neuron in the intersection is 60 ($= 100 - 0.1 \times 100 \times 4$), which is significantly lower than the response of the neuron in the street, 80 ($= 100 - 0.1 \times 100 \times 2$). This figure demonstrates why there are illusory dark spots in the center of the intersections.

mechanism fails to address our perceived experience of the periphery (figure 3). The reason for this failure is that the center of lateral inhibition in the peripheral area receives inhibition from all the surrounding directions, resulting in a weaker response than the intersections in the grid which only receive inhibition from four directions. The question that arose from this observation was whether a more elaborate mechanism exists in early visual processing, besides lateral inhibition.

2. The scintillating grid illusion

Another interesting visual illusion is the scintillating grid, which is a variation of the Hermann grid. The scintillating grid consists of bright discs superimposed on intersections of orthogonal gray bars on a dark background (figure 4) [3]. In this illusion,

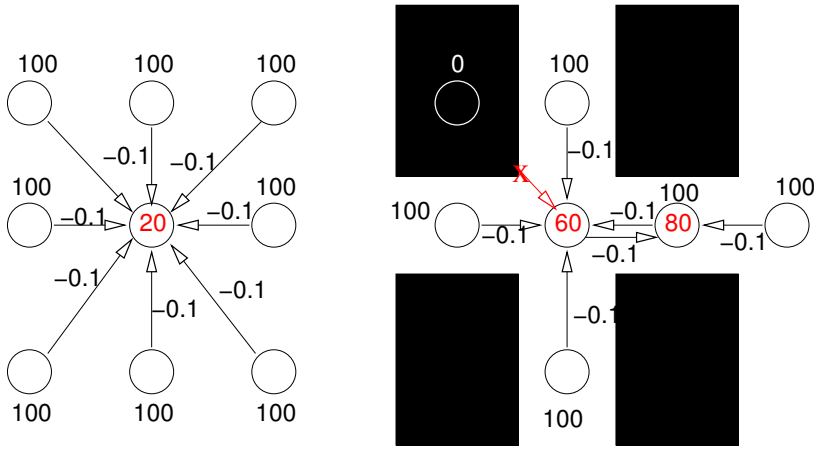


Fig. 3. **The periphery problem in the Hermann grid.** The center of lateral inhibition in the peripheral area (to the left) receives inhibition from all the surrounding directions which results in a weaker response than the intersections in the grid (to the right) which only receive inhibition from four directions. The neurons on the right are configured the same as in figure 2. The neuron in the center of the left side marked "20" receives inhibition from all the surrounding directions, and 20 ($= 100 - 0.1 \times 100 \times 8$) is its final response. According to the feedforward lateral inhibition model, the neurons in the periphery have a lower response than those in the street (the neuron marked with "80") and those at the intersection (the neuron marked with "60"). This result is contrary to our perception, because we perceive the periphery as brighter.

illusory dark spots are perceived as scintillating within the white discs. Several important spatiotemporal properties of the illusion have been discovered and reported in recent years. For example, the discs that are closer to a fixation show less scintillation [3], and the illusion is greatly reduced or even abolished both with steady fixation and by reducing the contrast between the constituent grid elements [3].

What kind of neural process could be responsible for such a dynamic illusion? The scintillating grid can be seen as a variation of the Hermann grid illusion where the brightness level of the intersecting bars is reduced. The illusory dark spots in

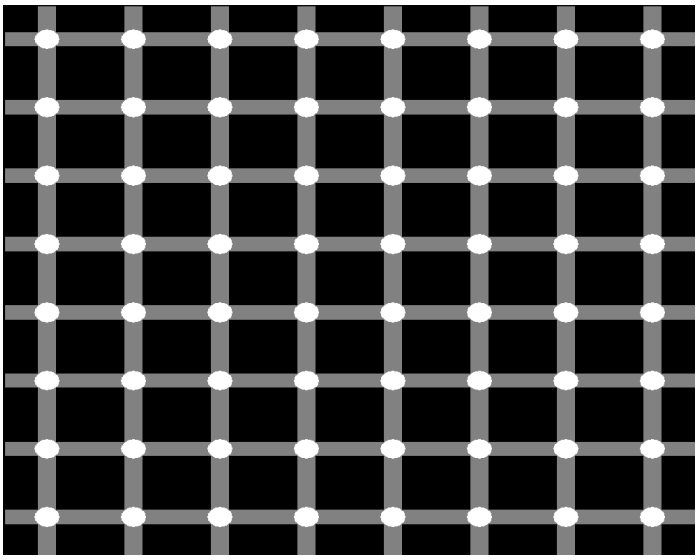


Fig. 4. **The scintillating grid illusion.** The scintillating grid illusion consists of bright discs superimposed on intersections of orthogonal gray bars on a dark background. In this illusion, illusory dark spots are perceived as scintillating within the white discs. See text for more details about this illusion (redrawn from [3]).

Hermann grid can be explained by feedforward lateral inhibition, commonly modeled with Difference of Gaussian (DoG) filters [2]. Thus, DoG filters may seem to be a plausible mechanism contributing to the scintillating grid illusion. However, DoG filters are not exactly fit to explain the complexities of the scintillating grid illusion because of the following reasons. (1) The DoG model cannot account for the change in the strength of scintillation over different brightness and contrast conditions, as shown in the experiments of Schrauf et al. [3]. (2) Furthermore, DoG cannot explain the basic scintillation effect which has a temporal dimension to it. Thus, the feedforward lateral mechanism represented by DoG fails to fully explain the scintillating effect.

The two above visual illusions (one in the previous section) suggests that there are other inhibitory mechanisms which are not captured by the lateral inhibition model.

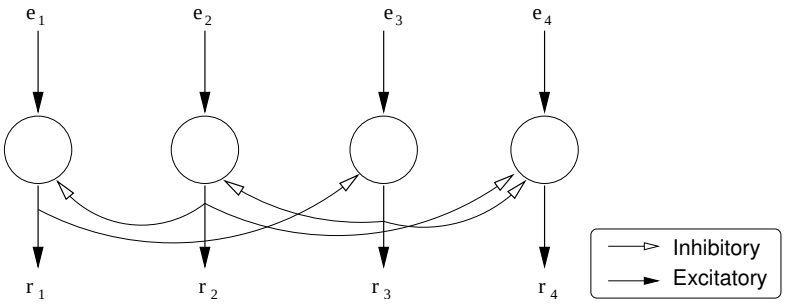


Fig. 5. **Type I disinhibition.** The input to node i is e_i, and the output through its axon is r_i. The link between each pair of nodes is inhibitory, and all the axons project back to other neurons. This type of disinhibition is recurrent. Lines with filled arrows represent excitatory synapses, while those with unfilled arrows inhibitory synapses.

One possible inhibition mechanism that can give rise to such effects is disinhibition. In the next section, I will introduce the concept of disinhibition and its categories. The explanation to the above two visual illusions will be fully discussed in Chapter IV and V, respectively.

B. The main research question: Role of disinhibition in the brain

Generally, disinhibition can be defined as inhibition of inhibitors. Based on the connectivity, in my observation, disinhibition can be categorized into recurrent negative feedback connection (I will call this type I disinhibition hence forth), negative feedforward connections (type II disinhibition), or a hybrid structure containing both of these (type III disinhibition).

Type I disinhibition refers to the structure where neurons receive recurrent inhibitory feedbacks from each other. Figure 5 gives an illustration of this kind of disinhibition. Here neurons are mutually cross inhibited. This type of disinhibition is found in the early visual pathway, such as in the optical cells in the Limulus [4; 1; 5].

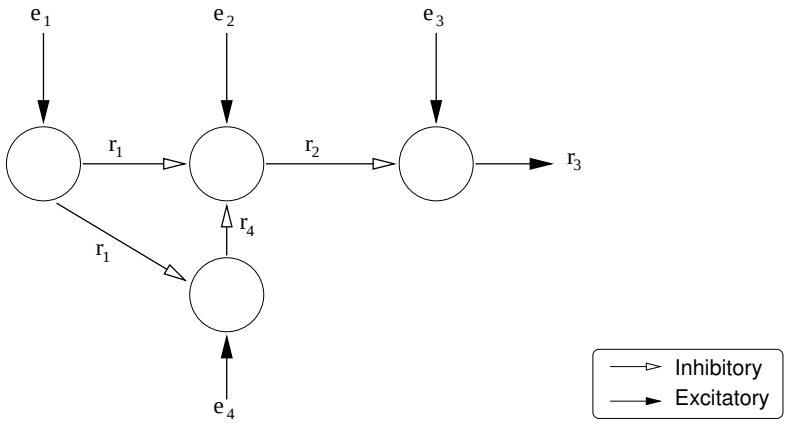


Fig. 6. **Type II disinhibition.** The input to node i is e_i, and the output of its axon is r_i. The link between each pair of nodes is inhibitory, and all the axons project in the same direction, i.e., there is no cycle in the graph. This type of disinhibition is feedforward. Lines with filled arrows represent excitatory synapses, while those with unfilled arrows inhibitory synapses.

The second type, feedforward structure (type II disinhibition), is illustrated in figure 6. In this structure, no pathway can come back to a neuron visited earlier, i.e. no cycle can be found. This type of feedforward disinhibition is found in the basal ganglia [6].

The third type is hybrid structure (type III disinhibition) which combines the recurrent and the feedforward topology. A sample of such a connection schema is shown in figure 7. Neurons numbered 1 to 4 have recurrent connections, while neuron 5 and 6 feedforward connections. An example of such a connection type in the brain is the thalamocortical circuit. It contains recurrent network layer (thalamic reticular nucleus neurons), and feedforward layers (relay cells and cortical neurons) [7; 8; 9].

Disinhibition can be widely found in the brain. For example, in the retina, bipolar cells are interconnected by amacrine cells, through which bipolar cells are mutually inhibited (type I disinhibition). In the thalamus, the thalamic reticular neurons are

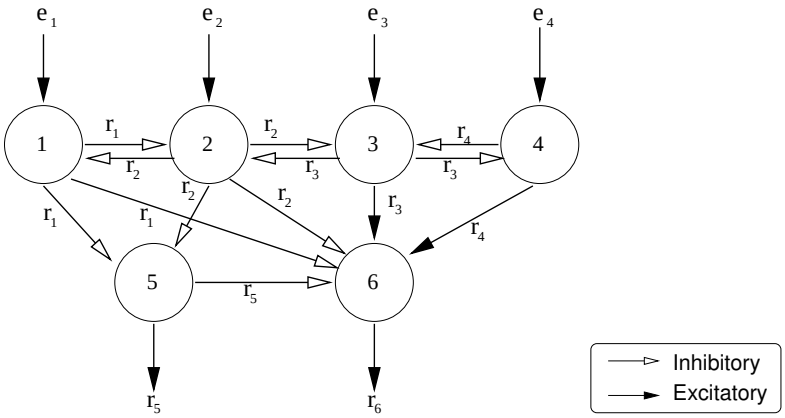


Fig. 7. **Type III disinhibition.** Neurons numbered 1 to 4 have recurrent connections, while neuron 5 and 6 feedforward connections. The input to node i is e_i, and the output of its axon is r_i.

mutually connected with each others with inhibitory synapses (type I disinhibition). Moreover, the inhibition of the reticular cell to the relay cell can be disinhibited by a second reticular cell, thus demonstrating type III disinhibition. In the cerebellum, feedforward disinhibition is widely observed, e.g. a series of inhibitory pathway (type II disinhibition) including basket cells, Purkinje cells, interpositus nucleus, and inferior olive [10]. The inhibitory pathways in basal ganglia also demonstrate type II disinhibition, from striatum to GPe (external segment of the globus pallidus) and STN (subthalamic nucleus), etc. (See Chapter 10, [11].) Finally, in the visual cortex, disinhibition between simple cells and complex cells are found to significantly affect angle perception. Here only a few are listed, but there are many more disinhibitory circuits in the brain. Thus, studying the role of disinhibition in the brain can help us gain insights into how the complex behavior (e.g., visual illusions) of various brain organizations can be realized through simple excitation and inhibition between neurons.

C. Approach

In order to analyze the computational role of disinhibition in brain function, first we need a general computational model to describe the disinhibition effect. As a first step in analyzing a non-linear system, an equilibrium point equation (IDoG model, Chapter III) is derived from the Hartline-Ratliff equation of disinhibition in the Limulus retina [4] (Chapter II). A further extended model (IDoGS, see Chapter III) is proposed to address the temporal behavior of disinhibition together with a self-inhibition mechanism.

I will apply the computational models to a set of basic circuits to study the local circuits where it has a disinhibitory circuit, e.g. the retina, the cortex, the thalamus, etc. and treat the neural system as a non-linear system in general, so that we can analyze its behavior through a system approach (Chapter VIII). For example, stability, and controllability will be analyzed to understand the computational role of disinhibition.

D. Outline

The organization of this dissertation is illustrated in figure 8. The next chapter will introduce the background of my research (Chapter II). The computational models of disinhibition (Chapter III) are the core of the research. I will apply the framework to different subsystems in the brain, such as in the early visual pathway including the retina (Chapter IV and V) and the primary visual cortex (Chapter VI), the thalamus (Chapter VII and VIII), and the basal ganglia (Chapter VIII). Furthermore, I will model and frame disinhibition in a nonlinear system perspective, then analyze it under a system approach (Chapter VIII) by studying issues such as stability and controllability. Finally, Chapter IX concludes the whole dissertation.

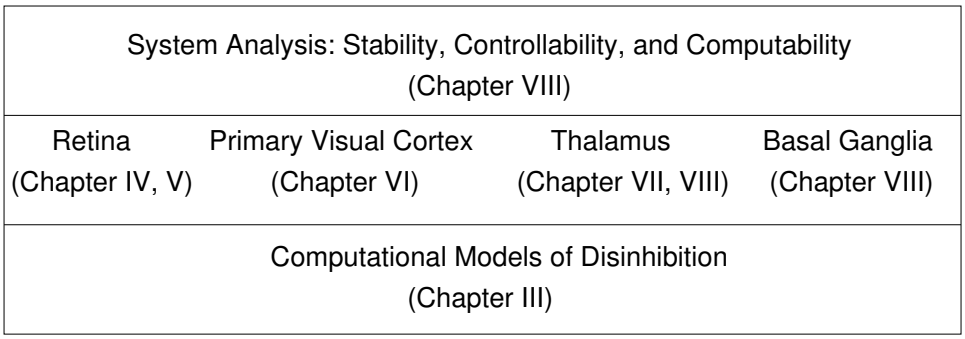


Fig. 8. **Outline.** This dissertation is organized into three layers as shown above. The base of the research is to build a computational model of disinhibition and apply it to various brain subsystems as shown in the second layer. The top layer contains a framework for disinhibition in a nonlinear system perspective, under a system approach based on observations and results from the second layer.

CHAPTER II

BACKGROUND

In this chapter, I will first introduce various disinhibition phenomena at different levels of brain organization (section A). The Hartline-Ratliff equation (section B) and a Fourier model (section C) of disinhibition will be introduced next. Based on the Hartline-Ratliff equation, a general model of disinhibition (Inverse DoG filter; IDoG) will be derived in Chapter III.

A. Various disinhibition phenomena at different levels of brain organization

Disinhibition can be found at various levels of brain organization, such as in the retina, the simple cells in the visual cortex, the thalamus, the basal ganglia, and the cerebellum.

In the early visual pathway, anatomical and physiological observations show that the center-surround property in early visual processing involves a recurrent inhibition, i.e. disinhibition (type I). For example, Hartline and colleagues used Limulus (horseshoe crab) optical cells to demonstrate disinhibition and self-inhibition in the retina [1]. Disinhibition in the early visual pathway has been discovered in mammals and other vertebrates as well. For example, disinhibition has been found in the retina of cats [12; 13], tiger salamanders [14], and mice [15]. The amacrine cells can vigorously rectify the signals in the retina [16]. For example, [13] has shown that the A2 amacrine cells in the cat retina contribute to lateral inhibition among ganglion cells, and they can play a role in disinhibition.

In the primary visual cortex, as shown in figure 9, it is suggested that the horizontal neuronal connectivity between the orientation detectors are recurrent, and thus, it has a disinhibitory effect [17]. Examples of the disinhibition effect in the early visual

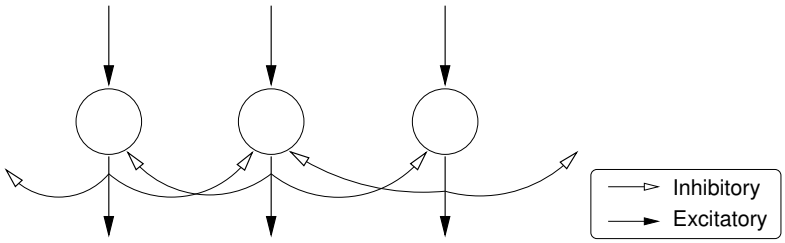


Fig. 9. **A possible configuration of lateral inhibition between orientation detectors.** The synapse with filled arrow is excitatory, and the synapse with unfilled arrow is inhibitory (c.f. [17]).

pathway include the brightness contrast (B-C) visual illusions [18] (e.g. Hermann grid and Mach band, discussed in chapters IV and V) and orientation illusion [19] (e.g. modified Poggendorff illusions in chapter VI).

Deeper inside the brain, disinhibitory circuits have been observed in the thalamus. The thalamus (chapter VIII) consists of relay cells and the thalamic reticular nucleus (TRN). The thalamic reticular nucleus is a collection of GABAergic neurons that form a shell surrounding the dorsal and the lateral aspect of the thalamus. The neurons in TRN are recurrently connected, and their inhibition mechanism is disinhibition in two senses [20]: (1) A TRN cell, say R_1, inhibits another, R_2, and R_2 inhibits the relay cell; (2) and there exists mutual inhibition among TRN cells (see Chapter VIII).

In the motor pathway, for example in the cerebellum, disinhibition is found in a series of inhibitory neurons. The feedforward pathway includes basket cells, Purkinje cells, interpositus nucleus, and inferior olive [10]. Similar type I disinhibition also exists in the basal ganglia, e.g. the pathway from the striatum to the external segment of the globus pallidus (GPe) and substantia nigra (STN), etc. (see also Chapter 10 in [11]).

B. Hartline-Ratliff equation

The Hartline-Ratliff equation [1], the first computational model of disinhibition based on the Limulus optical cells, describes the equilibrium state of a recurrent disinhibitory network. Thus, it serves as a good starting point to study the dynamics of disinhibition.

Experiments by Hartline and Ratliff on Limulus optical cells showed that the disinhibition effect is recurrent (figure 10). The final response of a specific neuron can be considered as the overall effect of the response from itself and from all other neurons. The response resulting from a light stimulus can be enhanced or reduced due to the interactions through inhibition from its neighbors.

The Hartline-Ratliff equation describing disinhibition in Limulus can be summarized as follows [1; 5; 21]:

$$r_m = \epsilon_m - K_s r_m - \sum k_{m \leftarrow n}(r_n - t_{m \leftarrow n}), \tag{2.1}$$

where r_m is the response, K_s the self-inhibition constant, ϵ_m excitation of the m-th ommatidia, $k_{m \leftarrow n}$ the inhibitory weight from other ommatidium, and $t_{m \leftarrow n}$ the threshold. The Hartline-Ratliff equation describes the equilibrium state for a single neuron. I will further extend the equation to calculate neural activity in a large-scale network in a more convenient way (Chapter III).

C. Fourier model

Brodie et al. [22] extended the Hartline-Ratliff equation to derive a spatiotemporal filter, where the input was assumed to be a sinusoidal grating in the form of $I(x,t) = e^{-i(\xi x + \omega t)}$ with spatial frequency ξ and temporal frequency ω (as shown in figure 11).

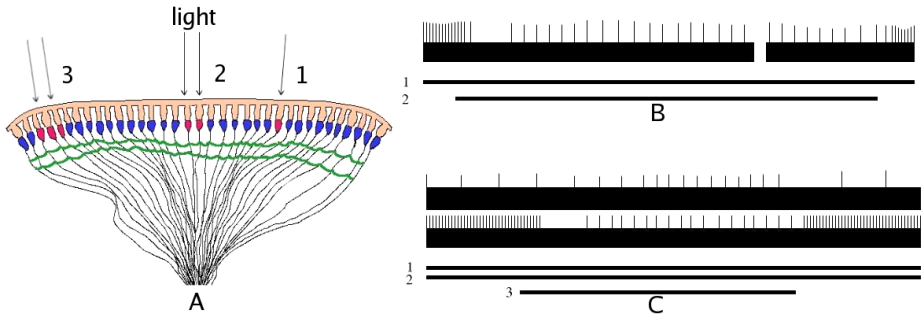


Fig. 10. **Lateral inhibition in Limulus optical cells (Redrawn from [1]).** The figure shows the disinhibition effect in Limulus optical cells. **A**. The retina of Limulus. Point light is presented to three locations (1, 2 and 3). **B**. The result of lighting position 1 and 2. The top trace shows the spike train of the neuron at 1, and the two bars below show the duration of stimulation to cell 1 and 2. When position 2 is excited, the neuron response of position 1 gets inhibited. **C**. Both 1 and 2 are illuminated, and after a short time, position 3 is lighted. The top two traces show the spike trains of cell 1 and cell 2. The three bars below are input duration to the three cells. As demonstrated in the figure, when position 3 is lighted, neurons at position 2 get inhibited by 3, so its ability to inhibit others get reduced. As a result, the firing rate of neuron at position 1 gets increased during the time neuron at position 3 is excited. This effect is called disinhibition.

The output $R(x,t)$ can then be written as

$$R(x,t) = F(\xi,\omega)e^{i(\xi x+\omega t)}, \tag{2.2}$$

where

$$F(\xi,\omega) = \int\int e^{-i(\xi x+\omega t)}\phi(x,t)dxdt, \tag{2.3}$$

and $\phi(x,t)$ is the response of the system to a spatiotemporal impulse (a vertical line at $x = 0$ flashed at the instant $t = 0$).

This model is perfect for explaining the Limulus retina as a filter with a single spatial frequency channel, which means that only a fixed spatial frequency input is

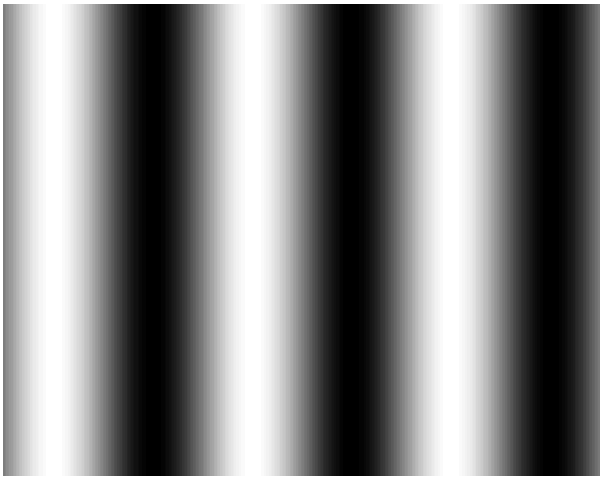


Fig. 11. **Sinusoidal input.** The single spatial frequency input stimulus as shown was used in Brodie's experiments.

allowed [22]. Because of this, their model cannot be applied to a complex input (e.g., visual illusions such as the Hermann grid illusion), as various (or even infinite in many cases) spatial frequencies could coexist in the input. On the other hand, even if we decompose the input into multiple single frequency sources, all the final response from these sources cannot be directly added up: Supposing the system function is $f(x)$, where x is the input, if we decompose x into x_1 and x_2, we have $f(x) = f(x_1 + x_2)$, but $f(x_1+x_2) \neq f(x_1)+f(x_2)$. This is due to the non-linearity of the neural network, whose output is not a simple sum of the inputs. Moreover, the sinusoidal input signal assumed in equation 2.2 is continuous, so it is not suitable for inputs with discrete values (i.e. discontinuities).

D. Summary

In this chapter, I introduced some local circuits inside the brain that include disinhibition. I also introduced the Hartline-Ratliff equation and Brodie's Fourier model of disinhibition. However, those models have their own limitations. The Hartline-Ratliff equation is for single neuron's equilibrium, while Brodie's model is not flexible enough to deal with multiple spatial frequency input. Hence, their models of disinhibition cannot be directly applied to complex images, such as the Hermann grid or the scintillating grid. In the following chapter, I will build upon the Hartline-Ratliff equation and derive a filter that can avoid these problems.

CHAPTER III

COMPUTATIONAL MODELS OF DISINHIBITION*

In this chapter, I will introduce computational models of disinhibition in the form of inverse difference of Gaussians (IDoG, section A), which addresses the spatial properties of disinhibition. In section B, I will introduce a spatio-temporal filter, IDoG with self-inhibition (IDoGS).

A. Spatial disinhibitory filter: The IDoG model

Rearranging the Hartline-Ratliff equation (2.1) and generalizing to n inputs, the responses of n cells can be expressed in a simple matrix form as shown below by assuming the threshold and the self-inhibitory constant to be zero (at this point, we only care for spatial properties of visual illusion, so the assumption of zero self-inhibition rate is reasonable):

$$\mathbf{r} = \mathbf{e} + \mathbf{W}\mathbf{r}, \tag{3.1}$$

where $\mathbf{r}$ is the output vector, $\mathbf{e}$ the input vector and $\mathbf{W}$ the weight matrix:

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ . \\ r_n \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ . \\ e_n \end{bmatrix}. \tag{3.2}$$

---

* Parts of this chapter have been reprinted with permission from "A neural model of scintillating grid illusion: Disinhibition and self-inhibition in early vision" by Yingwei Yu and Yoonsuck Choe, 2006. *Neural Computation*, vol. 18, pp. 501-524. Copyright 2006 by MIT Press.

The weight matrix $\mathbf{W}$ can be assigned its weights following the classic two-mechanism DoG distribution by Marr and Hildreth [23]:

$$W_{ij} = \begin{cases} w(|i,j|) & \text{when } i \neq j \\ \\ 0 & \text{when } i = j \end{cases}, \tag{3.3}$$

$$w(x) = \text{DoG}(x) = k_c e^{-(x/\sigma_c)^2} - k_s e^{-(x/\sigma_s)^2}, \tag{3.4}$$

where $|i,j|$ is the Euclidean distance between neuron i and j; k_c and k_s the scaling constants that determine the relative scale of the excitatory and inhibitory distributions; and σ_c and σ_s their widths.

The response vector $\mathbf{r}$ can finally be derived from equation 3.1 as follows:

$$\mathbf{r} = (\mathbf{I} - \mathbf{W})^{-1}\mathbf{e}. \tag{3.5}$$

Figure 12 shows a single row (corresponding to a neuron in the center) of the weight matrix $(\mathbf{I} - \mathbf{W})^{-1}$ plotted in 2D. Empirically, with DoG profile, it is invertible in all of our experiments in the later chapters. This weight matrix is a symmetric Toeplitz matrix. The complexity of the inverse operation for the symmetric Toeplitz matrix is $O(n \log n)$ as reported by Heinig and Rost [24], thus, computation of relatively large weight matrices can still be efficient. The plot shows that the neuron in the center can have an excitatory influence far away.

Any recurrent lateral inhibitory neural network automatically implements disinhibition (i.e., Type I disinhibition, see section I.B), and IDoG model should be able to model. When there is no recurrent feedback (e.g., the "on" and "off" ganglia cells in retina), the feedforward model (e.g., DoG filter) should be sufficient.

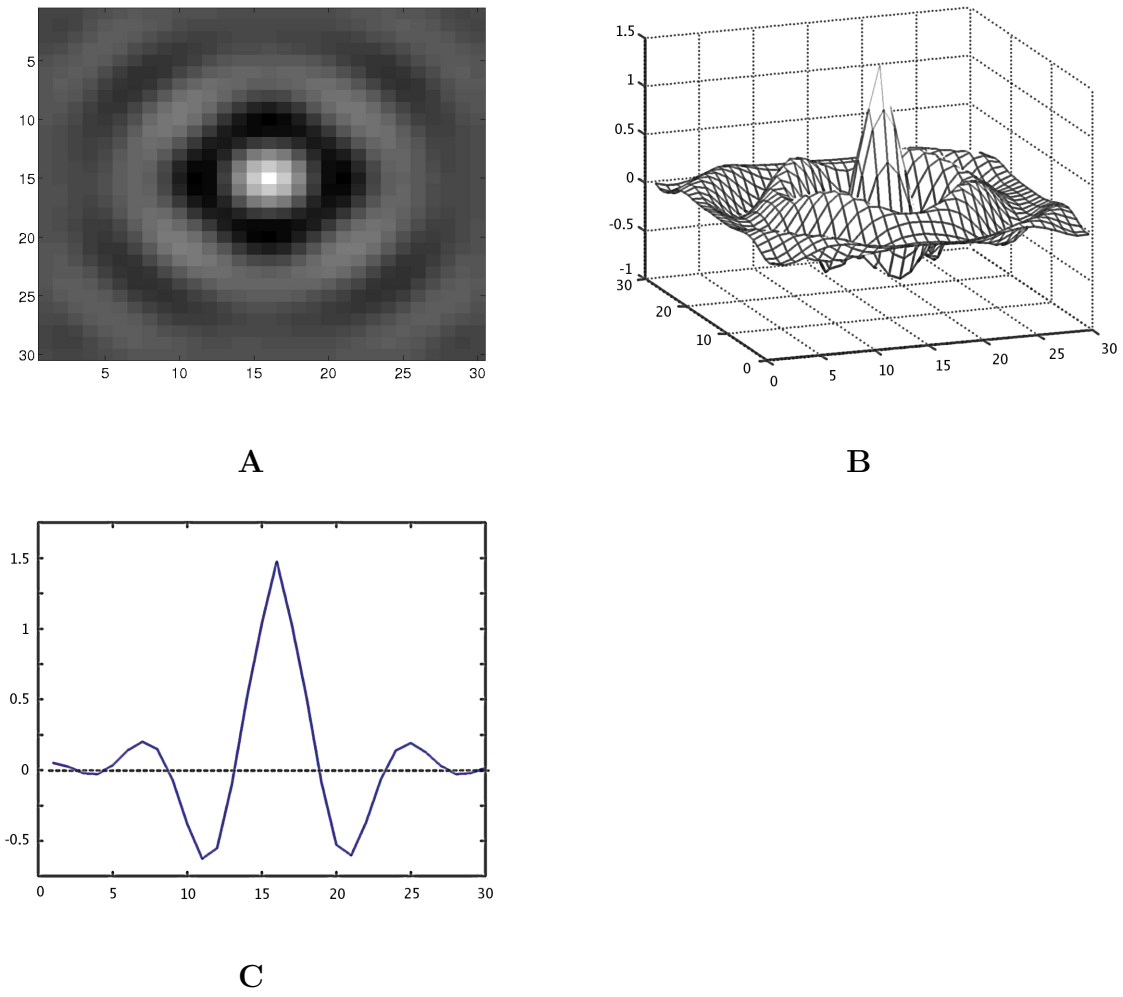


Fig. 12. **An inverse DoG filter (IDOG).** The filter (i.e., the connection weights) of the central neuron is shown. **A**. A 2D plot of the filter. **B**. A 3D mesh plot of the filter. **C**. The plot of the central row of the filter. Note the multiple concentric rippling tails.

B. Spatio-temporal disinhibitory filter: The IDoGS model

The IDoG model can be further extended with self-inhibition. This kind of mechanism adds dynamics to the filter, which can be used to model visual illusions such as the scintillating grid [25], as will be introduced in Chapter V. We can simply add a

self-inhibition factor in the matrix **W** to obtain such a feature, as follows.

$$W_{ij} = \begin{cases} w(|i,j|) & \text{when } i \neq j \\ \\ -K_s(t) & \text{when } i = j \end{cases}, \tag{3.6}$$

where $K_{\mathrm{s}}(t)$ is the self-inhibition rate at time t.

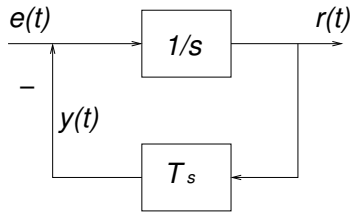


Fig. 13. **The dynamics of self-inhibition.** The input signal is $e(t)$, the output $r(t)$, and the feedback $y(t)$. The block $1/s$ is an integrator, and the feedback block T_s is described by equation 3.9.

For the convenience of calculation, we assume $K_{\mathrm{s}}(t)$ here approximately equals the self-inhibition rate of a single cell (the dynamics are illustrated in figure 13). The exact derivation of $K_{\mathrm{s}}(t)$ is as follows [22]:

$$K_{\mathrm{s}}(t) = \frac{y(t)}{r(t)}, \tag{3.7}$$

where $y(t)$ is the amount of self-inhibition at time t, and $r(t)$ the response at time t for this cell. We know that the Laplace transform $y(s)$ of $y(t)$ has the following property:

$$y(s) = r(s)T_{\mathrm{s}}(s), \tag{3.8}$$

$$T_{\mathrm{s}}(s) = \frac{k}{1 + s\tau}, \tag{3.9}$$

where k is the maximum value $K_{\mathrm{s}}(t)$ can reach, and τ the time constant of decay. By

assuming that the input $e(t)$ is a step input to this cell, the Laplace transform of $e(t)$ can be written as:

$$e(s) = \frac{I_0}{s}, \tag{3.10}$$

where I_0 is a constant representing the strength of the light stimulus. The response $r(t)$ of this cell can be calculated in the following manner:

$$r(s) = \left(\frac{I_0}{s} - r(s) \frac{k}{1 + s\tau} \right) \frac{1}{s}. \tag{3.11}$$

Solving this equation, we get

$$r(s) = \frac{I_0}{s} \frac{s\tau + 1}{\tau s^2 + s + k}. \tag{3.12}$$

By substituting $r(s)$ and $T(s)$ in equation 3.8 with equations 3.9 and 3.12, we get

$$y(s) = \frac{I_0}{s} \frac{(s\tau + 1)}{(\tau s^2 + s + k)} \frac{k}{(1 + s\tau)}. \tag{3.13}$$

Then, by inverse Laplace transform, we can get $y(t)$ and $r(t)$, and finally the exact expression for $K_s(t)$ can be obtained by evaluating equation 3.7. The exact formula for $K_s(t)$ can be derived as follows:

$$K_s(t) = \frac{y(t)}{r(t)}, \tag{3.14}$$

where $y(t)$ is the amount of self-inhibition at time t, and $r(t)$ the response at time t for this cell. We know that the Laplace transform $r(s)$ of $r(t)$ has the following property:

$$r(s)s = e(s) - r(s)T_s(s), \tag{3.15}$$

$$T_s(s) = \frac{k}{1 + \tau}, \tag{3.16}$$

where $T_s(s)$ is a transfer function, k the maximum value $K_s(t)$ can reach, τ the time

constant, and $e(s)$ the Laplace transform of the step input of this cell:

$$e(s) = \frac{1}{s}. \tag{3.17}$$

By rearranging equation 3.15, we can solve for $r(s)$ to obtain

$$r(s) = e(s)\frac{1}{s + T_s}. \tag{3.18}$$

Therefore, $r(t)$ can be treated as the step input function $e(t)$ convolved with an impulse response function:

$$r(t) = e(t) * f(t), \tag{3.19}$$

where "$*$" is the convolution operator, and

$$f(t) = L^{-1}\left[\frac{1}{s + T_s}\right] \tag{3.20}$$

where L^{-1} is the inverse Laplace transform operator. Solving equation 3.20, we get $f(t)$ as a superposition of two exponential functions:

$$f(t) = \frac{1}{C}\left(C_1 \exp(C_2 t) + C_2 \exp(C_1 t)\right), \tag{3.21}$$

where $C = \sqrt{1 - 4\tau k}$, $C_1 = (C + 1)/2$, and $C_2 = (C - 1)/2$. The function $y(t)$ can also be obtained in a similar manner as shown above:

$$y(s) = r(s)T_s(s). \tag{3.22}$$

By substituting $r(s)$ with the right-hand side in equation 3.18, we have

$$y(s) = e(s)\frac{T_s}{s + T_s}. \tag{3.23}$$

Therefore, $y(t)$ can also be treated as the step input function $e(t)$ convolved with an

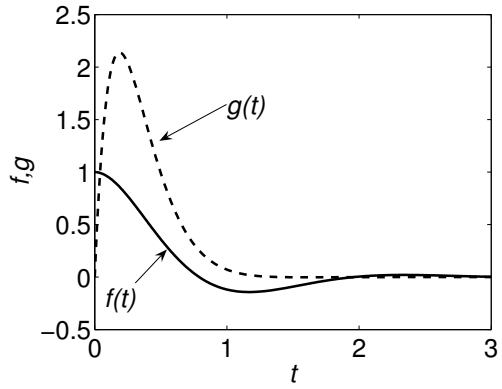


Fig. 14. **The impulse response functions $f(t)$ and $g(t)$.** The final form of self-inhibition function $K_s(t)$ can be calculated as a division of two convolutions of these functions as shown in equation 3.26

impulse response function $g(t)$ in time domain:

$$y(t) = e(t) * g(t) \tag{3.24}$$

where $g(t)$ is a sine-modulated, exponentially decaying function:

$$g(t) = L^{-1}\left[\frac{T_s}{s + T_s}\right] = 6\sqrt{5}\exp(-5t)\sin(\sqrt{5}t). \tag{3.25}$$

Hence the final form of $K_s(t)$ can then be calculated as a division of two convolutions as follows:

$$K_s(t) = \frac{e(t) * g(t)}{e(t) * f(t)}. \tag{3.26}$$

Figure 14 shows the impulse response functions $f(t)$ and $g(t)$. The above derivation gives the exact formula in equation 3.7.

Figure 15 shows several curves plotting the self-inhibition rate under different parameter conditions. As discovered in the Limulus [1; 5], self-inhibition is strong ($k = 3$), while lateral contribution is weak (0.1 or less). These values were experimentally determined by Hartline and Ratliff [1; 5], where τ was left as a free parameter.

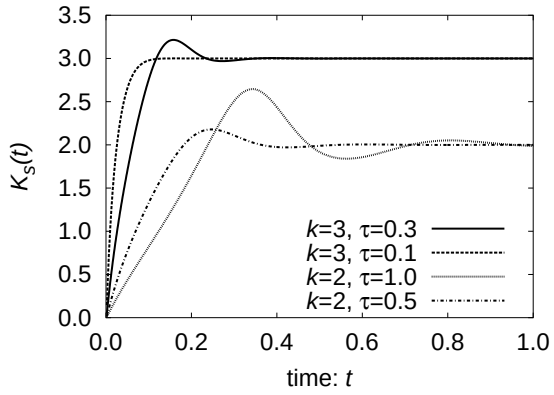
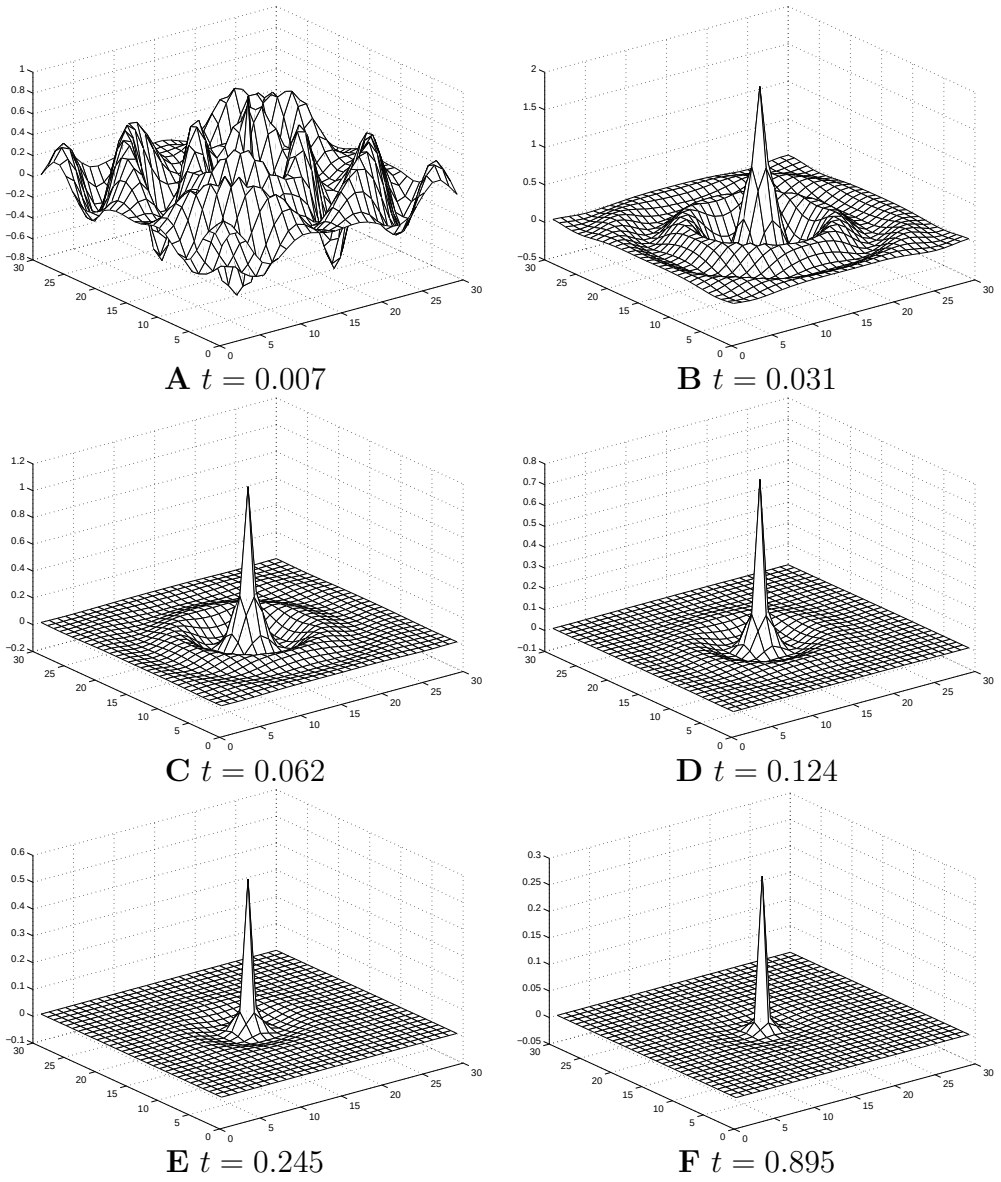


Fig. 15. **Self-inhibition rate.** Evolution of the self-inhibition rate $K_s(t)$ (y-axis) over time (x-axis) is shown for various parameter configurations (see equations 3.7–3.9). The parameter k defines the peak value of the curve, and τ determines how quickly the curve converges to a steady state. For all computational simulations in this dissertation, the values $k = 3$ and $\tau = 0.3$ were used.

Figure 16 shows a single row of the weight matrix $\mathbf{W}$, corresponding to a weight matrix (when reverse serialized) of a single cell in the center of the 2D retina, at various time points. The plot shows that the cell in the center can be influenced by the inputs from locations far away, outside of its classical receptive field area, and the range and magnitude of influence dynamically change over time.

C. Summary

DoG filter only accounts for feedforward lateral inhibition. If we add recurrent feedback to the neural networks with DoG profile, then it will automatically implement disinhibition. In this chapter, I derived two computational models of disinhibition: IDoG and IDoGS. Each model has its own specialty. The IDoG model is a spatial model of disinhibition, and it calculates the equilibrium state of a single layer neural network. The IDoGS is a spatiotemporal model of disinhibition and self-inhibition.



A $t = 0.007$ **B** $t = 0.031$ **C** $t = 0.062$ **D** $t = 0.124$ **E** $t = 0.245$ **F** $t = 0.895$

Fig. 16. **Inversed DoG with self-inhibition filter (IDoGS) at various time points.** The filter (i.e., the connection weights) of the central optical cell is shown at different time steps ($k = 3, \tau = 0.3$). The self-inhibition rate evolved over time as follows: **A** $K_{\mathrm{s}}(t) = 0.0299$, **B** $K_{\mathrm{s}}(t) = 0.1463$, **C** $K_{\mathrm{s}}(t) = 0.2855$, **D** $K_{\mathrm{s}}(t) = 0.5438$, **E** $K_{\mathrm{s}}(t) = 0.9890$, and **F** $K_{\mathrm{s}}(t) = 2.5940$. Initially, a large ripple extends over a long distance from the center (beyond the classical receptive field), but as time goes on the long-range influence diminishes. In other words, the *effective* receptive field size reduced over time due to the change in self-inhibition rate.

The next few chapters (Chapters IV-VII) will employ these two model to predict static and dynamic perception. These two computational models gave a general framework of disinhibition, and they made the analysis of disinhibition in brain function mathematically accessible.

CHAPTER IV

ROLE OF SPATIAL DISINHIBITION IN STATIC BRIGHTNESS-CONTRAST PERCEPTION*

Brightness-contrast (B-C) illusions allow us to understand the basic processes in the early visual pathway, and many of them can be explained by disinhibition. B-C illusions can become very complex, and a complete explanation may have to be based on a multi-stage, multi-channel model, with considerations of top-down influences [26; 27; 28]. In the following, however, we will focus on the very early stages of visual processing, and see how far we can exploit low-level mechanisms observed in biological vision systems toward explaining B-C illusions.

In the Hermann grid, the illusory dark spots at the intersections in the Hermann grid (Figure 17A) are due to lateral inhibition [2]. Lateral inhibition is the effect observed in the receptive field where the surrounding area inhibits the central area. The visual signal in the eye is generated by the photoreceptor cells, and then it is passed through bipolar, horizontal, and amacrine cells and finally sent to LGN. When the stimulus is given in the receptive field, the central receptors produce an excitatory signal, while the cells in the surrounding area send inhibition through the bipolar cells to the central area [29]. Difference of Gaussian (or DoG), filter [23] is commonly used to simulate such a process.

However, DoG filters alone cannot account for more complex visual B-C illusions. For example in the Hermann grid illusion, although the illusory spots are explained

---

* Parts of this chapter have been reprinted with permission from "Explaining low level brightness-contrast visual illusion using disinhibition" by Yingwei Yu, Takashi Yamauchi, and Yoonsuck Choe, 2004. In A. J. Ijspeert, M. Murata, and N. Wakamiya, editors, *Biologically Inspired Approaches to Advanced Information Technology*, *Lecture Notes in Computer Sciences 3141*, pp.166-175.

pretty well by the conventional DoG model, it cannot explain why the periphery (figure 17A, to the left) appears brighter than the illusory dark spots (figure 17A, to the right). The output is counter to our perceived experience. The reason for this failure is that the center of DoG in the peripheral area receives inhibition from all the surrounding directions which results in a weaker response than the intersections in the grid which only receive inhibition from four directions. Moreover, the White's effect [30] (figure 19A) cannot be explained using the conventional DoG filter. As shown in figure 19B, the output using conventional DoG filters gives an opposite result: The left gray patch on the black strip has a lower output value than the one on the white strip. On the contrary, we perceive that the left gray patch on the black strip as brighter than the one on the right.

Anatomical and physiological observations show that the center-surround property in early visual processing may not be strictly feedforward, involving lateral inhibition and, moreover, disinhibition. Note that disinhibition has been found in vertebrate retina such as in tiger salamanders [14] and in mice [31], and it can effectively reduce the amount of inhibition in case we have a large area of bright input. This might be a potential solution to the unsolved visual illusion problems above.

A. The Hermann grid illusion

Figure 17B and C show Hermann grid under DoG filter. When the size of the "on" center (positive Gaussian) in DoG profile matches the width of the streets in the grid, it shows the strongest illusory effect [2]. Hence, for any given receptive field size, there exists a corresponding size of Hermann grid that optimally gives rise to the illusion. Figure 17C shows the brightness level of the middle row in Figure 17B, where the dark illusory spots are clearly visible (P_1, P_2 and P_3).

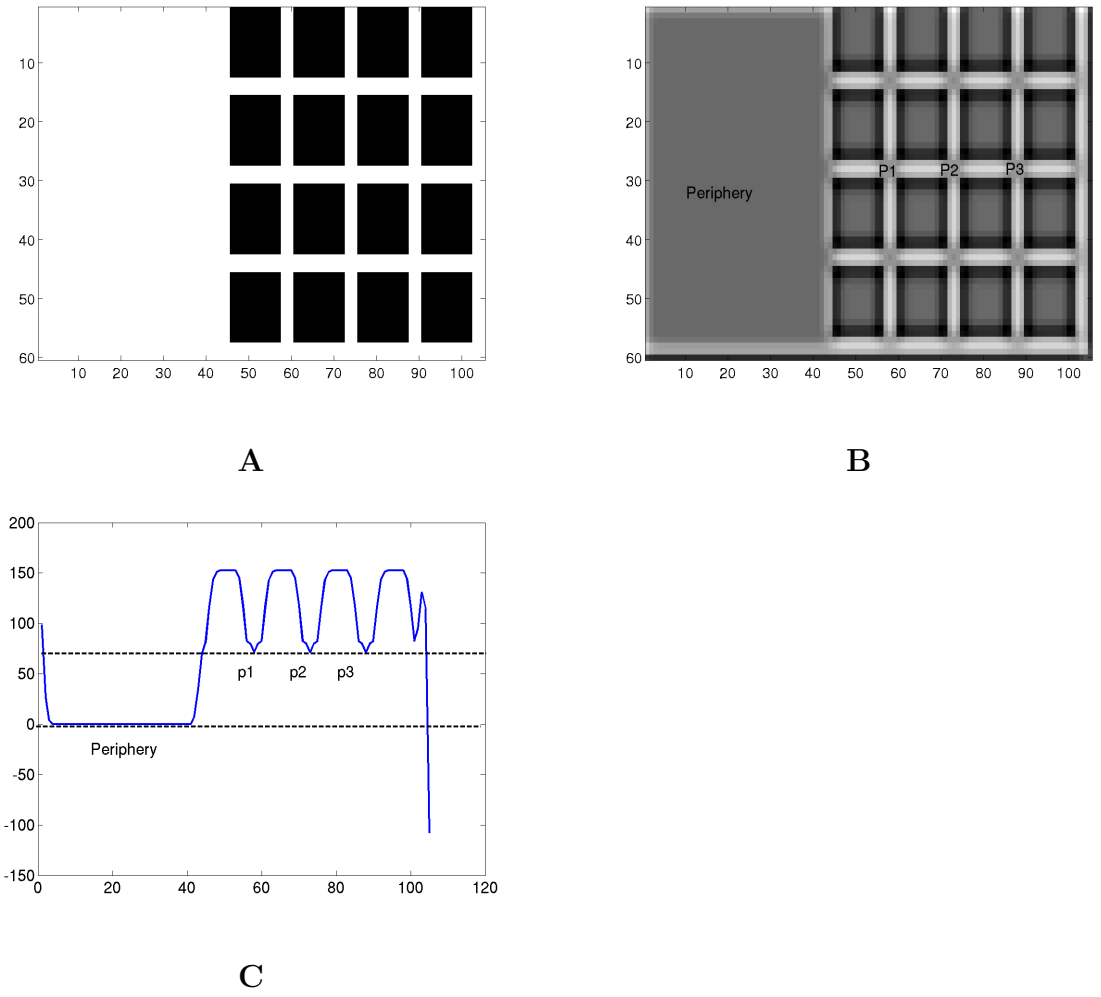
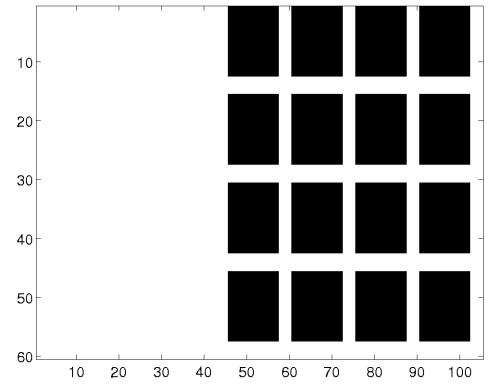
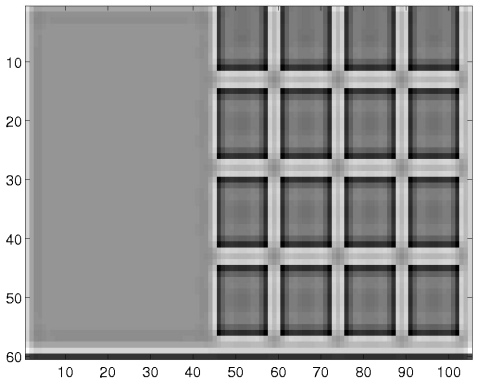


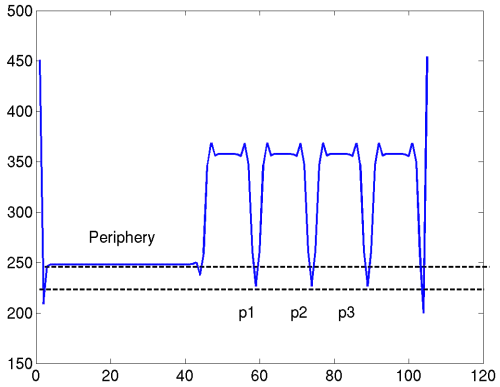
Fig. 17. **The Hermann grid illusion under DoG filter.** **A**. The Hermann grid illusion. The intersections look darker than the streets. **B**. The output using a conventional DoG filter. **C**. Quantified brightness level in B. To measure the average response, the column-wise sum of rows 27 to 29 was computed. Note that the illusory spots (at positions P_1, P_2 and P_3) have a brightness value much higher than the periphery. The conventional DoG operation cannot explain why we perceive the periphery to be brighter than the dark illusory spots.



A



B



C

Fig. 18. **The Hermann grid illusion under IDoG filter.** **A**. The Hermann grid illusion. **B**. The output response of IDoG. **C**. The prediction using the IDoG filter (from B). The illusory spots are at position P_1, P_2 and P_3, which have a brightness value lower than the periphery. (The curve shows the column-wise sum of rows 27 to 29.)

To account for the peripheral brightness in the Hermann grid, the IDoG filter was used. Our IDoG filter which explicitly models disinhibition provides a plausible explanation to this problem. Figure 18 shows the result of applying our filter to the Hermann grid image: C is the plot of the column-wise sum (rows 27-29) of the filter response in B. The periphery is indeed brighter than the dark illusory spots, showing that disinhibition (and hence IDoG) can account for the perceived brightness in this particular example.

As shown in the figure 18B, the IDoG filter can capture the brightness-contrast phenomenon in three spatial scales: (1) in the high spatial frequency scale (about 4 pixels), the contrast at the borders of black blocks are enhanced; (2) in the moderate spatial frequency scale (about 8 pixels), illusory dark spots appear; and (3) in the low spatial frequency scale (about 20 pixels), the periphery is brighter than the black blocks and the illusory dark spots. Compared to IDoG, the feedforward DoG filter can only preserve the brightness-contrast in one spatial scale. As shown in figure 17B, the results by DoG filter only predicted the illusory dark spots (moderate spatial frequency), but missed the brightness-contrast information in both high and low spatial frequencies.

B. The White's effect

The White's effect [30] is shown in figure 20A: The gray patch on the black vertical strip appears brighter than the gray patch on the right. As shown in figure 19, DoG cannot explain this illusion. Disinhibition plays an important role in this illusion: While the gray patch on the black strip receives inhibition from the two surrounding white strips, compared to the gray patch on the right side, disinhibition is relatively stronger. Because of this, the gray patch on the right appears darker than that on

the left (figure 20C).

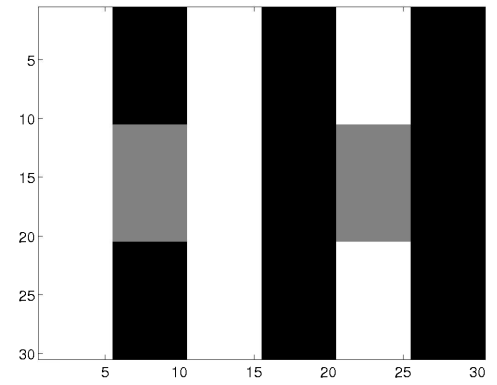
C. The Mach band

Compared to the conventional DoG filter, one advantage of the IDoG model is that it preserves the different level of brightness and also enhances the contrast at the edge. As demonstrated in figure 21, the four shades of gray are clearly separated using IDoG. These different shades are not preserved using a conventional DoG filter. Note that this can be simply because the sum of the DoG matrix equals zero, and scaling up k_c in equation 3.4 can correct the problem. However, there is one subtle point not captured in the conventional DoG approach: the wrinkle (figure 21E) near the Mach bands observed in Limulus experiments [32]. Compared to the IDoG result, we can clearly see that this wrinkle is absent in the DoG output (figure 21C).

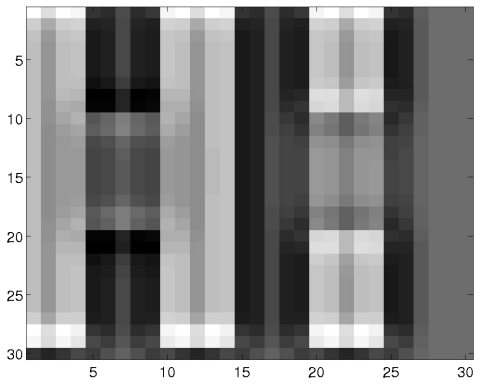
D. Summary

We have shown that certain limitations of DoG filters can be overcome by explicitly modeling disinhibition, and that a disinhibitory filter (IDoG) can be used to explain several brightness-contrast illusions (e.g., the Hermann grid, the White's effect, and the Mach band). The functional benefit of disinhibition in the early visual pathway is to preserve brightness and enhance contrast in multiple spatial scales as shown in the Hermann grid and the Mach band experiments.

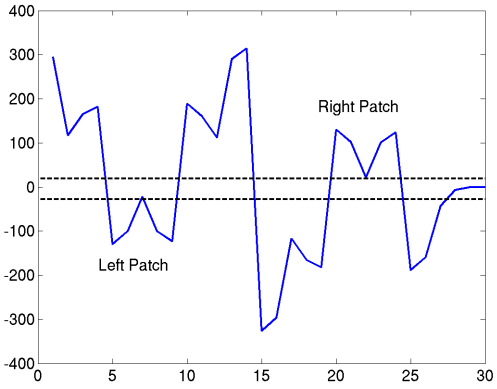
In the next chapter, we will apply the dynamic model of disinhibition, IDoGS, to predict the scintillating grid illusion, which is a variation of the Hermann grid illusion.



A



B



C

Fig. 19. **The White's effect under DoG filter.** **A**. The White's effect. The gray patch on the left has the same gray level as the one on the right, but we perceive the left to be brighter than the right. **B**. The output using a conventional DoG filter. **C**. The brightness level of the two gray patches calculated using conventional DoG filter. As in the previous figure, rows of 10 to 19 in the output were added to get the average response. Note that the left patch has a lower average value (below zero) than the right patch (above zero). The result contradicts our perceived brightness.

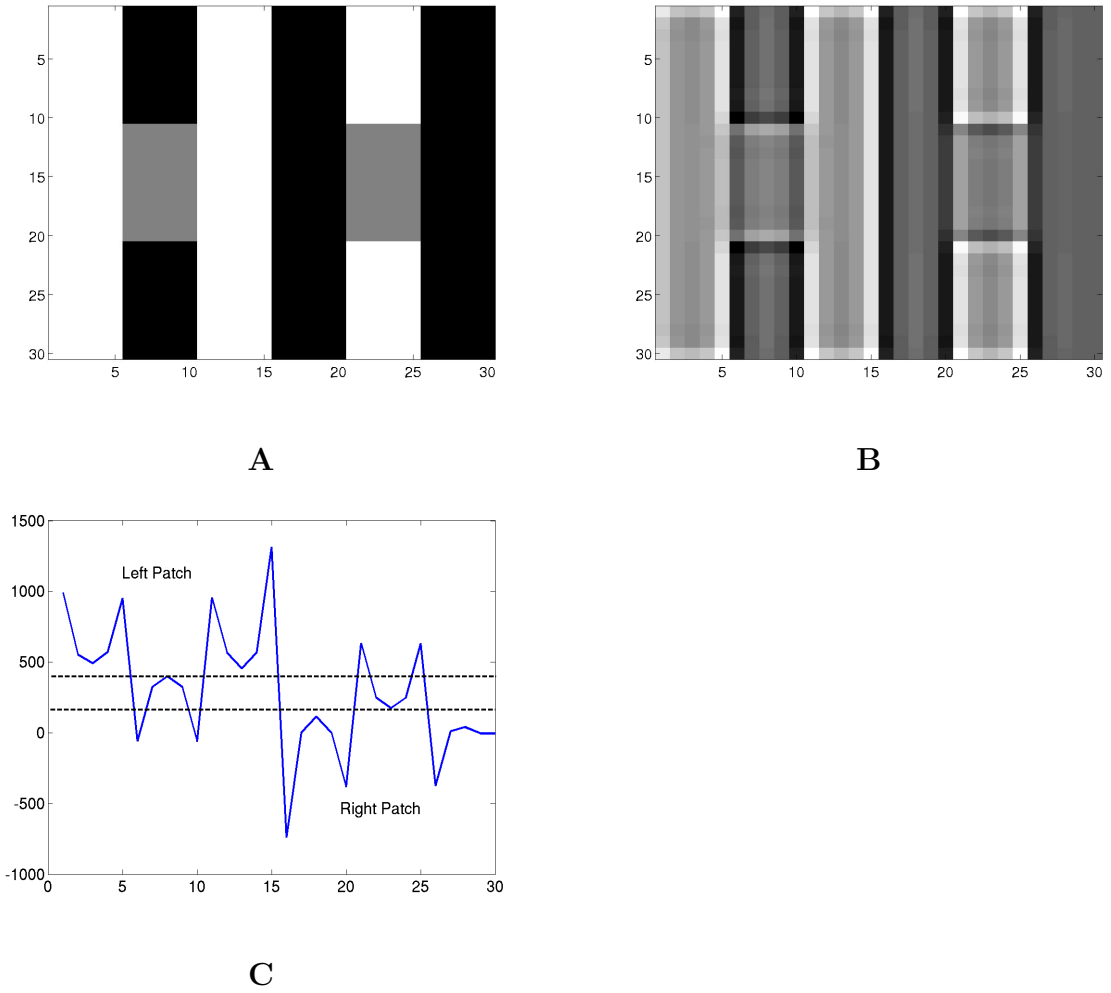


Fig. 20. **The White's effect and prediction under IDoG filter.** **A**. The White's effect stimulus. **B**. The output using IDoG. **C**. The prediction using the IDoG model. The gray patch on the left results in a higher value than that in the right. The curve shows the column-wise sum of rows 11 to 19.

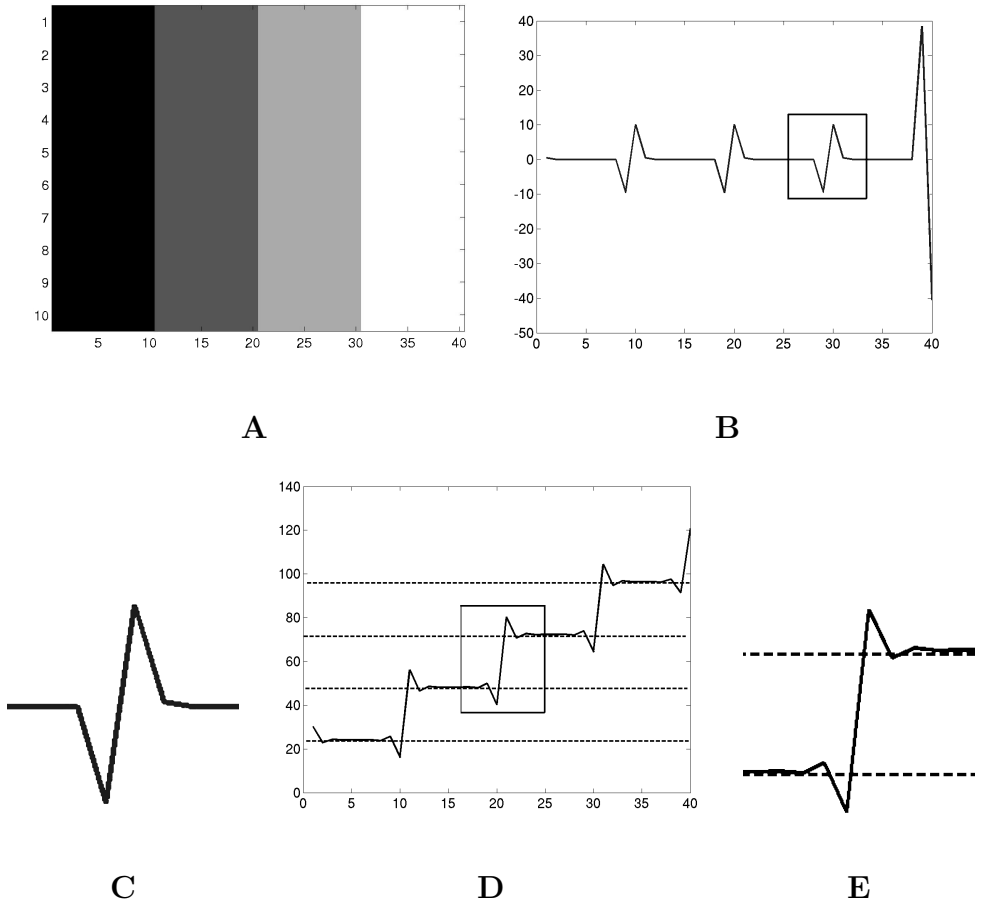


Fig. 21. **The Mach band under DoG and IDoG.** **A**. The Mach band input image. **B**. The output using a conventional DoG filter. The different brightness levels are not preserved. **C**. An expanded view of the inset in B. **D**. The output using IDoG. The different brightness levels are preserved. **E**. An expanded view of the inset in D, which shows wrinkles near luminance edge, unlike in **C**.

CHAPTER V

ROLE OF SPATIO-TEMPORAL DISINHIBITION IN DYNAMIC BRIGHTNESS-CONTRAST PERCEPTION*

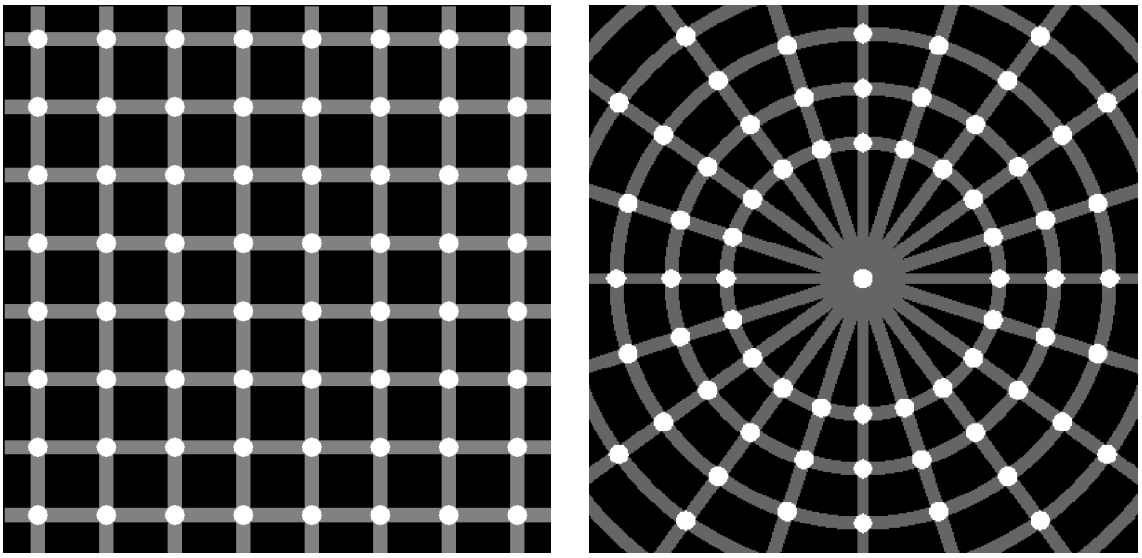
In the previous chapter I used the IDoG model to predict static brightness-contrast perception, such as the Hermann grid or the White's effect. In this chapter, I will analyze in detail an interesting visual illusion, scintillating grid, which is a static figure that can give rise to a dynamic brightness-contrast perception. To explain this phenomenon, I employed a dynamic model of disinhibition, IDoGS, to explain the striking visual effect.

A. Scintillating grid illusion

As we introduced in Chapter I, section A.2, the scintillating grid illusion consists of bright discs superimposed on intersections of orthogonal gray bars on a dark background (figure 22A) [3]. Several important properties of the illusion have been discovered and reported in recent years: (1) The discs that are closer to a fixation show less scintillation [3], which might be due to the fact that receptive fields in the periphery are larger than those in the fovea. As shown in figure 23, if the periphery of the scintillating grid is correspondingly scaled up, the scintillation effect is diminished. Note that the diminishing effect is not due to the polar arrangement alone, as can be seen in figure 22B. (2) The illusion is greatly reduced or even abolished both with steady fixation and by reducing the contrast between the constituent grid elements

---

* Parts of this chapter have been reprinted with permission from "A neural model of scintillating grid illusion: Disinhibition and self-inhibition in early vision" by Yingwei Yu and Yoonsuck Choe, 2006. *Neural Computation*, vol. 18, pp. 501-524.



A Scintillating Grid **B** Scintillating grid in polar arrangement

Fig. 22. **The scintillating grid illusion and its polar variation. A** The original scintillating grid illusion is shown (redrawn from [3]). **B** A polar variation of the illusion is shown. The scintillating effect is still strongly present in the polar arrangement (cf. [34]).

[3]. (3) As speed of motion is increased (either efferent eye-movement or afferent grid movement), the strength of scintillation decreased [33]. (4) The presentation duration of the grid also plays a role in determining the strength of illusion. The strength first increases when the presentation time is less than about 220 ms, but it slowly decreases once the presentation duration is extended beyond that [33].

What kind of neural process may be responsible for such a dynamic illusion? Anatomical and physiological observations show that the center-surround property in early visual processing involves disinhibition and self-inhibition, inhibition of the cell itself [1; 12; 13; 14; 15; 16]. For self-inhibition, it is found that depolarization of a rod bipolar cell in the rat retina evokes a feedback response to the same cell [35], thus indicating that a mechanism similar to those in the Limulus may exist in

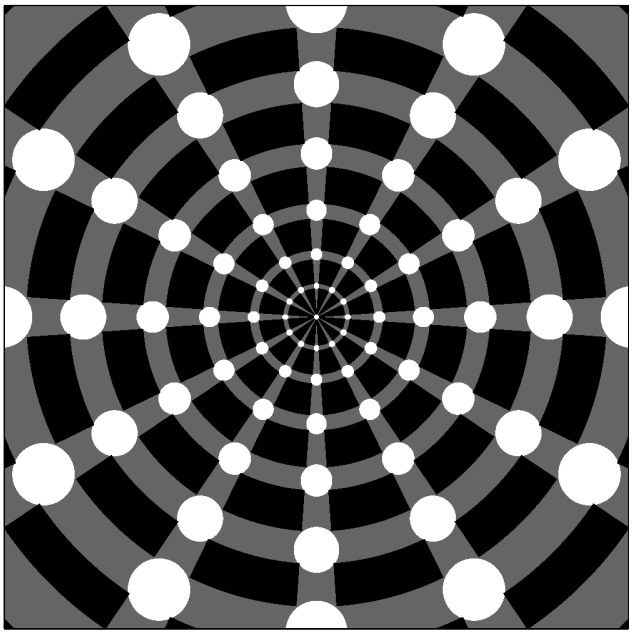


Fig. 23. **A variation without the scintillating effect.** The grids toward the periphery are significantly scaled up, which results in the abolishment of the scintillating effect when stared in the middle. This is because the scintillating grid illusion highly depends on the size of the receptive fields. In the fovea, the receptive field size is small and in the periphery, the receptive field size is relatively larger. (Note that Kitaoka [34] presented a similar plot, but there the periphery was not significantly scaled up such that the scintillating effect was preserved.)

mammalian vision. Other computational models also suggested that self-inhibition may exist in cells sensitive to light-dark contrast [36]. Disinhibition can effectively reduce the amount of inhibition in the case where there is a large area of bright input, and self-inhibition can give rise to oscillations in the response over time. Thus, the combination of those two mechanisms, i.e., disinhibition and self-inhibition, may provide an explanation to the intriguing Scintillating grid illusion.

B. Methods

To match the behavior of the model to psychophysical data, we need to measure the degree of the illusory effect in the scintillating grid. More specifically, we are interested in the change over time in the relative contrast of the disc vs. the gray bars:

$$S(t) = C(t) - C(0), \tag{5.1}$$

where $S(t)$ is the perceived strength t time units from the last eye movement or the time of initial presentation of the scintillating grid stimulus (time t in our model is on an arbitrary scale) , and $C(t)$ is the contrast between the disc and the gray bars in the center row of the response matrix:

$$C(t) = \frac{R_{\text{disc}}(t) - R_{\min}(t)}{R_{\text{bar}}(t) - R_{\min}(t)}, \tag{5.2}$$

where $R_{\text{disc}}(t)$ is the response at the center of the disc region, $R_{\text{bar}}(t)$ the response at the center of either of the gray bar regions, and $R_{\min}(t)$ the minimum response in the output at time t. In other words, the function of perceived strength of illusion $S(t)$ is defined as the relative disc-to-bar contrast at time t as compared to its initial value at time 0.

Using this measure, in the experiments below, we tested our model under various experimental conditions, mirroring those in [3; 33]. In all calculations, the effect of illusion was measured on an image consisting of a single isolated grid element of size 30×30 pixels. The disc at the center had a diameter of 8, and the bars had a width of 6. The model parameters $k = 3$ and $\tau = 0.3$ were fixed throughout all experiments and so was the pattern where the background luminance was set to 10, the gray bar to 50, and the white disc to 100, unless stated otherwise. Dependent on the experimental condition under consideration, the model parameters (receptive

field size ρ) and/or the stimulus conditions (such as the duration of exposure to the stimulus and/or brightness of different components of the grid) were varied. The units of the receptive field size, the width of the bar, and the diameter of the disc were all equivalent; in pixels on the receptor surface, where each pixel corresponds to one photo receptor. The details of the variations are provided in the experiments section below.

C. Experiments and results

1. Experiment 1: Perceived brightness as a function of receptive field size

In the scintillating grid illusion, the scintillating effect is most strongly present in the periphery of the visual field. As we stated earlier, this may be due to the fact that the receptive field size is larger in the periphery than in the fovea, thus matching the scale of the grid. If there is a mismatch in the scale of the grid and the receptive field size, the illusory dark spot would not appear. For example in figure 23, the input is scaled up in the periphery, thus creating a mismatch between the peripheral receptive field size and the scale of the grid. As a result, the scintillating effect is abolished. Conversely, if the receptive field size is reduced in size with no change to the input, the perceived scintillation would diminish (as it happens in the center of gaze in the original scintillating grid: figure 22).

To verify this point, we tested our model with different receptive field sizes while the input grid size was fixed. As shown in figure 24A, smaller receptive fields results in almost no darkening effect in the white disc.

In sum, these results could be an explanation to why there is no scintillating effect in figure 23. In the original configuration, the peripheral receptive fields were large enough to give rise to the dark spot, however, in the new configuration, they

are not large enough, and thus no dark spot can be perceived.

2. Experiment 2: Perceived brightness as a function of time

In this experiment, the response of the model at different time steps was measured. In figure 25A–E, five snapshots are shown. In the beginning, the dark spot can clearly be observed in the center of the disc, but as time goes on, it gradually becomes brighter. Figure 25F plots the relative brightness of the disc compared to the bars as a function of time, which shows a rapid increase to a steady state. Such a transition from dark to bright corresponds to a single scintillation (Note that the opposite effect, bright to dark, is achieved by refreshing of the neurons via saccades). Figure 25G shows the actual response level in a horizontal cross section of the response matrix shown in figure 25A–E. Initially, the response to the disc area shown as the sunken plateau in the middle is relatively low compared to that to the gray bars represented by the flanking areas (bottom trace, white ribbon). However, as time passes by the difference in response between the two areas dramatically increases (top trace, black ribbon). Again, the results show a nice transition from a perception of a dark spot to that of a bright disc.

3. Experiment 3: Strength of scintillation as a function of luminance

The strength in perceived illusion can be affected by changes in the luminance of the constituent parts of the scintillating grid, such as the gray bar, disc, and the dark background (figure 26A and C) [3]. Figure 26B and D show a variation in response in our model under such stimulus conditions. Our results show a close similarity to the experimental results by Schrauf et al. [3]. As the luminance of the gray bar increases, the strength of illusion increases, but after reaching about 40% of the disc brightness, the strength gradually declines (figure 26B), consistent with experimental

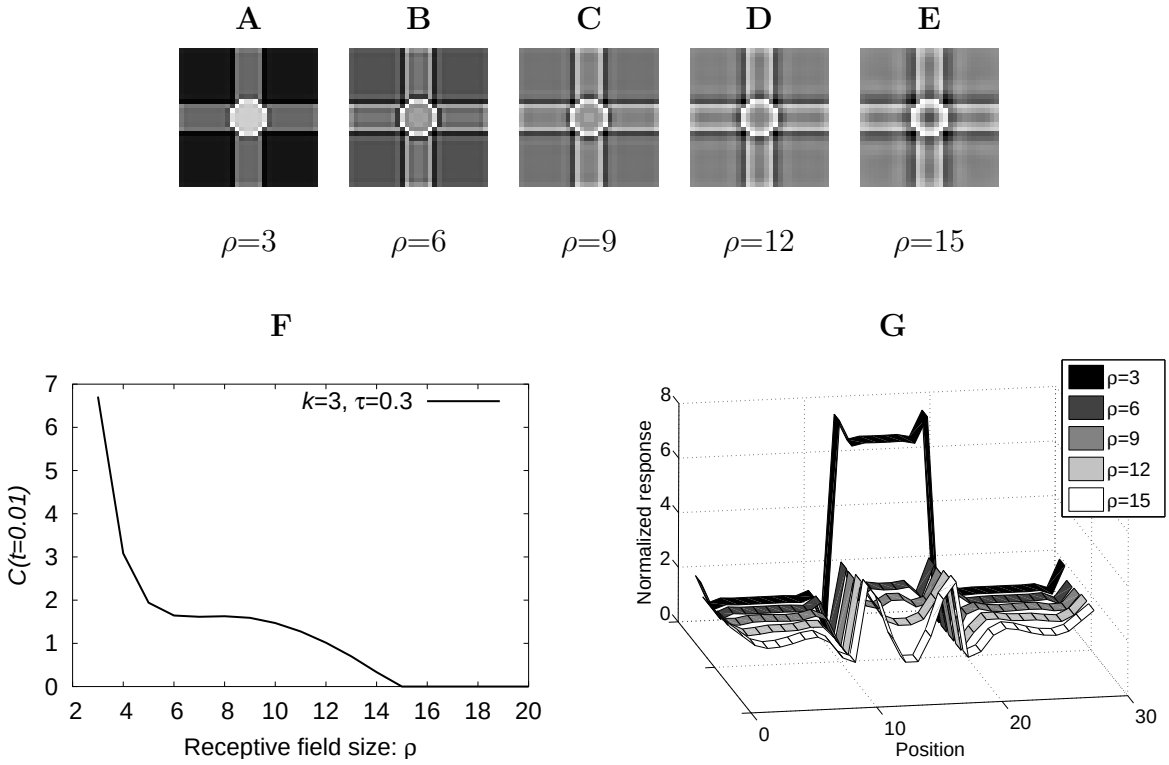


Fig. 24. **Response under various receptive field sizes.** The response of our model to a single grid element in the scintillating grid is shown, under various receptive field sizes at $t = 0.01$. **A**–**E** The responses of the model are shown, when the receptive field size was increased from $\rho = 3$ to 6, 9, 12, and 15. Initially, the disc in the center is bright (simulating the fovea), but as ρ increases, it becomes darker (simulating the periphery). **F** The relative brightness level of the central disc compared to the gray bar $C(t)$ is shown (equation 5.2). The contrast decreases as ρ increases, indicating that the disc in the center becomes relatively darker than the gray bar region. The contrast drops abruptly until around $\rho = 6$ and then gradually decreases. **G** The normalized responses of the horizontal cross section of **A**–**F** are shown. For normalization, the darkest part and the gray bar region of the horizontal cross section were scaled between 0.0 and 1.0. When ρ is small (=3), the disc in the center is very bright (the plateau in the middle in the black ribbon), but it becomes dark relative to the gray bars as ρ increases (white ribbon).

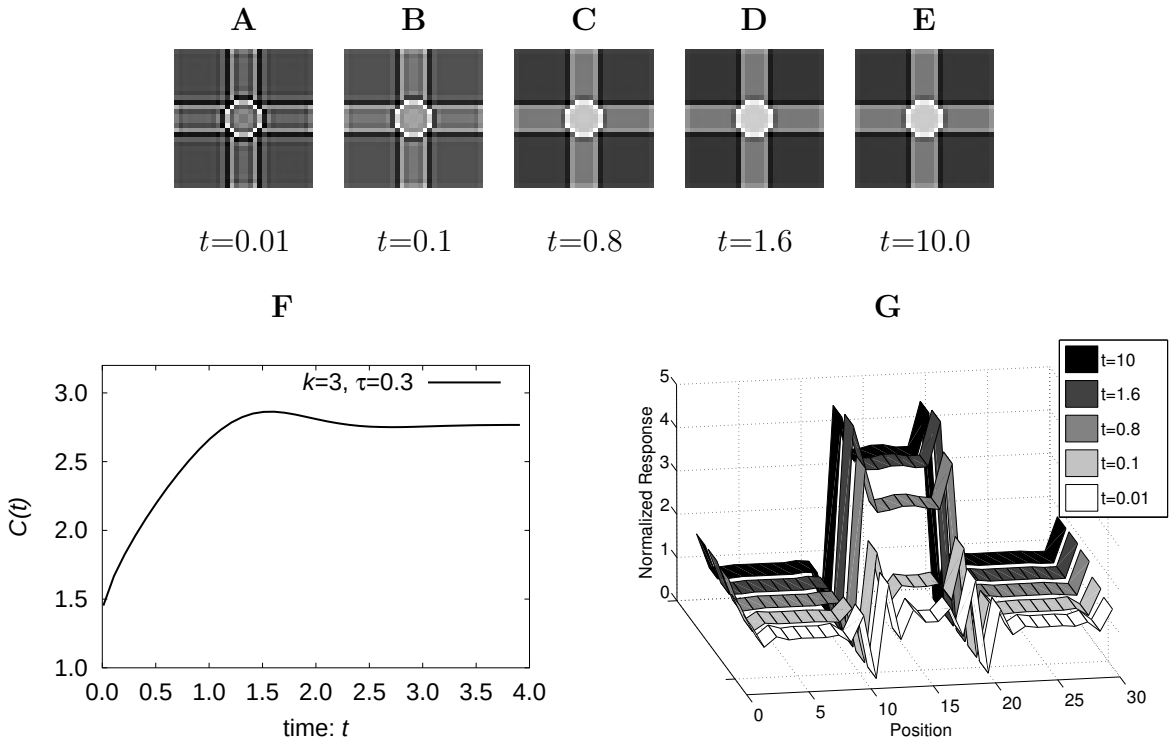


Fig. 25. **Response at various time points.** The response of the model to an isolated scintillating grid element is shown over time. The parameters used for this simulation were: receptive field size = 6 (represents the periphery), $k = 3$, and $\tau = 0.3$. The plots demonstrate a single blinking effect of the white disc. **A** In the beginning when the self-inhibition rate is small, the illusory dark spot can be seen in the central disc ($K_s(t) = 0.0495$). **B**–**E** As time goes on, the illusory dark spot disappears as the self-inhibition rate increases $K_s(t) = 0.04521$, $K_s(t) = 2.4107$, $K_s(t) = 3.2140$, and $K_s(t) = 3$, respectively. **F** The relative brightness level of the central disc compared to the gray bar $C(t)$ is shown (equation 5.2). The results demonstrate an increase in the relative perceived brightness of the center disc as time progresses. **G** The normalized response of the horizontal cross section of the **A**–**E** are shown. Normalization was done as described in figure 24G. In the beginning ($t = 0.01$), the disc region in the middle is almost level with the flanking gray bar region (white ribbon near the bottom). However, as time goes on, the plateau in the middle rises, signifying that the disc in the center is becoming perceived as brighter.

results (figure 26A). Such a decrease is due to disinhibition, which cannot be explained by DoG [18].

When the luminance of the disc was increased, the model (figure 26D, right) demonstrated a similar increase in the scintillating effect as in the human experiment (figure 26C, right). When the disc has a luminance lower than the bar, Hermann grid illusion occurs [3]. Both the human data (figure 26C, left) and the model results (figure 26D, left) showed an increase in the Hermann grid effect when the disc became darker.

Note that disinhibition plays an important role here, especially for the bar luminance experiments (figure 26A–B). In standard DoG, which lacks the recurrent inhibitory interaction, the illusory effect will monotonically increase with the increase in the luminance of the gray bars. However, with disinhibition, the increasing illusory effect will reach a critical point followed by a decline. (See the section V.D for more discussion.)

4. Experiment 4: Strength of scintillation as a function of motion speed and presentation duration

As we have seen above, the scintillating effect has both a spatial and a temporal component. Combining these two may give rise to a more complex effect. Schrauf et al. demonstrated that such an effect in fact exists [33]. They conducted experiments under three conditions: (1) smooth pursuit movements executed across a stationary grid (efferent condition); (2) grid motion at an equivalent speed while the eyes are held stationary (afferent condition); (3) brief exposure of a stationary grid while the eyes remained stationary. For conditions 1 and 2, both afferent and efferent motion produced very similar results: The strength of scintillation gradually decreased as the speed of motion increased (figure 27A). For condition 3, the strength of illusion

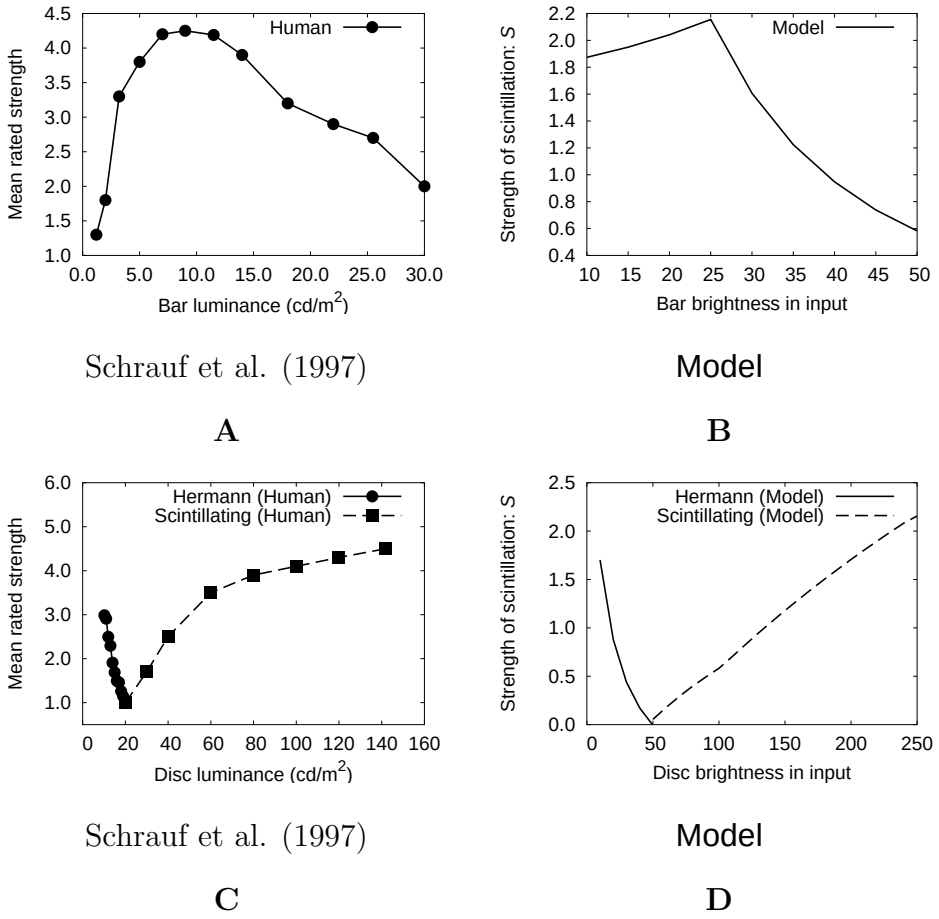


Fig. 26. **Strength of scintillation under various luminance conditions. A** Mean rated strength of scintillation in human experiments is shown as a function of disc luminance [3]. **B** Scintillation effect in the model is shown as a function of bar luminance. **C** Mean rated strength of scintillation in human experiments is plotted as a function of bar luminance [3]. The plot shows results from two separate experiments: Hermann grid on the left, and scintillating grid on the right. **D** Hermann grid and scintillation effects in the model are plotted as functions of disc luminance. Under both conditions, the model results closely resemble those in human experiments. For **B** and **D**, the strength of the scintillation effect in the model was calculated as $S = C(\infty) - C(0)$, where $C(\infty)$ is the steady state value of $C(t)$ (see equation 5.1). The illusion strength in the Hermann grid portion in **D** was calculated as $S = 1/C(\infty) - 1$. The reciprocal was used because in the Hermann grid, the intersection is darker than the bars, whereas in the scintillating grid, it is the other way around (disc is brighter than the bars).

abruptly increases, coming to a peak at around 200 ms and then slowly decreases (figure 27C). We tested our model under these conditions, to verify if temporal dynamics induced by self-inhibition can accurately account for the experimental results.

First, we tested the model when either the input or the eye was moving (assuming that condition 1 and 2 above are equivalent). In our experiments, instead of directly moving the stimulus, we estimated the effect of motion in the following manner. Let v be the corresponding speed of motion, either afferent of efferent. From this, we can calculate the amount of time elapsed before the stimulus (or the eye) move on to a new location. For a unit distance, the elapsed time t is simply an inverse function of motion speed v, thus the effect of illusion can be calculated as $S(v^{-1})$. Figure 27B shows the results from our model, which closely reflects the experimental results in figure 27A.

Next, we tested the effect of stimulus flash duration on our model behavior. Figure 27D shows our model's prediction of the brightness as a function of the presentation duration. In this case, given a duration of d, the strength of illusion can be calculated as $S(d)$. The perceived strength initially increases abruptly up till around $t = 1.5$, then it slowly decreases until it reaches a steady level. Again, the computational results closely reflect those in human experiments (figure 27C). The initial increase might be due to the fact that the presentation time is within the time period required for one scintillation, and the slow decline may be due to no new scintillation being produced after the first cycle as the eyes were fixated, so that the overall perception of the scintillating strength declines.

In summary, our model based on disinhibition and self-inhibition was able to accurately replicate experimental data under various temporal conditions.

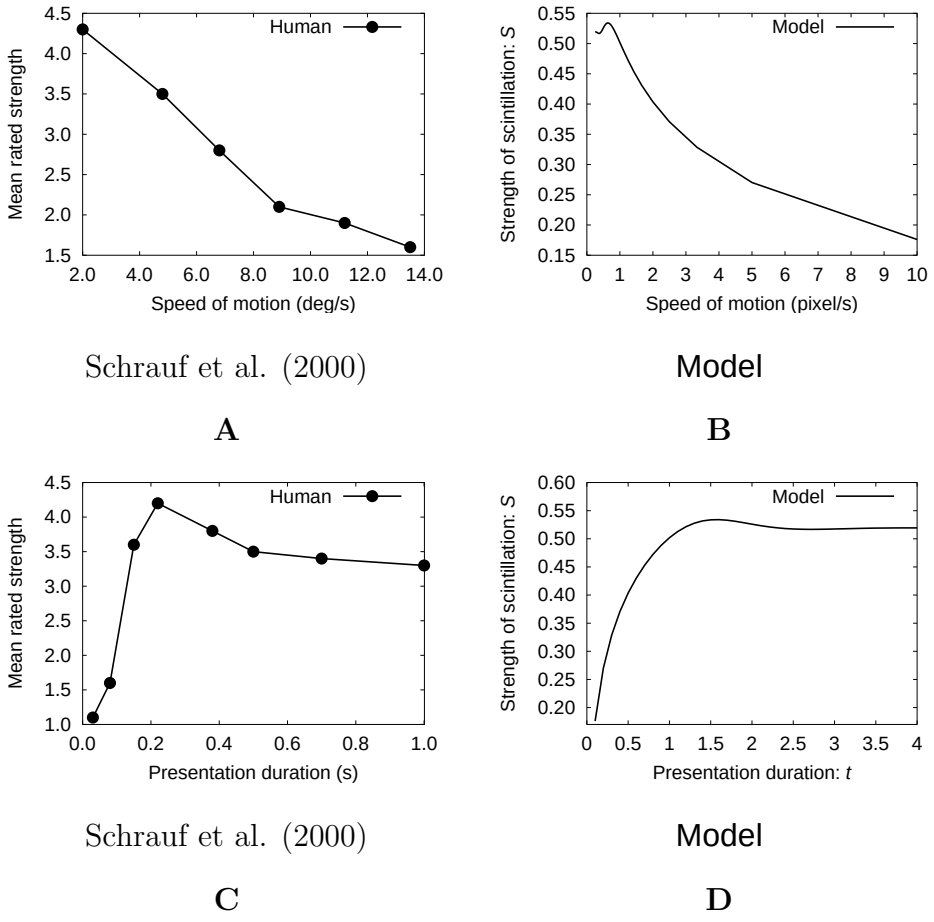


Fig. 27. **Strength of scintillation under varying speed and presentation duration.** **A** Mean rated strength of the illusion as a function of the speed of stimulus movement is shown [33]. **B** Scintillation effect as a function of the speed of motion (v) in the model is shown. The receptive field size was 6, and the strength of scintillation was calculated as $S(v^{-1}) = C(v^{-1}) - C(0)$. **C** Mean rated strength of the illusion as a function of the duration of exposure is shown [33]. **D** Scintillation effect as a function of presentation duration (t) in the model is shown. The receptive field size was 6, and the strength of scintillation was computed as $S(t) = C(t) - C(0)$. In both cases **A**–**B** and **C**–**D**, the curves show a very similar trend.

D. Discussion

The main contribution of this part of the dissertation was to provide, to our knowledge, the first neurophysiologically grounded computational model to explain the scintillating grid illusion. We have demonstrated that disinhibition and self-inhibition are sufficient mechanisms to explain a broad range of spatio-temporal phenomena observed in psychophysical experiments with the scintillating grid. DoG filter failed to account for the change in the strength of scintillation, because it does not incorporate the disinhibition mechanism nor the dynamics of self-inhibition. Disinhibition can effectively reduce the amount of inhibition in the case where there is a large area of bright input [18]. Therefore, DoG filter without disinhibition mechanism cannot explain why the dark illusory spots in the scintillating grid are perceived to be much darker than those in the Herman grid. The reason is, DoG filter predicts that the white bars in the Hermann grid should give stronger inhibition to its intersection than the gray bars in the scintillating grid to its disc. Thus, according to DoG, the intersection in the Herman grid should appear darker than that in the scintillating grid, which is contrary to the fact. However, with a disinhibition mechanism, since disinhibition is stronger in the Hermann grid than in the scintillating grid (because the bars are brighter in the Hermann grid, there is more disinhibition), the inhibition in the center of the Hermann grid is weaker than that in the scintillating grid. Thus, the center appears brighter (because of weaker inhibition) in the Hermann grid than in the scintillating grid, due to disinhibition. Regarding the issue of dynamics, the lack of self-inhibition mechanism in DoG filter makes it fail to explain the temporal properties of the scintillation.

There are certain issues with our model which may require further discussion. In our simulations, we used a step input with an abrupt stimulus onset. In a usual

viewing condition, the scintillating grid as a whole is presented and when the gaze moves around the scintillating effect is generated. All the while, the input is continuously present, without any discontinuous stimulus onset. Thus, the difference in the mode of stimulus presentation could be a potential issue. However, as Schrauf [33] observed, what causes the scintillation effect is not the saccadic eye movement *per se*, but the transient stimulation which the movement brings about. Thus, such a transient stimulation can be modeled as a step input, and the results of our model may well be an accurate reflection of the real phenomena.

Another concern is about the way we measured the strength of the scintillation effect in the model. In our model, we were mostly concerned about the change in the perceived brightness of the disc over time (equation 5.1), whereas in psychophysical experiments, other measures of the effect have been incorporated, such as the perceived number of scintillating dark spots [33]. However, one observation is that the refresh rate of the stimulus depends on the number of saccades in a given amount of time. Considering that a single saccade triggers an abrupt stimulus onset, we can model multiple saccades as a series of step inputs in our simulations. Since our model perceives one scintillation per stimulus onset, the frequency of flickering reported in the model can be modulated exactly by changing the number of stimulus onsets in our simulations. A related issue is the usage of a single grid element (instead of a whole array) in our experiments. It may seem that the scintillation effect would require at least a small array (say 2×2) of grid elements. However, as McAnany and Levine have shown [37], even a single grid element can elicit the scintillating effect quite robustly, thus, the stimulus condition in our simulations may be sufficient to model the target phenomenon.

Besides technical issues as discussed above, there are more fundamental questions that need to be addressed. Our model was largely motivated by the pioneering

work by Hartline et al. in the late fifties. However, the animal model they used was the Limulus, an invertebrate with compound eyes, thus the applicability of our extended model in human visual phenomena may be questionable. However, disinhibition and self-inhibition, the two main mechanisms in the Limulus, have been discovered in mammals and other vertebrates as we mentioned in the introduction. Mathematically, the recurrent inhibitory influence in the disinhibition mechanism and the self-inhibitory feedback are the same in both the limulus and in mammals. Therefore, our model based on the Limulus may generalize to human vision.

Finally, an important question is whether our bottom-up model accounts for the full range of phenomena in the scintillating grid illusion. Why should the scintillating effect only originate from such a low level in the visual pathway? In fact recent experiments have shown that part of the scintillating effect can arise based on top-down, covert attention [38]. The implication of Van Rullen and Dong's study [38] is that even though the scintillation effect can originate in the retina, it can be modulated by later stages in the visual hierarchy. This is somewhat expected because researchers have found that the receptive field properties (which may include the size) can be effectively modulated by attention (for a review, see [39]). It is unclear how exactly such a mechanism can affect brightness-contrast phenomena which depend on the receptive field size at such a very low level, thus it may require further investigation. Schrauf and Spillmann [40] also pointed out a possible involvement of a later stage, by studying the illusion in stereo-depth. But, as they admitted, the *major* component of the illusion may be retinal in origin. Regardless of these issues, modeling spatio-temporal properties at the retinal level may be worthwhile, by serving as a firm initial stepping stone upon which a more complete theory can be constructed.

E. Summary

In this chapter, I presented a neural model of the scintillating grid illusion, based on disinhibition and self-inhibition in early vision. The two mechanisms inspired by neurophysiology were found to be sufficient in explaining the multi-facetted spatio-temporal properties of the modeled phenomena. I expect the IDoGS model to be extendible to the latest results that indicate a higher-level involvement in the illusion, such as that of attention.

CHAPTER VI

ROLE OF DISINHIBITION IN ORIENTATION PERCEPTION*

Besides the misinterpretation of the brightness-contrast level, human vision may misinterpret the location or orientation of visual objects. The Poggendorff illusion is a good example to demonstrate this category of visual illusion. As shown in figure 28A and B, our perception of an angle is usually greater than the actual angle (expansion effect), but when there are multiple lines and thus multiple angles, the expansion effect can either be enhanced or reduced. In this chapter, we will examine the interference effect in a modified Poggendorff illusion (figure 28C and D).

A. Poggendorff illusion

In the original Poggendorff illusion (see, e.g., [41; 42]), the top and the bottom portions of the penetrating thin line are perceived as misaligned (Figure 28A). Figure 28B shows how such a perception of misalignment can occur. The line on top forms an angle α with the horizontal bar, but the perceived angle α' is greater than α. As a result, the line on top in Figure 28A is perceived to be collinear with line 4 at the bottom, instead of line 3 which is physically collinear. However, when an additional bar is added, the illusory angular expansion effect is altered: the effect is either reduced (Figure 28C) or enhanced (Figure 28D) depending on the orientation of the newly added bar. Understanding the functional organization and the low-level neurophysiology underlying such a nontrivial interaction is the main aim of this chapter.

---

* This chapter is a significantly expanded version of [19] "Angular disinhibition effect in a modified Poggendorff illusion" by Yingwei Yu and Yoonsuck Choe, in *Proc. 26th Annual Conference of the Cognitive Science Society*, Kenneth D. Forbus, Dedre Gentner, and Terry Regier, Eds., 2004, pp. 1500-1505.

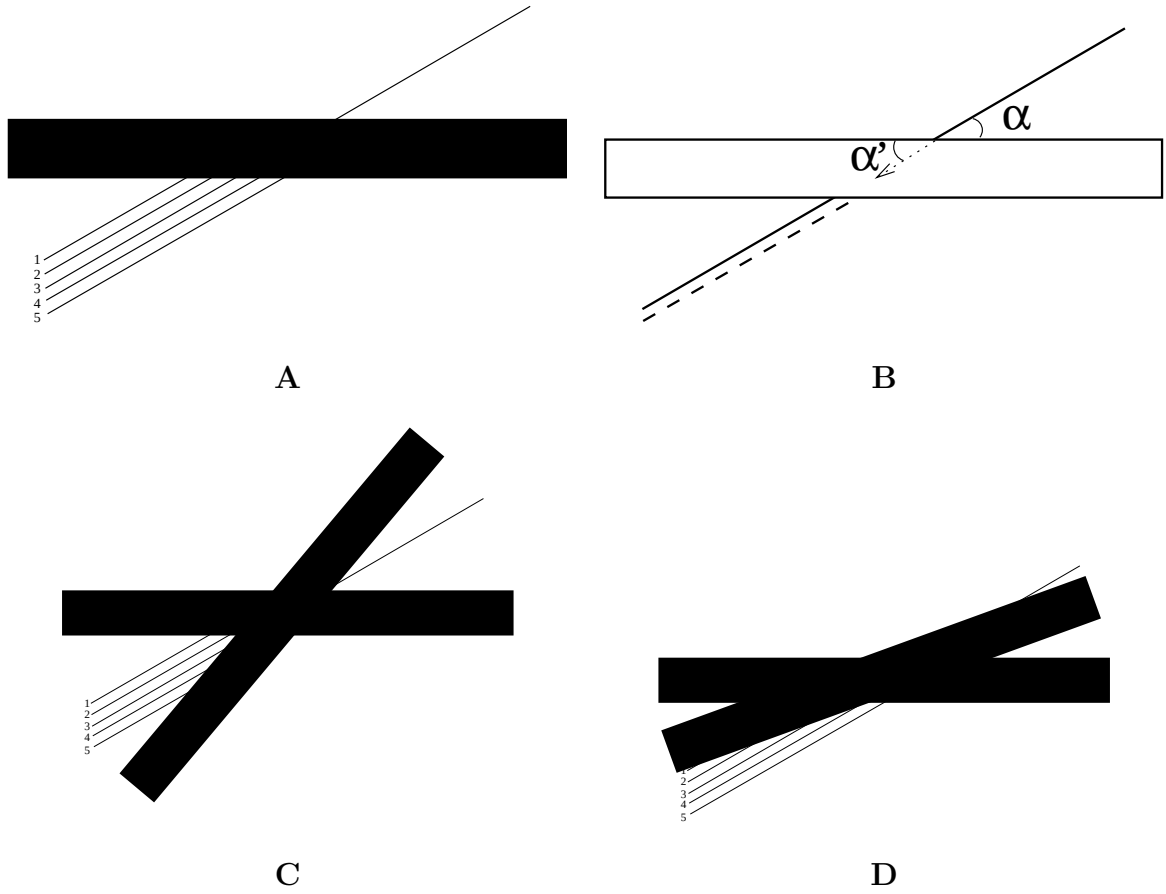


Fig. 28. **The Poggendorff Illusion. A** The original Poggendorff illusion is shown. The five lines below the horizontal bar are labeled 1 to 5 from top to bottom. Line 3 is physically collinear with the line on top. However, line 4 is perceived to be collinear. **B** The actual angle α ($= 30^{\circ}$) and the perceived angle α' ($> 30^{\circ}$) are shown. The solid line shows the straight line penetrating the bar. The dashed line below shows the perceived direction in which the line on top seemingly extends to. **C** The Poggendorff figure with an additional bar at 50° is shown. In this case, line 2 is perceived to be collinear (i.e., $\alpha' < 30^{\circ}$). **D** The Poggendorff figure with an additional bar at 20° is shown. For this case, unlike in **C**, line 5 is perceived to be collinear ($\alpha' > 30^{\circ}$). (The angle α' in this case is slightly greater than in the original Poggendorff figure.)

Neurophysiologically, in the original case where only two orientations interact, lateral inhibition between orientation-tuned cells in the visual cortex can explain the exaggeration of perceived angle. However, as we have seen in Figures 28C and D, with an additional orientation response, lateral inhibition is not enough to explain the resulting interference effect. Our observation is that this complex response is due to disinhibition [4; 1; 5; 21; 22]. Unlike models using simple lateral inhibition, we explicitly accounted for disinhibition in our computational model to describe the complex interactions between multiple orientation cells. The resulting model based on the neurophysiology of the early visual system was able to accurately predict the perceptual performance in the modified Poggendorff illusion.

The rest of the chapter is organized as follows. The next section demonstrates our experimental methods. Then, a neurophysiological motivation for our computational model is presented, followed by a detailed mathematical description of the model. Next, the results from computational experiments with the model is presented and compared to psychophysical data we gathered, followed by discussion and summary.

B. Methods

To quantify the interference effect in the modified Poggendorff illusion, we conducted a psychological experiment. Two subjects with normal vision participated in the experiment. A CRT display panel with a 1600×1200 resolution was used to display the stimuli at a distance of 30 cm. The computer program displayed two thick bars and one thin line on the screen, similar to the stimuli in Figure 28C. The first thick bar was fixed in the center of the screen at 0^o, with a width of 100 pixels. The thin line, 5 pixels in width, intersected the horizontal bar at a fixed angle of 30^o. The second thick bar, 100 pixels in width, intersected at the same point as the other two, whereas

the angle was varied from trial to trial. The stimulus display program also displayed up to 10 thin lines (all at 30°) below the horizontal bar, from which the subjects were asked to choose the one that is the most collinear to the thin line above the bar. The subjects were allowed to click on the line of choice, and then the perceived angle was recorded for each click. Afterward, a new stimulus was generated. A total of 101 trials were recorded for each subject. The experimental results are reported later in "Results" section VI.D, together with computational results.

C. Model

Let us first consider how orientation columns in the visual cortex interact in response to several intersecting lines. For each line at the intersection, there are corresponding orientation columns that respond maximally, which can be approximated by a Gaussian response distribution. As multiple simple cells are activated by different lines at the intersection, the response levels will interact with each other through lateral connections. Thus, there are two issues we want to address in our model: (1) what exactly is the activation profile (or the response distribution) of the orientation-tuned cells, and (2) how these cells interact with each other through lateral connections.

1. Activation profile of orientation columns

Each simple cell in the primary visual cortex responds maximally to visual stimuli with a particular orientation, say θ. The response of these cells y_θ to different orientations x can be modeled as a Gaussian function:

$$y_\theta(x) = y_0 + \frac{a}{\sigma\sqrt{\pi/2}} e^{-2\frac{(x-\theta)^2}{\sigma^2}}, \tag{6.1}$$

where y_0 is the response offset; θ the center (or mean); σ the standard deviation; and a a scaling constant [43].

It also comes to our attention that the cell tuned to a certain orientation, say α, should respond to the opposite orientation, which is $+180^o$. However, experiments have shown that the peak at the position $+180^o$ is somewhat smaller than the peak at α [44]. To accurately model this, we need two Gaussian curves to fit the response of a cell to a full range of orientations from -180^o to 180^o.

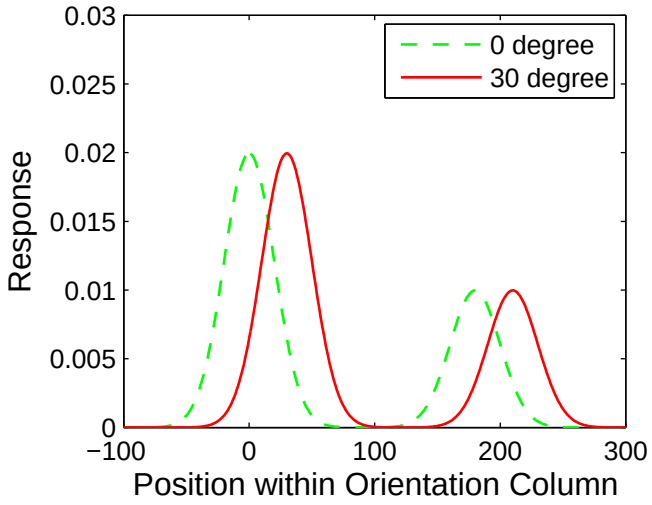
The fitting curve can be written as follows:

$$y_\theta(x) = y_0 + \frac{a}{\sigma\sqrt{\pi/2}} e^{-2\frac{(x-\theta)^2}{\sigma^2}} + \frac{ak}{\sigma\sqrt{\pi/2}} e^{-2\frac{(x-\theta-\pi)^2}{\sigma^2}}, \tag{6.2}$$

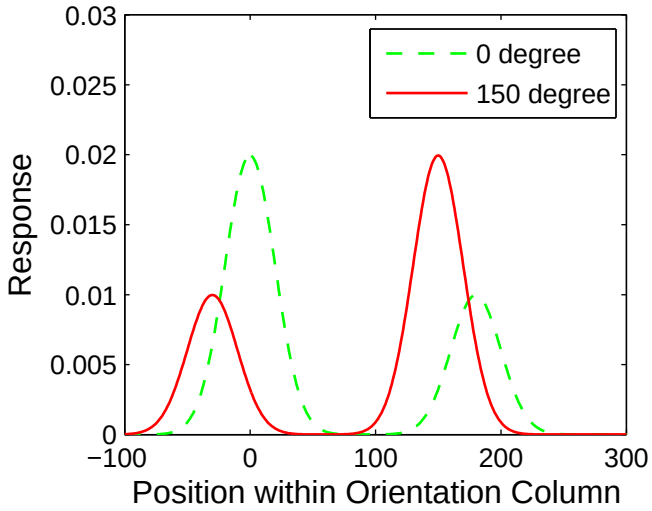
where k is the degree of activation for the opposite direction ($k < 1$). All other terms have the same definition as in Equation 6.1. Such an asymmetric response enables the simple cells to be sensitive to the direction, as well as the orientation of the stimulus.

Using the equation, we can now visualize the response profile of simple cells tuned to orientations ranging from 0^o to 360^o. Figure 29A shows the responses of orientation columns tuned to 90^o, given inputs of two different orientations, 0^o and 30^o. Figure 29B shows the responses of the same set of orientation columns to inputs of two orientations, 0^o and 150^o. From these two figures, we can observe that for each specific orientation input, the excitation is tuned at that value with a peak in the Gaussian curve, and at the same time, the opposite direction-tuned cell shows a lower peak response. The asymmetry in responses occurs in both acute (Figure 29A) and obtuse angles (Figure 29B). Note that even though the difference in orientation between 0^o vs. 30^o (Figure 29A) and 0^o vs. 150^o (Figure 29B) is 30^o in both cases, the response profile greatly differs in the 0^o vs. the 150^o case.

Next, we will investigate how response profiles in multiple orientation columns



A



B

Fig. 29. **Activation profile.** (**A**) The activation of simple cells in response to an acute angle is shown. The dashed curve is the response of the cells in an orientation column (x-axis) to a horizontal line of 0°, and the solid curve that to a 30° line. (**B**) The activation of simple cells in response to an obtuse angle is shown. The dashed curve is the response of the cells in an orientation column to a horizontal line of 0°, while the solid curve is the response to a 150° line.

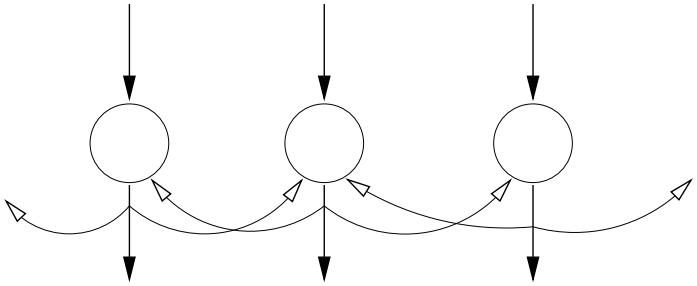


Fig. 30. **A possible configuration of lateral inhibition between orientation detectors.** The lines with unfilled arrows illustrate mutual inhibition between cells, and the lines with filled arrow are excitatory synapses. (Adapted from [45].)

can interact.

2. Column level inhibition and disinhibition

Our observation that angular enlargement sometimes seems to be weakened when there are more than two bars or lines in the Poggendorff illusion (Figure 28) led us to hypothesize about the potential role of a recurrent inhibition effect, i.e., disinhibition. Figure 30 shows the recurrent feedback network structure proposed by Carpenter and Blakemore [45] which can account for the observed properties of angle expansion. They suggested that the horizontal neuronal connectivity between the orientation detectors are recurrent in humans, and thus it can implement disinhibition.

3. Applying disinhibition to orientation cells

Orientation sensitive cells in the cat visual cortex are known to inhibit each other [46; 47; 48]. From this, we can postulate that a group of cells tuned to the same orientation representing different lines (e.g., intersecting lines) may compete with each other through inhibition.

Now let us consider the mathematical description for inhibition at the column level. Suppose there are n lines with orientations $\{\theta(1), \theta(2), ...\theta(n)\}$ intersecting at one point. Let the initial responses of orientation columns to each line be $\{e_1, e_2, ..., e_n\}$, where e_i is the column response vector to input line i. In the orientation column for the i-th line input, let α be the position in the orientation column whose cell is tuned to the orientation α. The initial excitation $e_i(\alpha)$ can be calculated as

$$\mathbf{e}_i(\alpha) = d_i y_{\theta(i)}(\alpha) \tag{6.3}$$

where d_i is the width of the i-th input line, $\theta(i)$ the orientation of the i-th input line. In this way, we can calculate the initial excitation e of the cell which is tuned to α, responding to the i-th input line.

By the definition of disinhibition, the final response r_i of orientation columns i can be obtained as follows:

$$\mathbf{r}_i = \mathbf{e}_i - \mathbf{W}\mathbf{r}_i, \tag{6.4}$$

where $\mathbf{W}$ is a constant matrix of inhibition strengths (or weights), controlled by a parameter: $\mathbf{W}_{ij} = \eta$ if $i \neq j$, and 0 otherwise. From this, we can rearrange the terms to derive the response equation which accounts for the disinhibition effect:

$$\mathbf{r}_i = (\mathbf{I} + \mathbf{W})^{-1}\mathbf{e}_i, \tag{6.5}$$

where $\mathbf{I}$ is the identity matrix.

By applying disinhibition, the response of the columns to the lines should shift a little depending on the strength of the response to each line. Thus, the final perceived line orientation γ can be obtained by finding the maximum response within each

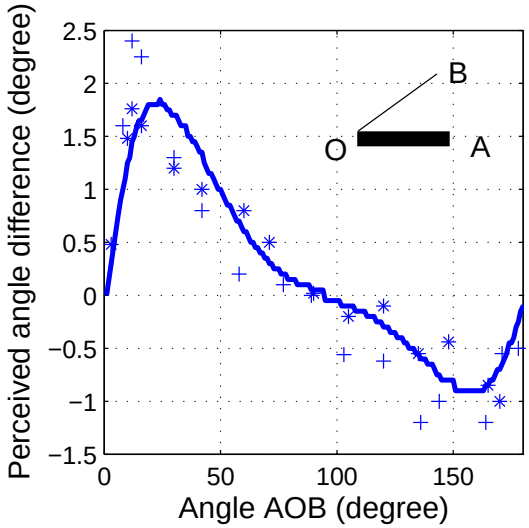


Fig. 31. **The variations of perceived angle between two intersecting lines.** The x-axis corresponds to the angle AOB (inset), from 0° to 180°. The y-axis is the difference between the perceived angle and the actual angle. The solid line is the result predicted by our model, and the data points $*$ and $+$ are data from human subjects in experiments by Blakemore et al. [49]. The curve was generated in two iterations with the following parameters: $\eta = 0.009$ and $\sigma = 1.0$ for the first pass; $\eta = 0.005$ and $\sigma = 0.5$ for the second. The other parameters remained the same for both iterations: $y_0 = 0.0$ and $k = 0.5$.

column after the inhibition process:

$$\gamma_i = \arg\max_{\alpha \in C} \mathbf{r}_i(\alpha), \tag{6.6}$$

where γ_i is the perceived orientation for the i-th line, $\mathbf{r}_i(\alpha)$ is the response of i-th orientation column's neuron tuned to orientation α, and C is the set of all the orientations within each column (from 0° to 180°) in layer 4 of the visual cortex.

D. Results

1. Experiment 1: Angle expansion without additional context

To test our computational model in the simplest stimulus configuration, we used stimuli consisting of one thick bar and one thin line. The thick bar was fixed at 0°, and the thin line was rotated to various orientations while the perceived angle was measured in the model. The enlargement effect of the angle varied depending on the orientation of the thin line. As shown in Figure 31, we can observe that there are three major characteristics of this varying effect. First, for the acute angles, there is an increase in the angle of the perceived compared to the actual angle, but for the obtuse angles, the perceived is less than the actual angle. Second, the peak is around 20° for the largest positive displacement, and around 160° for the largest negative displacement. Third, there is a clear asymmetry in the magnitude of the displacement between the acute angles and the obtuse angles: the peak at 20° is greater in magnitude than the dip at 160°. As compared in Figure 31, these computational results are consistent with results obtained in psychophysical experiment by Blakemore et al. [49]

2. Experiment 2: modified Poggendorff illusion

Disinhibition effect is the key observation leading to our extension to the angular expansion model, which is based on lateral inhibition alone. Because of disinhibition, when more than two lines or bars intersect, the perceived angle of the thin line will deviate from the case where only two lines or bars are present. The computational model is compliant with the human experiment. In the human experiment, the widths of thick bars were twenty times that of the thin line, so we kept the same ratio to address the thickness of the input lines in the model. The thick bars were assigned 40 units in width, while the thin line 2 units. The thickness of input lines can be

controlled by the constant a in equation 6.3. The offset y_0 in equation 6.3 was set to 0 in all experiments. Other parameters in equation 6.3 are free parameters, and the best fitting ones used in the resulting fit (as shown in Figure 33) were as follows: $y_0 = 0$, $\eta = 0.009$, $\sigma = 0.56$, and $k = 0.5$. The resulting curve is generated by two fixed inputs: thin line (width: 2 units) with fixed orientation at 30° and a bar (width: 40 units) with fixed orientation at 0° and a bar (width: 40 units) with a changing orientation from 0° to 180°.

Figure 32 shows the experiment where a 20° bar was added. Figures 32A-C show the initial activation of orientation columns to three lines (first thick bar is 0°, the thin line is 30°, and the second thick bar is at 20°). Figures 32D-F show the final response of the orientation columns after the disinhibition process. Note that the perceived thin line's orientations (the red line in Figure 32E inset) is slightly increased compared to that of the original input (green line in Figure 32E inset), but the perceived bars' orientation (the green line in Figure 32D and F inset) are barely affected. It is because the bars' input responses are much stronger than the thin line due to their thickness. Therefore, the proportion of change in the bars relative to their initial response is significantly smaller than that of the line, so the peak positions of the bars' responses are not changed after disinhibitory process. This experiment shows that the displacement of the peak positions before and after the disinhibition process can explain the amount of angular perception at a neuronal level. Using the model of disinhibition applied on orientation columns, the angular displacement can be estimated mathematically.

As shown in the model prediction results (blue curve, Figure 33), the effect demonstrated in Figure 28C is accurately predicted by the peak near 20°, and the effect in Figure 28D by the valley near 50°. In a similar manner, the model can predict the perceived angle when the angle between the thin line and horizontal bar is reduced.

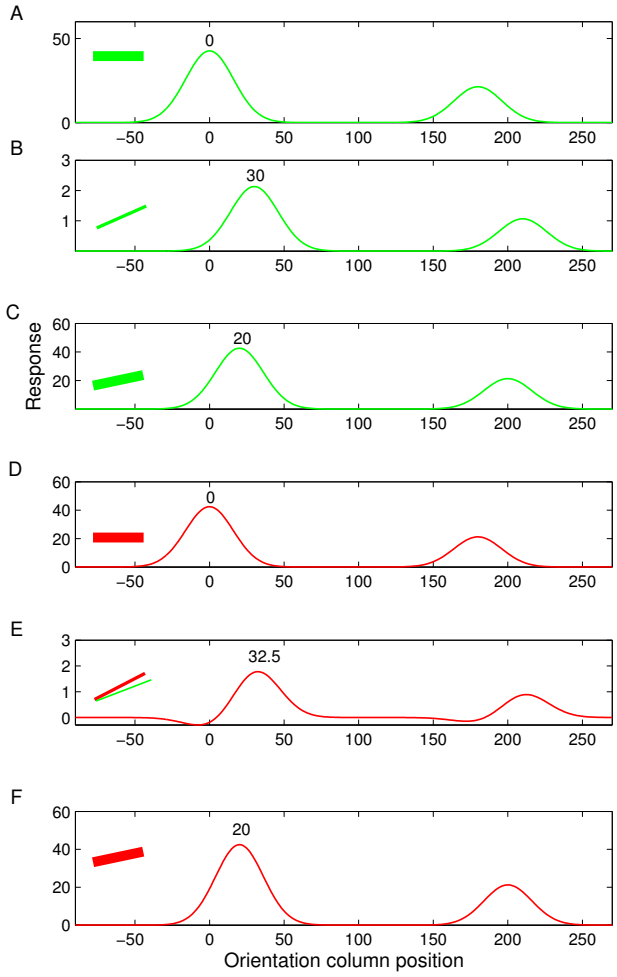


Fig. 32. **Initial orientation column activations (green) and final responses of orientation columns (red) after disinhibition.** **A** Initial excitation of the first bar. **B** Initial excitation of the thin line. **C** Initial excitation of the second thick bar. **D** Response to the first bar after disinhibition. **E** Response to the thin line after disinhibition. **F** Response to the second thick bar after disinhibition. This figure shows the perceived orientation of the thin line is enhanced (from 30° in **B** to 32.5° in **E**) through disinhibition that is introduced by the second thick bar at 20°. See text for details.

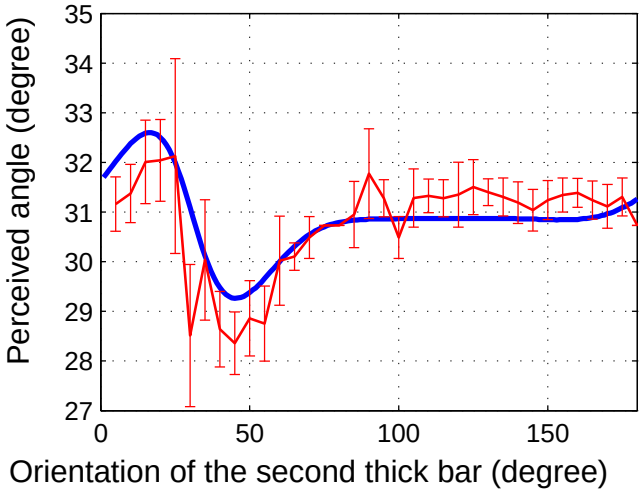
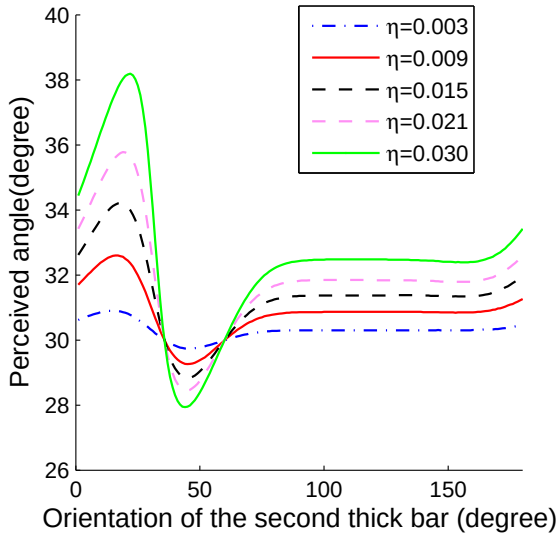


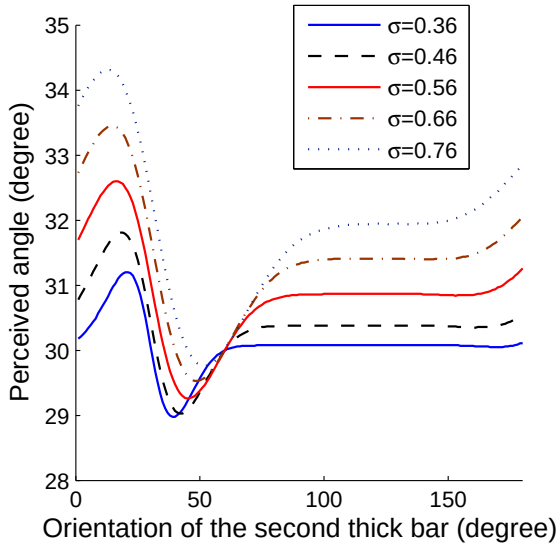
Fig. 33. **Perceived angle in a modified Poggendorff illusion.** The results from the computational model (blue line) and human experiments (red line with error bars representing the standard deviation of six samples on average) on a modified Poggendorff illusion (Figure 28C) are plotted. The second thick bar was rotated while the perceived angle was measured. The x-axis indicates the angle of the second bar. The y-axis shows the perceived angle of the thin 30° line. The model prediction and the human data are in close agreement. The parameters used in this experiment were as follows: $y_0 = 0$,$\eta = 0.009$,$\sigma = 0.56$, and $k = 0.5$.

So, at least for these two cases, we can say that our disinhibition-based explanation is accurate. However, does the explanation hold for an arbitrary orientation? To test this, we conducted a psychophysical experiment to measure human perceptual performance and compare the results to the model prediction (see the Methods section above for details). The human results are shown as a red curve with error bars in Figure 33.

The peak (near 20°) and valley (near 50°) in Figure 33 are apparent in the experimental data, and the overall shape of the curve closely agrees with the model prediction. The results show that our model of angular interaction based on disin-



A



B

Fig. 34. **Perceived angles under various values of η and σ.** (**A**) Perceived angles under various values of inhibition strength η ($\sigma = 0.56$ in all trials). (**B**) Perceived angles under various values of interaction width σ ($\eta = 0.009$ in all trials). See text for details.

hibition can accurately explain the modified Poggendorff illusion, and that low-level neurophysiology can provide us with insights into understanding the mechanisms underlying visual illusions with complex interactions.

The standard deviation of the Gaussian σ and the inhibition strength η are two free parameters that can be used for the curve fitting in Figure 33. The values of these two parameters are necessary in modulating the angles perceived from multiple lines, such as the sensitivity to the small angles, and the amount of distortion in orientation perception. Experiments with ferrets [50] showed that the strength of orientation tuning in the cortex can be changing during development, and therefore mature orientation cells will be both sensitive to the small angles and at the same time minimize the distortion. These parameters are tuned throughout development, and we can also test similar effects in our simulations. The two experiments as shown in Figures 34A and 34B tested different configurations of these two parameters in order to gain some insight into how those parameters can affect orientation perception.

Figure 34A shows how the inhibition strength η defines the magnitude of the curve predicted by the model. As η increases, the peak value of perceived angle becomes larger. In the tests of inhibition strength in Figure 34A, we held σ to a constant (0.56 in those trials), and in the final curve fitting in Figure 33, we picked a value of 0.009 for η as the best fitting one. Note that the locations of the peak and the valley do not change in this computational experiment.

Figure 34B shows that the standard deviation for the orientation column's activation profile defines the shape of the curve, for example, the positions of the peak and valley, or the direction of the tail in the curve. When σ is small, which means the Gaussian curve of the orientation column excitation profile is narrow, there would be less interactions across orientation columns. As a consequence, the cross-column inhibition will be limited only to a relatively short range. If σ is larger, the cross-

column inhibition can be effective in a wider range. Based on the peak and the valley positions of our experimental data, we chose $\sigma = 0.56$ as the appropriate value to fit the data.

The above observations suggest that the shape of the orientation column activation profile and cross-column inhibition strength could be the key factors which define human angular perception. Therefore, our model predicts that the effect of angular disinhibition will differ depending on which part of the visual field the stimulus is present, e.g. fovea vs. periphery, because the σ is larger in the periphery than in the fovea.

E. Discussion

The study of Poggendorff illusion has a long history. One existing explanation of the angle expansion phenomenon in Poggendorff illusion is that it is due to lateral inhibition between orientation cells [49; 46]. The explanation is also known as angular displacement theory [51]. Our model is an extension to the angular displacement theory. The angular displacement theory has been disputed (e.g., as pointed out by Robinson [52], and Howe et al. [53]) because it seems that it cannot explain the case where only the acute or the obtuse components are present. The Poggendorff illusion is apparently reduced when only acute angle components are present (Figure 35A), but it is maintained when only obtuse angle components are shown (figure 35B). One explanation to this puzzling phenomenon as proposed by Zarándy [51] was that there is a illusory shift or overestimation of the end position of the acute angles by endpoint detectors in the visual cortex (figure 35C), but the shift is not perceived with the obtuse angles. The apparent shift of the acute angles' tip positions (the points marked as A and B in figure 35D) move the edges along the directions as indicated by

the arrows in figure 35D. Therefore, when the two mechanisms of angular displacement and endpoint shifting are combined, the Poggendorff illusion is reduced, and thus the line components appear collinear (as shown in figure 35D, demonstrated by the solid line between the dashed lines). Their discovery suggested that the endpoint filter can neutralize the effects of angle expansion under special configurations, and that the angular displacement theory is still valid under usual configurations of the Poggendorff illusion. However, the illusory shifting mechanism by the cortical endpoint detector cells [51], which may account for the reduced illusory effect of acute angle components, is not included in our model.

There are several other existing theories explaining the Poggendorff illusion. For example, Gilliam proposed a depth processing theory, and suggested that the Poggendorff illusion was due to the bias from three dimensional perception [54; 55]. Morgan explained the illusion based on bias in the estimation of the orientation of virtual lines by second-stage filters [42]. On the other hand, Fermüller proposed that noise and uncertainty in the formation and processing of images caused a bias in perception of the line orientation [56]. Howe et al. explained the illusion based on natural scene geometry using statistics of natural images [53]. They showed that the location of a thin line segment across a thick bar in natural environments has the highest possibility away from the collinear point. The bias in the geometric perception in natural scene matched well with the shifting of the thin line in Poggendorff illusion. Indeed, the above theories successfully explain possible sources of bias formed in the Poggendorff illusion, but none of those provide an explanation of Poggendorff illusion at a neurophysiological level, nor did they explicitly model the disinhibitory effect as presented here.

The model we have presented here is based on angular inhibition which takes into account the disinhibition effect, and the soundness of the theoretical extension lies

in physiological and psychological facts. First, our model was based on the Limulus visual system. However, it is also known that disinhibition exists in the vertebrate visual system, such as in the visual cortical column of cats [57; 46; 47; 48], tiger salamanders [58] and in mice [59]. It is also known that the opposite directions of the same orientation evoke an asymmetric response [44]. Our model of the angle variations for acute and obtuse angles shows asymmetric properties and matches these experiments well. Second, our model can correctly replicate disinhibition caused by more than two lines intersecting and the results match our own experimental data obtained by the same kind of stimuli.

Besides the Poggendorff illusion, our model has the potential for explaining other geometric illusions, such as the café-wall illusion. Fermüller and Malm showed a variation of the café-wall illusion where adding some dots in strategic places significantly reduced the perceived distortion [56]. Such a correctional effect can potentially be explained by our model. Because the newly introduced dots give rise to a new orientation component (as the second thick bar did in our modified Poggendorff illusion), the disinhibitory effect caused by that new orientation can reduce the distortion formed by the existing orientation components.

F. Summary

In this chapter, a neurophysiologically based model of disinhibition to account for a modified version of the Poggendorff illusion was presented. The model was able to accurately predict a subtle orientation interaction effect, closely matching the psychophysical data we collected. We expect the model to be general enough to account for other kinds of geometrical illusions as well.

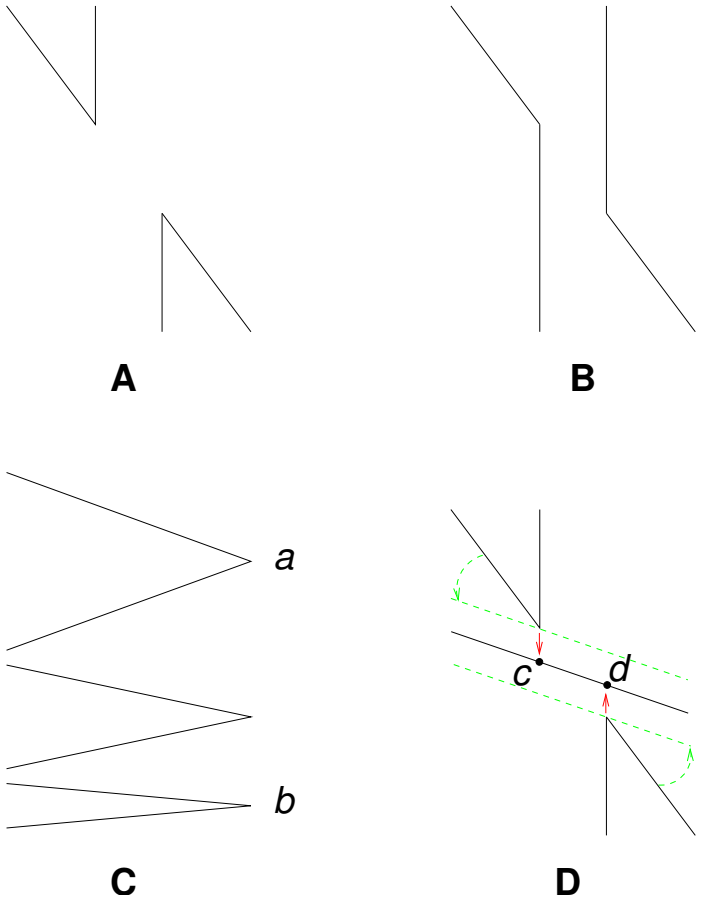


Fig. 35. **The endpoint effect in the Poggendorff illusion.** **A** The reduced Poggendorff illusion with only acute angle components is shown. **B** The illusion is not reduced with obtuse angle components. **C** The positions of the acute angle endpoint can be overestimated. For example, the point b appears to be on the right of point a, but actually they are on the same vertical line. Redrawn from [51]. **D** The combination of shift in endpoint and angle expansion mechanisms may explain why the illusion is reduced in **A**. The overestimated endpoints are labeled as c and d, and the expanded angles are illustrated by the green dashed line. (The amount of angle expansion by the dashed lines may not be accurate, because we just use them as a demonstration of the angle expansion effect.) If we shift the expanded angle edges (the green dashed line) to the overestimated endpoints c and d (shifted as the red arrows indicated), the line components appear collinear (as illustrated by the solid line in the middle of the two dashed line). Therefore, the overall illusory effect is reduced.

CHAPTER VII

ROLE OF DISINHIBITION IN ATTENTIONAL CONTROL

In this chapter, we will analyze the role of disinhibition in attentional control. I will first introduce the thalamocortical circuit in section A. The thalamocortical circuit is rich disinhibitory patterns, and it is believed to play a role in enhancing and suppressing sensory inputs to the cortex. Based on this function of thalamocortical circuit, I will propose a neural network model of input-modulation, and test it with the Stimulus Onset Asynchrony (SOA) effect in the Stroop task (section B) [60].

A. Disinhibition in the thalamocortical circuit

The thalamus is about the size of the end segment of the little finger and is located at the top of the brainstem in the interior region of the brain [61]. It consists of the relay nuclei and the thalamic reticular nucleus (TRN). The thalamic reticular nucleus is a collection of GABAergic neurons that form a shell surrounding the dorsal and the lateral aspects of the thalamus. Figure 36 is a schematic drawing of the thalamic connectivity. Sensory input is received by relay cells, and then relay cells forward the activity to pyramidal cells in the cortex and also to the reticular nucleus. The pyramidal cells have excitatory feedback to both the relay cells and the reticular nucleus. The TRN neurons send inhibition to each other as well as to the relay cells.

The function of the thalamus can be summarized as follows (see, e.g., [9; 62]). First, it transmits signals from the sensory periphery to the cortex. Second, it transfers signals from deep motor nuclei to the cortical motor centers. Third, it controls the signals to select which input and output will be permitted to pass to and from the cortex and how the signals will be sequenced. The thalamic reticular nucleus (TRN) plays a major role in such a signal selection function. Forth, it modulates (i.e. control

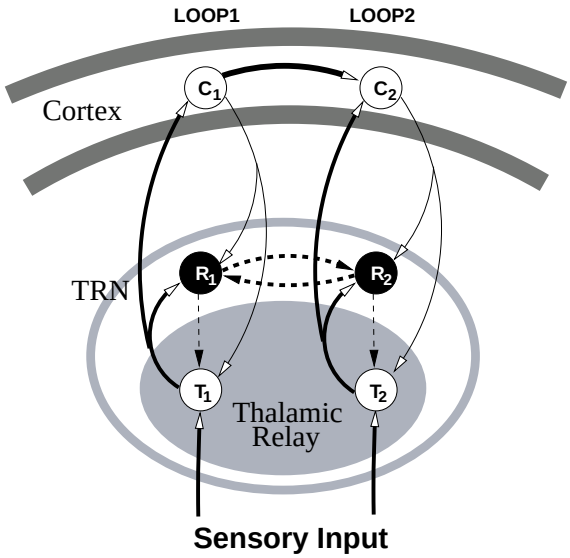


Fig. 36. **A schematic drawing of thalamus connectivity.** Adapted from [20]. T_1 and T_2: relay cells. R_1 and R_2: thalamic reticular nucleus cells. C_1 and C_2: cortical cells. See text for details. Filled circles indicate inhibitory, and open circles excitatory neurons.

the intensity) and synchronizes (for grouping) the signal transmissions.

In the following, I will first summarize the roles of disinhibition in the thalamus (section A.1) , then propose a biologically accurate model of the thalamocortical circuit to verify the role of disinhibition (section A.2).

1. Role of disinhibition in the thalamus

In the thalamocortical model (figure 36) we can observe that the TRN neurons play an important role in controlling and modulating the signal from the relay cells to the cortex. Note that the TRN is recurrently connected within itself, and their inhibition mechanism implements disinhibition in two senses: (1) One TRN cell, say R_1, inhibits another R_2, and R_2 inhibits the relay cell T_2 ($R_1 \rightarrow R_2 \rightarrow T_2$); and (2) the inhibition is mutual among TRN cells. Since I am mainly interested in the function

of disinhibition, it will be worthwhile to study how these disinhibition mechanisms can control and modulate the signals to and from the cortex. The first disinhibition pathway is to disambiguate between the input and the output of cortical computation when the output was partially input driven. In this case, the more input-driven cortical representation will be inhibited through the thalamus-TRN feedback loop [20]. As for the second case, disinhibition within the TRN can allow for inhibitory and excitatory effects to ripple through further than the immediate physical connectivity radius, thus allowing for a large-scale coordination of activity within the TRN. Moreover, recurrent disinhibition can enhance the signal contrast over space, so this structure could be used in modulating the intensity of the signals. Research on attention (e.g. [63]) pointed out that the TRN neuron can control and modulate specific, localized, active parts of the response of the thalamus to the environmental input. For example, O'Connor et al. found that lateral geniculate nucleus (LGN) activity was enhanced when subjects attended to the stimulus, and it was suppressed when they ignored it [64]. Furthermore, directed attention to a spatial location in anticipation of the stimulus onset led to an increase in baseline activity in the LGN. Such attention regulations can be a result of the control from the cortex to the TRN; and through the disinhibition mechanism of the TRN neurons, the signals from the relay cells transmitted to the cortical cells can be controlled and modulated.

2. Biologically accurate model of the thalamocortical circuit

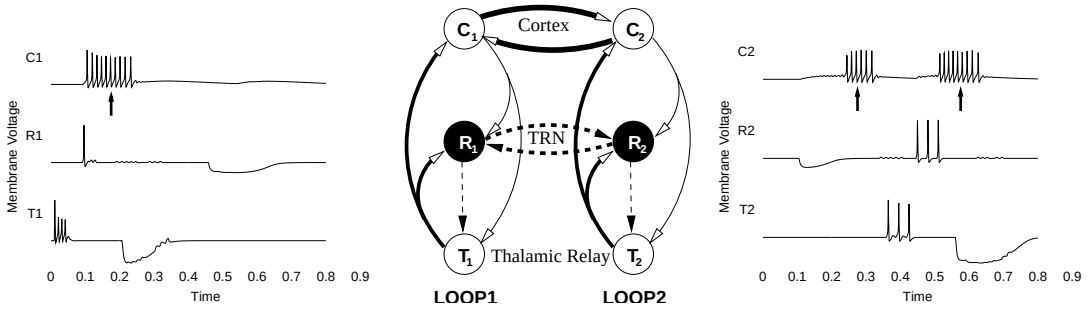
Choe employed integrate-and-fire neurons to simulate the thalamocortical circuit [20]. In Choe's experiment, the initial activation of the cortical cell driven by the input was suppressed, and only the cortex-driven cortical activity was able to reactivate the cortex through feedback to the thalamus. This kind of thalamocortical circuit behavior can also be simulated in a more biologically accurate way with Hodgkin-

Huxley neurons [65] (see results in figure 37 and 38). In figure 37, the input was only injected to the relay cell T_1. As a result the reticular cell (R_1) and the cortical cell (C_1) were both activated. The cortical cell C_2 was activated by the cortico-cortical connection from C_1. At this time point, the reticular cell R_2 was inhibited by R_1 and it allowed the reactivation of C_2 (the right arrow in figure 37C). This simple thalamocortical circuit demonstrates that the thalamus can control the propagation of activation from one cortical region to another while suppressing the originating region, even when the two regions are reciprocally connected. Figure 38 demonstrates a similar result with different level of input to the two loops (as shown in the figure, Loop_i is defined as the local circuit composed of T_i, R_i, and C_i). $Loop_1$ was injected with strong input (2.0) while $Loop_2$ with weak input (1.0). Due to disinhibition in the TRN (R_1 suppressing R_2), $Loop_2$ succeeds in reactivating the cortex C_2 (right arrow in figure 38C) even though it was input driven (i.e. weakly input-driven cortical activity gets promoted).

In this section, we reviewed the thalamocortical circuit. In the following sections, I will propose a neural network with disinhibition features. The proposed model can enhance or suppress the input through feedforward disinhibitory connection, and implement selective attention over time.

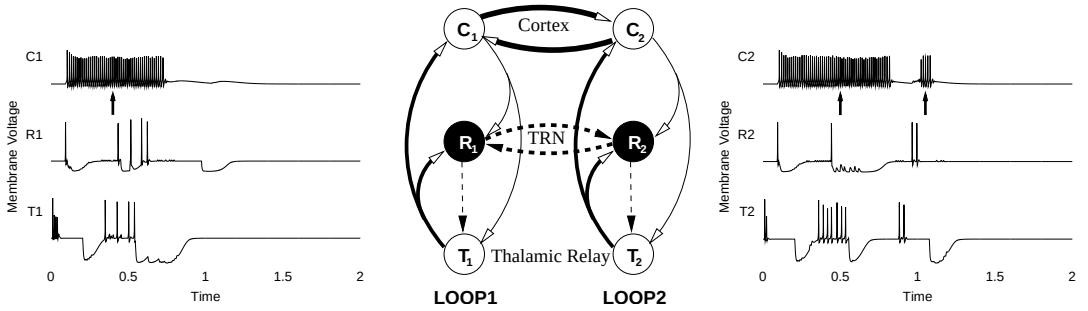
B. Selective attention over time

Selective attention refers to the competition between target resources (or relevant stimuli) and distracting resources (or irrelevant stimuli). As a result, the attended stimulus creates more reliable cortical activity than the unattended ones [66]. LaBerge [7; 8] explained selective attention as an enhancement of target site (the corresponding principal cells of a thalamic nucleus). Some other theories pointed out that selective



A *Loop*$_1$: Input=1.0 **B** **C** *Loop*$_2$: Input=0.0

Fig. 37. **Input vs. no-input condition.** The spike train for neurons in two connected loops are shown in **A** and **C**. The two loops are connected as shown in the middle **B**. **A**. A depolarizing current of duration 0.05 was injected in T_1. C_1 activates once (arrow) and goes silent. **C**. No current was injected anywhere in the loop, thus all activities were initially driven by the cortico-cortical connection from C_1 to C_2 at time $t = 0.3$. Only the cortex-driven cortical activity in C_2 (arrow on left) is able to reactivate the cortex through feedback to the thalamus ($t = 0.6$ in C, arrow on right). Adapted from [20].



A *Loop*$_1$: Input=2.0 **B** **C** *Loop*$_2$: Input=1.0

Fig. 38. **Strong vs. weak input condition.** A similar experiment as in fig 37 where both loops received input ($Loop_1 = 2.0$ and $Loop_2 = 1.0$). Due to disinhibition in the TRN (R_1 suppressing R_2), $Loop_2$ succeeds in reactivating the cortex (right arrow in C_2) even though it was input driven (i.e. weakly input-driven cortical activity gets promoted). Adapted from [20].

attention is object-based [67; 68; 69; 66] or space-based [70; 71; 66; 72]. These theories treated selective attention as a selection of What, Where, and Which [61]. Yet, selection in time has not been thoroughly investigated. In this section, we will focus on attentional control in the temporal domain: the selection of "When".

Attentional selection has been studied in visual search tasks (e.g. [73; 74]). In these tasks, target and distractor objects are presented simultaneously (figure 39A). Since the stimulus onset times are the same, there is no preference for a particular time period: no modulation is needed to magnify or reduce the signal during a particular time frame. However, if the target and the distractor are presented at a different time, temporal modulation may be needed if a certain time period is to be given preference (see e.g., [75]). As shown in figure 39B, C, and D, when a relevant stimulus and an irrelevant stimulus are presented asynchronously, the desired modulation is to enhance the signal during the relevant time frame, and reduce the signal during the irrelevant time frame. Here, we are particularly interested in the attentional control mechanisms that show modulation of input signals over time. Such an attentional control in the temporal domain can be seen as *the selection of "When"*, which is different from space-based or object-based attention [75]. The time-based modulation profile can be applied to all the stimuli, showing no preference over objects or locations. Time-based selection can provide an alternative explanation to the SOA effect in the Stroop task.

Stroop task [76] tests how humans respond to a compound stimulus where the color information conveyed by the printed words is incompatible with the ink color (i.e., *incongruent case*: for a comprehensive review, see [77]). In the color naming task, stimulus feature from one dimension (color) is a target, while that from another dimension (word) becomes a distractor. The control-condition cards were the same as the experimental cards except that the text was replaced with colored blocks. The results showed that there was a significant difference (almost twice) in response time

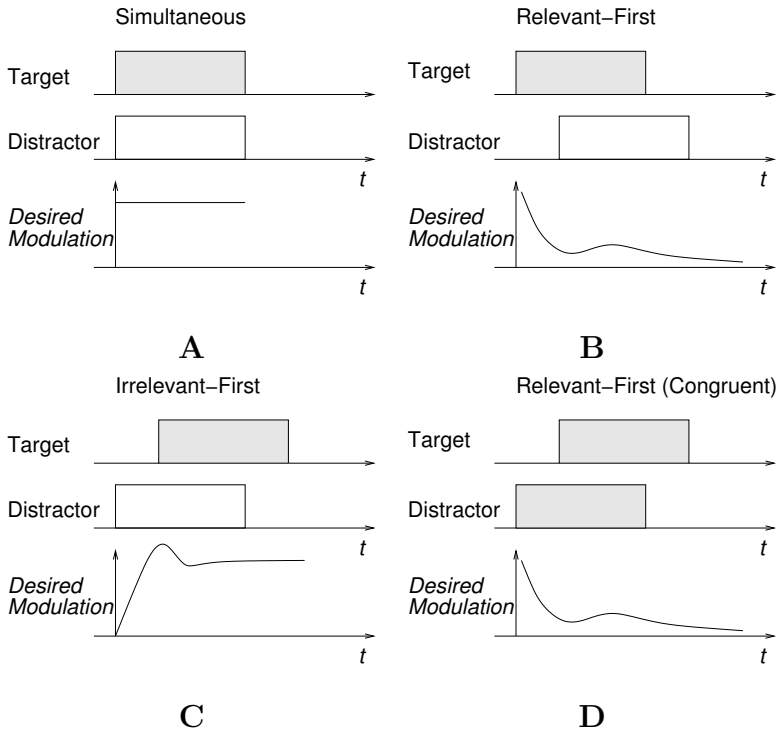


Fig. 39. **A schematic drawing of attentional control.** The target and the distractor are defined by the task (e.g. in color-naming Stroop task, the target is a color block and the distractor a word). Distractor can be congruent with the target (as in D: we mark the congruent stimuli pair in gray), or incongruent with the target (as in **A**-**C**: the target and distractor stimuli pair are colored in gray and white). If the stimulus can provide sufficient information (that does not necessarily have to be a target: e.g., a congruent distractor) to evoke a response, the stimulus is defined as relevant, and otherwise irrelevant. **A** Simultaneous onset of target and distractor: input-modulation over time is not needed. **B** Relevant-first: The target stimulus starts first, and the distractor (irrelevant) later. The desired modulation is to enhance the input in the early-stage and reduce in the later-stage. **C** Irrelevant-first: The distractor (irrelevant) stimulus starts first, and the target follows. The desired modulation is to reduce the input in the early-stage and enhance that in the later-stage. **D** Relevant-first (congruent): The distractor stimulus begins first, but unlike in case **C**, the distractor can be a relevant stimulus. The desired modulation is to enhance the input in the early-stage (with relevant-first profile) as in **B**.

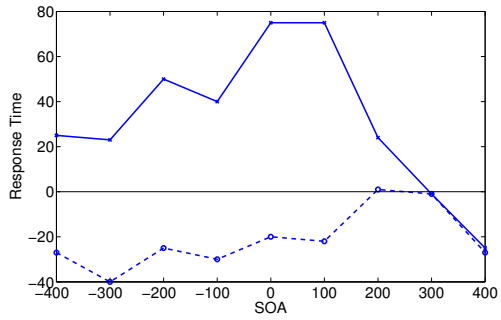
per item in the experimental case than in the control case [76].

Experiments on Stimulus Onset Asynchrony (SOA) investigated the time course of the Stoop effect [78; 79]. For example, Glaser and Glaser [79] presented words and colors with a set of target-first and distractor-first SOAs (figure 39B and C). In their configuration, the words were presented in white on a dark background, and the color in a colored block on the same background. The onset time of the word were 400, 300, 200, 100 or 0 ms before the time of color block presentation (distractor-first); or 0, 100, 200, 300, or 400 ms after the color block onset time (target-first).

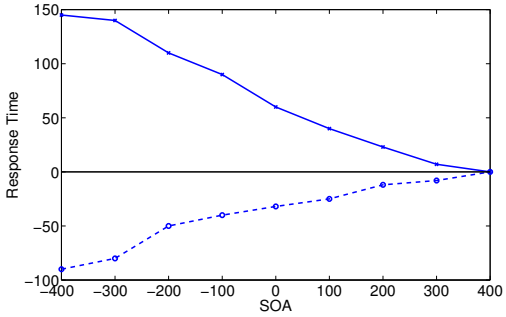
The results by Glaser and Glaser [79] indicated that the Stroop phenomenon was not caused by the relative speed of processing of word or color. Interestingly, as shown in figure 40A, the response time is shorter for the distractor-first task (incongruent case). However, neither models based on selection-through-accumulation [80; 81; 82] (figure 40B) nor selection-through-attraction [83] can explain the phenomenon (see [84] for a summary). What could be the mechanism underlying such a time-course property in the SOA effect?

Roelofs proposed the theory of selection-through-verification, which used a system named WEAVER++ to predict the SOA data [84]. Although WEAVER++ yielded better results than all previous models, this model was not a purely connectionist model (i.e., it was semi-rule-based). More importantly, it omitted the possibility of attentional control over time.

It is possible that attentional control over time can be learned: If the subject has experienced target (or the relevant cue) onset time that is always (or with a certain high probability) in a certain time offset from another stimulus onset, a neural process may adaptively adjust attention to enhance the relevant input or reduce the irrelevant input. From an attentional control perspective, we explore here, through attentional selection of "when", an alternative way to explain the SOA effect in the Stroop task.



A Experiment by Glaser and Glaser [79]



B Model [80; 81]

Fig. 40. **Human data and model result of SOA experiment.** In both **A** and **B**, the results are for the color-naming task in the Stroop effect. The x-axis is the stimulus onset time. The negative time is for distractor-first case, and the positive time for the target-first case. The y-axis is the response time of human compared to the control case. Throughout this chapter, the response time in the control case is used as a reference (solid line at $y = 0$). Therefore, positive response time means slower than the control, and negative means faster than the control. The solid lines are the response times of the incongruent case, while the dashed lines are those of the congruent case. **A** Human data [79]. Note that for the incongruent case, the peak of the curve is around time 0 and when the lag between the distractor and the target increases, the response time is reduced. (Redrawn from [79].) **B** Results from Cohen's model [80; 81]. Note that the incongruent cases (solid line) are not correctly predicted compared to human data. (Adapted from [84].)

The remainder of this chapter is organized as follows. The next section (section C) introduces the input-modulation model. Section D describes the experiments and results, followed by discussions in section E. We conclude with section F.

C. Model

1. Model architecture

The proposed model contains three modules as shown in figure 41. Module I is the attentional control module, which involves two parts. The first part functions as a temporal learner and it generates inhibition profiles to modulate the input over time through the attentional gateway (Module III). The second part plays the role of a conflict monitor, monitoring the conflict in the responses that arose in the processing modules. Module I outputs to Module III to control the input signal magnitude.

The second module employs the stimulus competition model, GRAIN, by Cohen and Huston [81]. There are two layers of neurons that are bidirectionally and recurrently connected as shown in figure 42. The processing network follows the GRAIN model's configuration as shown below:

$$\alpha_j(t) = \sum_i a_i(t) w_{ij} + e_j, \tag{7.1}$$

where t is time, α_j the post-synaptic potential of the j-th neuron, w_{ij} the synaptic weight as shown in the figure 42 (note that all the synapses are bi-directional, and $w_{ij} = w_{ji}$), and e_j the input from lower-level sensors. The pre-synaptic activity of neuron i is defined as:

$$a_i(t) = \sigma(\beta_i + \theta_i), \tag{7.2}$$

where σ is a sigmoid function (e.g. tanh in our experiments), θ the threshold (number inside the circles in figure 42, and 0 otherwise), β_i the running average of post-synaptic

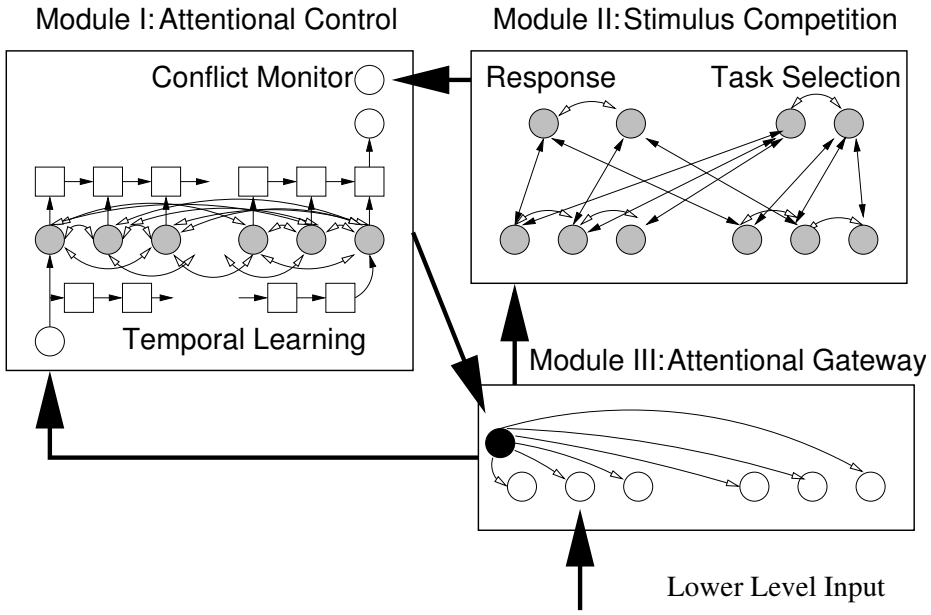


Fig. 41. **An overview of the input-modulation model.** There are three modules: Module I - Attentional Control Module, Module II - The Stimulus Competition Module, and Module III - Attentional Gateway. The circles represent neurons (black circle: inhibitory neuron, gray circle: neuron with both excitatory and inhibitory synapses, white circle: excitatory neuron), and the squares delay units. Filled arrows represent excitatory, and unfilled arrows inhibitory synapses. The thick lines with arrows are interconnections between the modules, representing multiple parallel connections. See text for details.

potential over time with an averaging rate τ:

$$\beta_i = \tau\alpha_i + (1-\tau)\beta_i. \tag{7.3}$$

The constant τ was 0.01 in all the experiments. The two response neurons marked "RESPONSE" send their outputs to the conflict monitor [85]. When the difference in the two outputs accumulates to reach a threshold (= 1 in our experiments), an output of the perceived color is announced and that time point is recorded as the response time (GRAIN model, [81]).

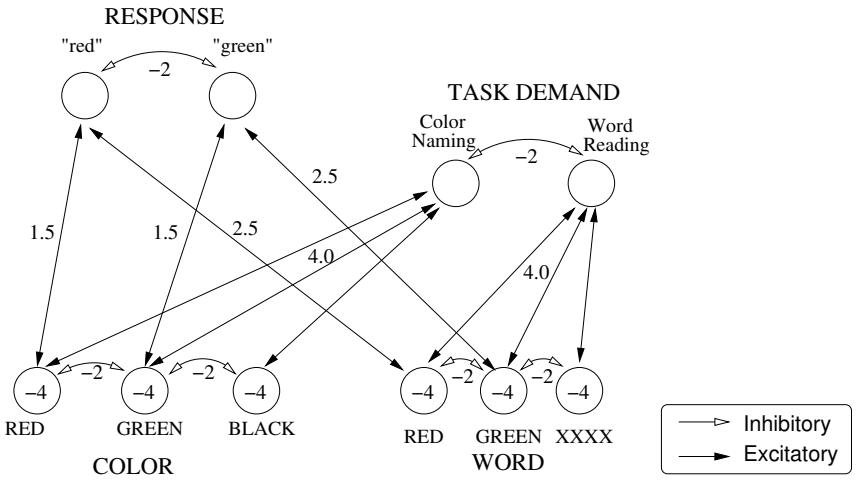


Fig. 42. **Module II: The stimulus competition model.** Two layers of neurons that are bidirectionally and recurrently connected are shown. The circles represent neurons. Filled arrows represent excitatory synapses, and unfilled arrows inhibitory synapses. Numbers in the circles represent the threshold of that neuron. (Redrawn from [81].)

The third module is an attentional gateway. It forwards the input from lower level visual pathway to module I and II. The inhibitory neuron tonically inhibits the sensory input to module II, and it reads the attentional control feedback from module I and regulates the input magnitude through an inhibitory neuron. When the output from Module I inhibit the inhibitory neuron in Module III, the sensory input is allowed to transfer to module II. The details of temporal control will be introduced in "Temporal Control Profile" subsection below.

This model architecture follows the "triangle circuit" theory proposed by LaBerge [61; 86]. The triangle circuit includes three aspects of attention: expression, enhancement mechanism, and control. The "expression aspect", as indicated by LaBerge, corresponds to the clusters of neurons in the posterior and the anterior cortex that serve cognitive functions. They map to the processing module (Module II) in our model. The "enhancement mechanism" maps to the thalamic nuclei as the atten-

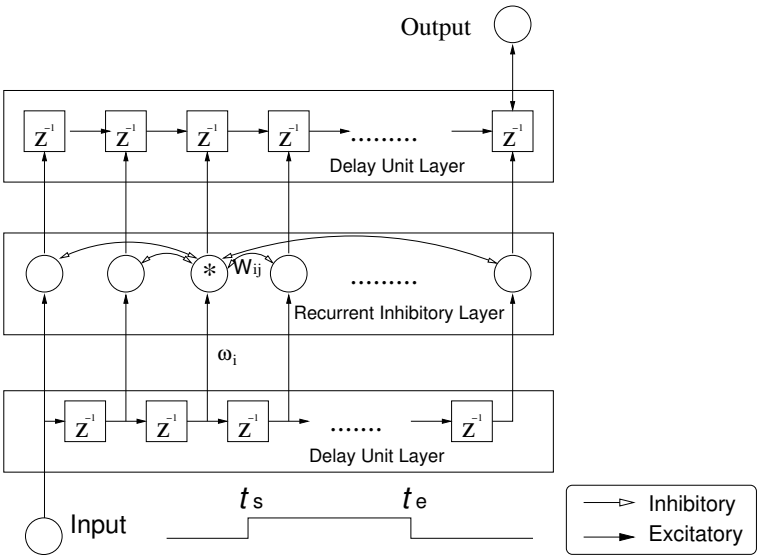


Fig. 43. **Module I: A solution for temporal learning.** There are three layers in this module: the bottom layer is a delay unit layer, which transfers the temporal sequence into spatial representations in the middle layer. The middle layer contains mutually inhibitory neurons. In the figure, only the connectivity of the neuron marked with a star is shown. The other neurons in this layer are similarly connected. The top layer is another delay unit layer, which replays the output from the middle layer into a temporal sequence. See text for more details.

tional gateway in our model (Module III). The "control" maps to the attentional control module (Module I).

2. Temporal control profile

The temporal input-modulation profile can be learned by a neural network as shown in figure 43. The circuit updates the connection weights for every instance of SOA stimulus. Therefore, the circuit becomes more accurate over time in predicting the onset time of a target stimulus. For example, let the stimulus (relevant or irrelevant) occur at time t_s and vanish at time t_e. (Note that the "relevant stimulus" does not

necessarily mean the "target", because in the congruent case, the "distractor" can also be relevant to the final response, and the brain may use this relevant information as a cue to predict.) The input neuron in figure 43 forwards the signal 1 for relevant stimulus or -1 for irrelevant stimulus during time period t_0 and t_e, and 0 at other times to the delay unit layer. The delay units in this layer then converts the temporal sequence into spatially distributed signals $\{s_0, s_1, ..., s_{n-1}\}$ and update the weights (ω_i, where i is the index of neuron) of the synapses between the feedforward delay unit layer and the recurrent inhibitory layer. For relevant stimulus, there will be an increase in the synaptic weight by a factor of γ (similarly for irrelevant stimulus, a decrease) in the synaptic weight. If there is no input, the weight remains the same. Therefore, the synaptic weights can be updated based on the input to the neuron in the recurrent inhibitory layer $i(t)$ as follows:

$$\omega_i(t+1) = \omega_i(t) + \gamma i(t) w_i(t). \tag{7.4}$$

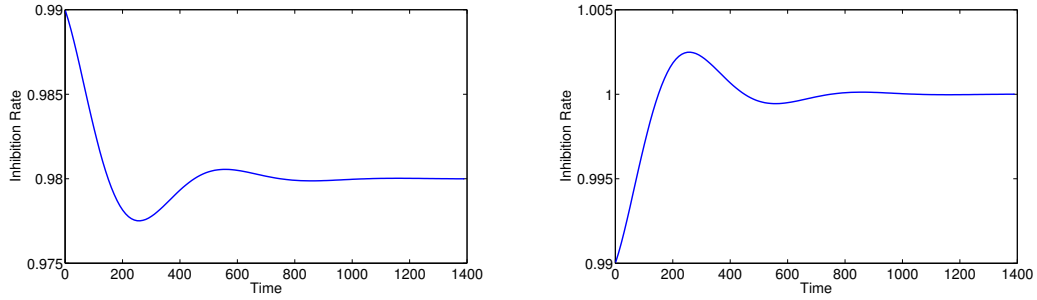
The synaptic weight W_{ij} from unit j to i in the middle layer employs a difference of Gaussian (DoG) neuronal interaction profile, which is defined as follows:

$$W_{ij} = G_{\sigma_c}(\mid i - j \mid) - G_{\sigma_s}(\mid i - j \mid), \tag{7.5}$$

where σ_c and σ_s are standard deviations for the center and the surround Gaussians, and the function G_σ is a Gaussian function with mean at zero and standard deviation of σ ($G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$). The Gaussian kernel in the inhibitory layer can smooth and increase the contrast. If we treat the input sequence $\{s_0, s_1, ..., s_{n-1}\}$ as a vector $\mathbf{s}$, the output $\mathbf{r}$ of the middle layer can be obtained by the equation below [18]:

$$\mathbf{r} = (\mathbf{I} - \mathbf{W})^{-1}\mathbf{s} \tag{7.6}$$

where $\mathbf{I}$ is an identity matrix, and $\mathbf{W}$ the weight matrix defined in 7.5.



A Relevant-first task **B** Irrelevant-first task

Fig. 44. **Temporal control profiles.** **A** Relevant-first case ($k = 0.01$, $\tau = 200$, $\omega = \pi/300$, $\eta_0 = 1 - 2k$). **B** Irrelevant-first case ($k = -0.03$, $\tau = 200$, $\omega = \pi/300$, $\eta_0 = 1$).

The third layer converts the spatial sequence **r** back into temporal control sequence $\eta(t)$ at the output neuron through a series of delay units. The output neuron disinhibits the sensory input in Module III. This way, the input signal gets modulated according to the temporal input-modulation profile.

In our experiments, for practical reasons, the shape of the temporal control sequence $\eta(t)$ at the output neuron (for both relevant-first and irrelevant-first cases) were approximated by $\eta(t) = \eta_0 + ke^{-t/\tau}\cos(\omega t)$, where η_0, k, τ, and ω are free parameters to control the shape of the profile. The parameter η_0 defines the baseline of the inhibitory profile, k the relevancy ($+1$: distractor-first; -1: target-first), and τ and ω the decay temporal factors. The inhibitory modulation rate $\eta(t)$ is plotted in figure 44.

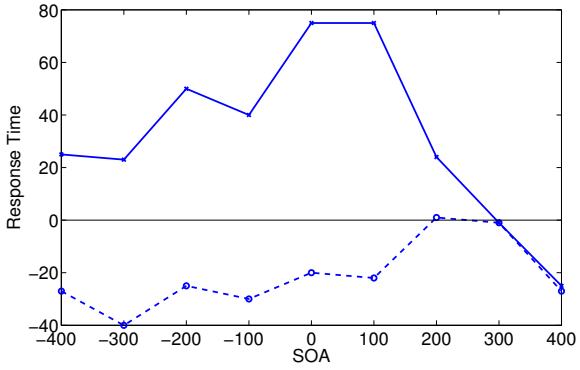
The learning site of the temporal attention control can be in the hippocampus, the basal ganglia, and the prefrontal cortex [87]. In the next section, we will show that the response time can be significantly changed by different temporal input-modulation profiles.

D. Experiments and results

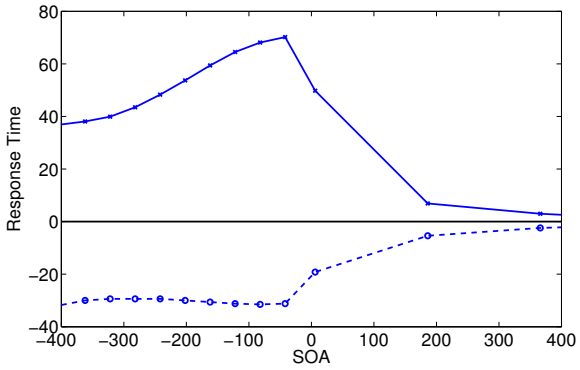
1. Experiment 1: Using irrelevant control profile for distractor

In the first experiment by Glaser and Glaser [79], they used a set of 48 SOA cases (with 1/3 congruent cases) which were randomly ordered. Due to the low rate of congruent cases, the distractor was more likely to be irrelevant to the response. For the simulation of this experiment, we employed the temporal input-modulation profile shown in figure 44A (relevant-first) for the target-first task, and that in figure 44B (irrelevant-first) for the distractor-first task. The model predictions and human data in color naming task are compared in figure 45. Similar to the human data, the response time of the incongruent case predicted by the model has a peak at around 0 ms SOA, and decreases in both positive and negative directions. In contrast, the model results by Cohen and Huston [80; 81] (figure 40B) only decreased monotonically toward positive SOA, and it achieves maximum response time at the negative end of SOA (at -400ms). In the congruent case, our model correctly predicted that the response time is below the control baseline ($y = 0$). The approximate response time of the model (solid curve below the control baseline) matches well with the human data (the dashed curve below the control baseline).

These results indicate that, through temporal attentional control, the response time is reduced when the stimuli have a longer lag between their onset time. Therefore, when two stimuli occur more separately over time (for both positive SOA and negative SOA), neural processes can discriminate the two by reducing the input magnitude during the presentation of the distractors.



A Experiment [79]



B Model

Fig. 45. **Result of experiment 1: Human data and model result of SOA experiment.** In both **A** and **B**, the results are for the color-naming task in the Stroop effect. The solid lines are incongruent case, and the dashed lines congruent case. **A** Human data by Glaser and Glaser [79]. Note that for the incongruent case, the peak of the curve is around time 0 and when the lag between distractor and target increases, the response time is reduced. (Adapted from [84].) **B** Model prediction. This result indicates that through temporal attentional control, the response time is reduced when the stimuli have a longer lag between their onset time points. The model response time is scaled by 0.3.

2. Experiment 2: Using relevant control profile for distractor

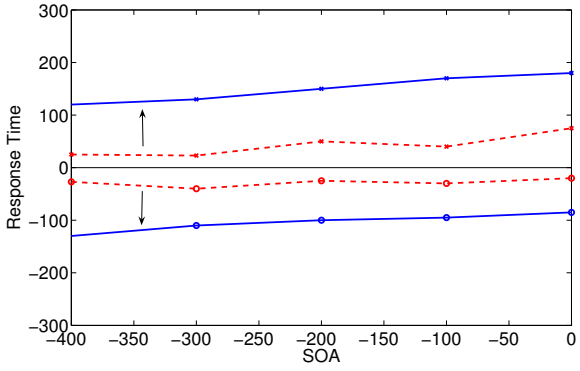
In the previous experiment, we used irrelevant-first input-modulation profile as shown in figure 44B for distractor-first SOA cases, and thus the distractor is treated as irrelevant to the response. Now we are interested in knowing how the choice of irrelevant-first or relevant-first profiles can affect the response time. In this experiment, we will apply relevent-first modulation profile for the distractor-first SOA cases.

As shown by the results of the model, the response time of the incongruent case (the solid curve above line $y = 0$ in figure 46B) takes longer than the result in experiment 1 (the dashed curve above line $y = 0$). The response time of the congruent case (the solid curve below line $y = 0$ in figure 46B) is shorter than the result in experiment 1 (the dashed curve below line $y = 0$). A comparable human experiment of "relevant distractor" in SOA was done by Glaser and Glaser [79]. They did an experiment with a 80% probability of congruent cases. Since there was high probability of congruent cases, in the distractor-first case, the distractor may not be a real "distractor" but more likely a cue for the "target". In figure 46, the changes in response time in those two experiments are marked by the arrows in both the human and the model data. The human data show that the subjects appeared to have reduced their response time for all the incongruent cases when compared to the results in the first experiment.

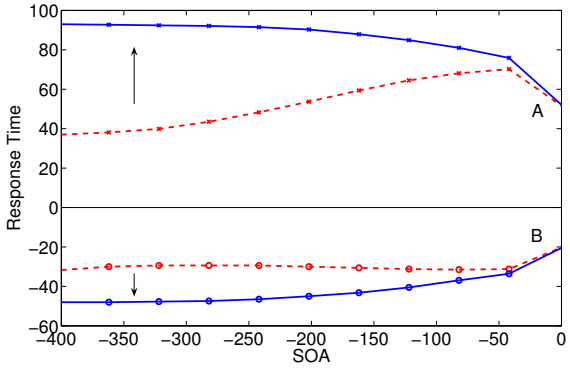
How can the response time of the congruent case become shorter while in the incongruent case it becomes longer? That is due to the increased probability of the congruent stimulus presenting, where the brain used the earlier appearing "distractor" as a cue and assign more confidence on the corresponding target onset time than that in the first experiment. However, in the incongruent case, the distractor stimulus is no longer a cue and becomes irrelevant. The brain mistakenly takes the irrelevant cue

at the wrong moment. As a result, it takes longer for the brain to give out a correct response. Thus the response time for the incongruent case is significantly increased. In this experiment, we simulated the process of "relevant" distractor-first case by using the "first-relevant" profile as shown in figure 29A. The model results are shown in figure 46B.

The model results match the human data pretty well, however, a discrepancy arises at SOA time 0 (marked "A" and "B" in figure 46B): the solid curve (data from experiment 2, high probability of congruent case) converges with the dashed curve (data from experiment 1, low probability of congruent case). This is because our model is not designed to handle probability of congruent case at the current stage. What the model demonstrated is how the relevancy profile of the distractor can affect the response time. Approximately, the higher probability of congruent cases in distractor-first task, the temporal input-modulation profile has a peak at an earlier position over time. Moreover, when the time lag decreases to zero, the temporal input-modulation profile does not make any difference to both of the stimuli, and therefore in such a case the whole system degrades to Module II (figure 42) which has no ability to handle the change of the probability of the congruent cases. To address the learning mechanism of non-temporal factors, which is purely contributed by the distributions of the training cases, more research may be needed. For our current model, it can demonstrate that the shift of attention input-modulation profile in time domain can affect the level of conflict between different stimuli, and when the brain pay more attention to the time period of the relevant stimulus onset, the overall response time is reduced. As a side effect, the response time in the incongruent case becomes longer due to the incorrect selection of time during which the stimulus is irrelevant.



A Experiment [79]



B Model

Fig. 46. **Result of experiment 2: Using relevant control profile for distractor.** **A** Human Data. (Redrawn from [79].) **B** Model Data. Similar to human data, the incongruent case's response time is increased than in the first experiment, while for the congruent case it is reduced. The model response time is scaled by 0.3. In both **A** and **B**, the solid lines are results of experiment 2 with high probability of congruent cases. The dashed lines are from experiment 1 with low probability of congruent cases. The curves above line $y = 0$ are for incongruent cases, while those below line $y = 0$ congruent cases. Near SOA$= 0$, the model shows a slightly different behavior compared to human's result. This is because our model is not modulated by the probability of congruent case. See text for details.

E. Discussion

Time-based selective attention is different from space-based or object-based attention control mechanisms. For object-based attention, the control mechanism gives preference to one of the simultaneously presented objects. Similarly, space-based attention is focus on the location of interest. Both the space-based and the object-based attention are spatial control mechanisms, and the selection does not have a preference over time [75]. In contrast, time-based attentional control gives preference in the temporal domain regardless of the stimuli's spatial locations. Unlike temporal attention defined in [75], our definition of time-based attention is an input-modulation mechanisms and it does not require that multiple items are presented at "the same location". The control mechanism adjusts the input magnitude of stimuli at all possible locations according to the temporal relevancy of the stimulus, e.g. the occurrence of the congruent stimulus, to the overall response to color in the Stroop task. Thus, the SOA in Stroop effect may not be due to the temporal modulation at the postperceptual level in visual short-term memory (cf. [88; 89]). It is simply due to input modulation, which is an ability to predict relevant stimulus onset time in order to reduce the irrelevant inputs. The temporal input-modulation profile can be given a priori from a higher level cognitive module, or obtained through reinforcement learning. The high peak in the temporal input-modulation profile means the "expected" moments and the input should be enhanced in such a case; the low valley can be interpreted as "noisy" moments and the input should be inhibited. Through sampling the stimulus and monitoring the internal response, the "expected" and the "noisy" time frames can be adaptively learned. Therefore the brain can use less computational resources for faster and more accurate responses for repeated tasks.

The time-based selection is neither an early-selection (e.g., filter theory by [90])

nor a late-selection (e.g., [91]). The brain not only controls the inputs at an early-stage as in Module III, but also evaluates and learns at a late-stage as in Module I. The evaluation of response relevancy must consider the relationship between high-level motor responses and the low-level sensory input. In this respect, our understanding of selective attention is similar to the "enhancement of target site" theory by LaBerge [7], where ours is "enhancement of stimulus in a relevant time frame".

One limitation of our current model is that it does not evaluate whether a stimulus is relevant or irrelevant. At the current stage, the relevancy of the input to the response was a given (1 for relevant, -1 for irrelevant input). However, this functional block of relevancy evaluation can be extended by checking if a stimulus is sufficient to evoke a correct response, and if so, it can be labeled as a relevant stimulus. For example, in congruent case of color-naming task, a first appearing word "red" that is followed by a red block is sufficient to invoke a verbal response "red", therefore, it is a relevant stimulus and the association can be learned.

The idea of selective attention over time can be verified through functional magnetic resonance imaging (fMRI) experiments (similar to the experiments for space-based attention [73] or object-based attention [66]). For example, we can design experiments to show subjects a sequence of words on a computer screen, and at some random time point play a bell sound. The subjects are to remember the words only at the point when the bell is heard. If the onset time of sound has a narrow distribution over time, it is expected that the brain activities of the thalamus and certain brain areas (e.g. prefrontal cortex or basal ganglia) will increase in a short period preceding the sound onset time. If so, it demonstrates that through learning, the brain has come to "expect" or "prefer" a certain time frame, i.e. selective attention of "when".

F. Summary

In this chapter, I discussed the role of disinhibition in attentional control. In the thalamocortical circuit, feedforward disinhibition (cortex-TRN-thalamus pathway) can implement temporal modulation while recurrent disinhibition (TRN-TRN) can implement spatial modulation. I further verified temporal modulation by designing a model of time-based attentional control and tested it with the SOA effect in the Stroop task.

In the model, the attention mechanism involves the thalamocortical circuit and other brains areas (e.g. prefrontal cortex, hippocampus, or basal ganglia) that carry out learning and higher level cognitive functions (e.g. defining the task and comparing the relevancy of input stimulus). The control can be realized by an internally learned temporal input-modulation profile, and inhibitions to the low-level sensory inputs. Although the idea of selective attention over time and the model suggested in this chapter awaits both more theoretical and empirical confirmations, it expanded the concept of selective attention into the temporal domain: the selection of "When". Time-based attentional control can enhance the overall system performance.

CHAPTER VIII

ROLE OF DISINHIBITION FROM A SYSTEM PERSPECTIVE

The previous chapters discussed the role of disinhibition in various brain organizations. In this chapter I will discuss the role of disinhibition from a dynamical system perspective. I will abstract the thalamocortical circuit introduced in Chapter VII to an interconnected Cohen-Grossberg network model to derive a set of sufficient conditions to ensure the stability of such layered circuits. I will continue on to analyze the controllability and computability within the disinhibition network structure, in sections C and D.

A. Stability analysis of thalamocortical circuit

Choe [20] showed that the thalamocortical circuit may be involved in the processing of analogy, where the results of the process are promoted through the cortico-thalamocortical loop. For this to work reliably, enduring oscillations need to be avoided. However, ensuring that oscillation does not occur in such a highly recurrently connected circuit with various cell-types is a non-trivial task. Brute-force search for the parameter may be infeasible due to the high dimensionality of the parameter space. Here, we adopted the Cohen-Grossberg (C-G) theorem to derive conditions that allow asymptotic stability in the thalamus-TRN-cortex circuit. The original C-G theorem requires that all connections are bidirectional and symmetric, thus it cannot be applied to the thalamocortical circuit in its original form. However, if the cortex, the TRN, and the thalamic relays are each treated as one instance of a C-G network, then the C-G theorem can be used to derive the conditions for stability by treating the whole network as interconnected C-G networks. In this section, we will provide a set of sufficient conditions for such an interconnected system to be asymptotically stable.

This allows us to greatly reduce the range of parameters to search. The framework for treating networks containing asymmetric connections as interconnected symmetric networks can also be of general interest to theorists studying stability in neural circuits.

As illustrated by Choe [20; 65], the initial activation of the cortical cell driven by input will be suppressed by the second iteration through cortical activation (see figure 38). The thalamocortical circuit can be abstracted as interconnected Cohen-Grossberg neural networks (CGNN). See [92; 93] and [94] for details about the CGNN. Note that Cohen-Grossberg neural network only allows symmetric synaptic connections, but the model of interconnected CGNNs (see, e.g., figure 47) allows asymmetric connections among CGNN networks. For example, the thalamocortical circuit can be seen as three local CGNNs ($CGNN_1$ as the relay-cell layer, $CGNN_2$ as the TRN layer, and $CGNN_3$ as the cortical layer), and these three layers are interconnected by a vertical column as shown in figure 47.

1. Interconnected Cohen-Grossberg neural networks

The Cohen-Grossberg neural network [92; 93; 94] is described by the following ordinary differential equation:

$$\dot{x}_i(t) = -a_i(x_i(t)) \left[b_i(x_i(t)) - \sum_{j=1}^{n} w_{ij}\sigma_j(x_j(t)) + I_i \right], \tag{8.1}$$

where I_i $(i = 1, 2, ..., n)$ denotes the external input, w_{ij} the connection weight, functions $a_i(x_i)$ and $b_i(x_i)$ the amplification functions, σ_j the activation function, C_i the membrane capacitance constant, and R_i the neuron resistance constant. The Cohen-Grossberg neural network has the following properties [92]:

- Existence of Lyapunov function (see equation 8.15).

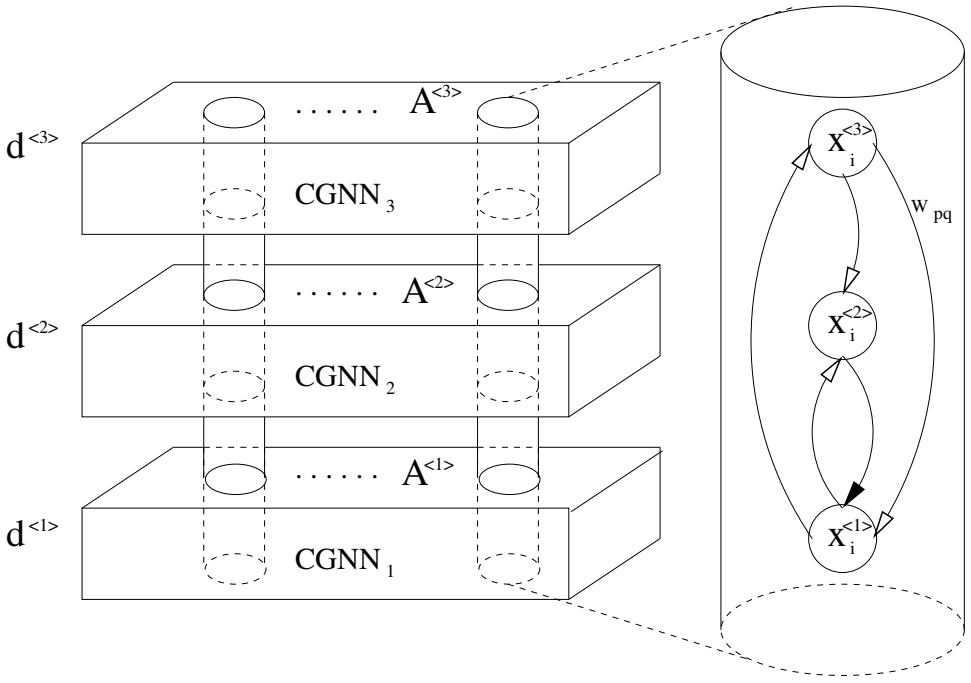


Fig. 47. **An illustration of multiple CGNNs with column-wise connection.** An interconnection of CGNNs is demonstrated on the left. The figure on the right shows the intracolumnar connections of the three CGNN modules. The constant $d^{<p>}$ is a positive number representing the weight of the Lyapunov function for each CGNN module p, $A^{<p>}$ the connection weight matrix inside each CGNN module, w_{pq} the inter-module connection weight, and $x_i^{<p>}$ the neuron activity in module p. See section B.5 for detailed description and dynamic equations of this system.

- Existence of equilibrium.
- Under some sufficient conditions (i.e., the activation function is monotonic and differentiable, and the connection weight matrix is symmetric), the system is asymptotic stable (also known as global pattern formation).

For neurons in the thalamocortical circuit, assuming the input is sustained between time 0 and t_0, and using tanh as the activation function, the circuit can be

described by the following ordinary differential equations:

$$\dot{x}_i(t) = \frac{1}{C_i}\left[-\frac{x_i}{R_i} + \sum_{j=1}^{n} w_{ij}\tanh(x_j(t-\delta_{ij})) + I_i\right], \quad \text{when } 0 \le t \le t_0 \tag{8.2}$$

$$\dot{x}_i(t) = \frac{1}{C_i}\left[-\frac{x_i}{R_i} + \sum_{j=1}^{n} w_{ij}\tanh(x_j(t-\delta_{ij}))\right], \quad \text{when } t > t_0 \tag{8.3}$$

where I_i $(i = 1, 2, ..., n)$ denote the external input that is sustained between time 0 and t_0, w_{ij} the connection weight, and δ_{ij} the transmission delay.

2. Dynamics of interconnected CGNN

We will consider the asymptotic stability of interconnected CGNNs composed by m CGNN components (or m modules), where each module contains n neurons. The derivation procedure treats each CGNN as an individual module, and then finds the Lyapunov function for such an interconnected system as a whole (for a review, see [95], pp. 358-361). For example, the thalamocortical circuit can be described by the following ordinary differential equation (when $t > t_0$):

$$\dot{x}_i^{<p>} = \frac{1}{C^{<p>}}\left[-\frac{x_i^{<p>}(t)}{R^{<p>}} + H(x^{<p>}, t) + V(x_i^{<\cdot>}, t)\right], \tag{8.4}$$

where x is the state variable representing the membrane potential, function $H(x^{<p>}, t)$ the horizontal connection,

$$H(x^{<p>}, t) = \sum_{j=1}^{n} A_{ij}^{<p>}\tanh(x_j^{<p>}(t - \delta_{ij}^{<p>})),$$

and $V(x_i^{<\cdot>}, t)$ the vertical connection,

$$V(x_i^{<\cdot>}, t) = \sum_{q}^{m} w_{qp}\tanh(x_i^{<q>}(t - \tau_{qp})).$$

The indices p and q are the indices of the CGNN modules (or layers); i, j the indices of neurons within each module; and $C^{<p>}$ and $R^{<p>}$ the membrane capacitance and

resistance constants for neurons in module p. Matrix $\mathbf{A}^{<p>}$ is the horizontal connection matrix within CGNN module p, and w_{qp} an element in the vertical connection matrix (across CGNN modules). The value $\delta_{ij}^{<p>}$ is the horizontal connection delay within CGNN module p, and τ_{qp} the vertical connection delay between module p and q, and $I_i^{<p>}$ the external input to neuron i in module p.

The interconnected CGNN is expected to have the following properties:

- Existence of Lyapunov function. The Lyapunov function is in the form of a weighted sum of the Lyapunov function of each CGNN modules.
- Existence of equilibrium.
- The connection weight matrix is symmetric in each CGNN module.
- Under some sufficient conditions, the system is asymptotic stable (which will be demonstrated as the main result of this section).

3. Assumptions

For simplicity of analysis, we assume that the following conditions (A1 to A3) are satisfied in the network.

- A1. The connection matrix $\mathbf{A}$ within each CGNN module is symmetric, and defined as

$$A_{ij}^{<p>} = A_{ji}^{<p>} = r^{<p>}, \text{ if i} \neq \text{j},$$

and

$$A_{ij}^{<p>} = 0, \text{ if i} = \text{j}.$$

- A2. Each of the CGNNs has n neurons, and they are interconnected by the same connectivity structure in each column, i.e. only neurons with same index

i across the modules are interconnected. The inter-module connection weight matrix is $\mathbf{W}$.

4. Existence of equilibrium point

To show a system is asymptotic stable, we have to prove that the system has at least one equilibrium point. For the interconnected Cohen-Grossberg Neural Network, we can claim that for every input I, there exists an equilibrium point for the system defined by equation 8.4. The proof is as follows.

The system in equation 8.4 can be rewritten into the form shown below if we assign each neuron a unique index:

$$\dot{x}_i(t) = -\frac{1}{C_i}\left[\frac{x_i(t)}{R_i} - \sum_{j=1}^{n} B_{ij}\tanh(x_j(t-\tau_{ij}))\right], \tag{8.5}$$

where B_{ij} is the corresponding connection weight. The system in equation 8.5 can be treated as a delayed CGNN, following [96]:

$$\dot{x}(t) = -a_i(x_i)\left[b_i(x_i) - \sum_{j=1}^{n} B_{ij}s_j(x_j(t-\tau_{ij}))\right], \tag{8.6}$$

if we let

$$a_i(x_i) = \frac{1}{C_i},$$
$$b_i(x_i) = \frac{x_i(t)}{R_i},$$

and

$$s_j(x) = \tanh(x).$$

Therefore,

- a_i is bounded, positive, and locally Lipschitz continuous; and
- b_i and b_i^{-1} are locally Lipschitz continuous;

- s_j is bounded (by -1 and 1) and Lipschitz continuous.

These conditions satisfy the assumptions of the equilibrium theorem in [96], and thus we can safely claim that the system 8.6 (or equivalently, system 8.5) has an equilibrium point. Because system 8.5 is just a rewrite of system 8.4 based on different neuron index, there must exist an equilibrium point for system 8.5 as well. Thus, the proof is complete.

5. Lyapunov function for interconnected systems in general

The system defined by equation 8.4 can be generalized as interconnected systems. Khalil proposed a method to derive the Lyapunov function of interconnected systems in general [95]. The steps proposed by Khalil [95] can be summarized as follows.

For each system module, we assume that its dynamics is as follows:

$$\dot{x}^{<p>} = f^{<p>}(x^{<p>}, t), \tag{8.7}$$

where function f defines the dynamic of the p-th module (CGNN), and the interconnection between modules (the uniform column connection) is described by a function $g^{<p>}(x, t)$. Then, the whole system can be written as

$$\dot{x}^{<p>} = f^{<p>}(x^{<p>}, t) + g^{<p>}(x, t). \tag{8.8}$$

When each module's Lyapunov function is known ($\Lambda^{<p>}$), it is reasonable to consider the following function as the Lyapunov function for the interconnected system.

$$V(x) = \sum_{p=1}^{m} d^{<p>} \Lambda^{<p>},$$

where $d^{<p>}$ is some positive number. Since $\Lambda^{<p>}$ is the Lyapunov function of module p, it is positive definite. Hence, their weighted sum $V(x)$ is also positive definite.

Therefore, the derivative of V can be derived as:

$$\dot{V}(x) = \sum_{p=1}^{m} d^{<p>} \left[\frac{\partial \Lambda^{<p>}}{\partial x^{<p>}} f^{<p>}(x^{<p>}) \right] + \sum_{p=1}^{m} d^{<p>} \sum_{i=1}^{n} \left(\frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}} g_i^{<p>}(x_i^{<\cdot>}, t) \right). \quad (8.9)$$

Once the $\dot{V}(x)$ is proved to be less than zero, we can assert that the interconnected system is stable. In the following, we will apply the procedure by Khalil [95] to the interconnected CGNNs, and derive the conditions of asymptotic stability for the case of no delay.

6. Asymptotic stability of interconnected CGNN

The interconnected CGNNs (8.4) without delay is defined as:

$$\dot{x}_i^{<p>} = \frac{1}{C^{<p>}} \left[-\frac{x_i^{<p>}(t)}{R^{<p>}} + \sum_{j=1}^{n} A_{ij}^{<p>} \tanh(x_j^{<p>}(t)) + \sum_{q}^{m} w_{qp} \tanh(x_i^{<q>}(t)) \right], \quad (8.10)$$

Now we define a new state variable y, where $y = \tanh(x)$. Therefore, the system defined by equation 8.10 can be rewritten:

$$\dot{y}_i^{<p>} = h(y_i^{<p>}) \dot{x}_i^{<p>} = \frac{h(y_i^{<p>})}{C^{<p>}} \left[-\frac{\tanh^{-1}(y_i^{<p>})}{R^{<p>}} + \sum_{j=1}^{n} A_{ij}^{<p>} y_j^{<p>} + \sum_{q}^{m} w_{qp} y_i^{<q>} \right], \quad (8.11)$$

where

$$h(y) = \frac{dy}{dx} = \frac{d}{dx} \tanh(x) = \operatorname{sech}^2(x) = \frac{2}{e^x + e^{-x}} \quad (8.12)$$

The interconnected CGNNs in 8.11 can be split into a form of interconnected system where

$$f^{<p>}(x^{<p>}, t) = \{ f_i^{<p>}(x_i^{<p>}, t),\ i \in 1...n \}$$

is an n-dimensional function, and each of its elements is defined as follows:

$$f_i^{<p>}(x_i^{<p>}, t) = \frac{h(x_i^{<p>})}{C^{<p>}} \left[-\frac{\tanh^{-1}(x_i^{<p>})}{R^{<p>}} + \sum_{j=1}^{n} A_{ij}^{<p>} x_j^{<p>} \right]. \tag{8.13}$$

The interconnecting-dynamics function g is defined as:

$$g_i^{<p>}(x_i^{<\cdot>}, t) = \frac{h(x_i^{<p>})}{C^{<p>}} \left[\sum_{q}^{m} w_{qp} x_i^{<q>} \right]. \tag{8.14}$$

According to Khalil's method [95], the candidate Lyapunov function of interconnected system can be a weighted sum of each module's Lyapunov function (as shown in equation 5). Cohen and Grossberg [92] gave the Lyapunov function for each CGNN module (8.13), which is as shown below:

$$\Lambda^{<p>}(x^{<p>}) = -\frac{1}{2} \sum_i \sum_j A_{ij}^{<p>} x_i^{<p>} x_j^{<p>} + \frac{1}{R^{<p>}} \sum_i \int_0^{x_i^{<p>}} \tanh^{-1}(y) dy \tag{8.15}$$

and

$$\dot{x}_i^{<p>} = -\frac{1}{C^{<p>}} h(x_i^{<p>}) \frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}} \tag{8.16}$$

The derivative of $\Lambda^{<p>}(x^{<p>})$ is given by:

$$\dot{\Lambda}^{<p>}(x^{<p>}) = \sum_i \frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}} \dot{x}_i^{<p>} = -\sum_i \frac{1}{C^{<p>}} \mathrm{sech}^2(x_i^{<p>}) \left(\frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}} \right)^2 \leq 0 \tag{8.17}$$

To look for the conditions under which equation 8.9 becomes less than zero, we need to find the upper bound of each term. We can begin with the first term:

$$\begin{aligned}
V_f(x^{<p>}) &= \frac{\partial \Lambda^{<p>}}{\partial x^{<p>}} f^{<p>}(x^{<p>}) \\
&= \dot{\Lambda}^{<p>}(x^{<p>}) \\
&= -\sum_{i=1}^{n} \frac{\mathrm{sech}^2(x_i^{<p>})}{C^{<p>}} \left(\frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}} \right)^2 \\
&= -\sum_{i=1}^{n} \frac{\mathrm{sech}^2(x_i^{<p>})}{C^{<p>}} \left(\sum_{j=1}^{n} A_{ij}^{<p>} x_j^{<p>} - \frac{\tanh^{-1}(x_i^{<p>})}{R^{<p>}} \right)^2 \\
&= -\alpha^{<p>} [\phi^{<p>}(x^{<p>})]^2
\end{aligned}$$

where

$$\alpha^{<p>} = \frac{1}{C^{<p>}},$$

and $\phi^{<p>}(x^{<p>})$ is the L-2 norm of an n-dimensional function, and it is defined as follows:

$$\phi^{<p>}(x^{<p>}) = \left\| \left\{ \operatorname{sech}(x_i^{<p>}) \frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}},\ i = 1...n \right\} \right\|, \tag{8.18}$$

where $\|\cdot\|$ is the L2-norm. Therefore, the first part $V_f(x^{<p>})$ has an upper bound as below:

$$V_f(x^{<p>}) = \frac{\partial \Lambda^{<p>}}{\partial x^{<p>}} f^{<p>}(x^{<p>}) \leq -\alpha^{<p>} \phi^{<p>}(x^{<p>})^2,$$

By equation 8.18, we have

$$\left\| \frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}} \right\| = \left\| \frac{1}{\operatorname{sech}(x_i^{<p>})} \right\| \phi^{<p>}(x^{<p>}).$$

Because $x_i^{<p>} \in (-1, 1)$,

$$\left\| \frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}} \right\| \leq \frac{1}{\operatorname{sech}(1)} \phi^{<p>}(x^{<p>}).$$

Let $\beta^{<p>} = \frac{1}{\operatorname{sech}(1)}$, then

$$\left\| \frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}} \right\| \leq \beta^{<p>} \phi^{<p>}(x^{<p>})$$

For the second term of equation 8.9,

$$V_s(x^{<p>}) = \sum_{i=1}^{n} \left(\frac{\partial \Lambda^{<p>}}{\partial x_i^{<p>}} g_i^{<p>}(x_i^{<\cdot>}, t) \right)$$

$$\leq \sum_{i=1}^{n} \left[\beta^{<p>} \phi^{<p>}(x^{<p>}) g_i^{<p>}(x_i^{<\cdot>}, t) \right]$$

$$= \beta^{<p>} \phi^{<p>}(x^{<p>}) \sum_{i=1}^{n} g_i^{<p>}(x_i^{<\cdot>}, t)$$

$$= \beta^{<p>} \phi^{<p>}(x^{<p>}) \sum_{i=1}^{n} \left(\frac{\operatorname{sech}^2(x_i^{<p>})}{C^{<p>}} (\sum_{q=1}^{m} w_{qp} x_i^{<q>}) \right)$$

$$= \beta^{<p>}\phi^{<p>}(x^{<p>})\left[\sum_{q=1}^{m}\left(\frac{w_{qp}}{C^{<p>}}\sum_{i=1}^{n}\operatorname{sech}^2(x_i^{<p>})x_i^{<q>}\right)\right]$$

Now consider the part

$$\sum_{q=1}^{m}\left(\frac{w_{qp}}{C^{<p>}}\sum_{i=1}^{n}\operatorname{sech}^2(x_i^{<p>})x_i^{<q>}\right)$$

$$\leq \sum_{q=1}^{m}\left(\left\|\frac{w_{qp}}{C^{<p>}}\right\|\sum_{i=1}^{n}\left\|x_i^{<q>}\right\|\right)$$

If there exists a constant $k^{<q>}$, such that

$$k^{<q>}\left\|x_i^{<q>}\right\| \leq \phi_i^{<q>}(x_i^{<q>}) \leq \phi^{<q>}(x^{<q>}),$$

then

$$\sum_{q=1}^{m}\left(\left\|\frac{w_{qp}}{C^{<p>}}\right\|\sum_{i=1}^{n}\left\|x_i^{<q>}\right\|\right) \leq \sum_{q=1}^{m}\left(\left\|\frac{nw_{qp}}{C^{<p>}k^{<q>}}\right\|\phi^{<q>}(x^{<q>})\right).$$

The constant $k^{<q>}$ can be set to

$$k^{<q>} = \frac{\operatorname{sech}(1)}{R^{<q>}},$$

so that

$$k^{<q>}\left\|x_i^{<q>}\right\| \leq \phi_i^{<q>}(x_i^{<q>}) = \left\|\sum_{j=1}^{n}A_{ij}^{<q>}x_j^{<q>} - \frac{\tanh^{-1}(x_i^{<q>})}{R^{<q>}}\right\|\operatorname{sech}(x_i^{<q>}).$$

Since $x_i^{<q>} \in (-1,1)$,

$$\operatorname{sech}(x_i^{<q>}) \in (\operatorname{sech}(1),1).$$

Then the derivative of V is bounded above as:

$$\dot{V}(x) = \sum_{p=1}^{m}d^{<p>}\left[\frac{\partial\Lambda^{<p>}}{\partial x^{<p>}}f^{<p>}(x^{<p>})\right] + \sum_{p=1}^{m}d^{<p>}\sum_{i=1}^{n}\left(\frac{\partial\Lambda^{<p>}}{\partial x_i^{<p>}}g_i^{<p>}(t,x)\right)$$

$$\leq \sum_{p=1}^{m}d^{<p>}\left[-\alpha^{<p>}\phi^{<p>}(x^{<p>})^2 + \sum_{q=1}^{m}\beta^{<p>}\left\|\frac{w_{qp}}{C^{<p>}k^{<q>}}\right\|(\phi^{<p>}(x^{<p>})\phi^{<q>}(x^{<q>}))\right],$$

$$= \sum_{p=1}^{m} d^{<p>} \left[-\alpha^{<p>} \phi^{<p>}(x^{<p>})^2 + \sum_{q=1}^{m} \beta^{<p>} \frac{\|w_{qp}\| \operatorname{sech}(1) R^{<q>}}{C^{<p>}} \phi^{<p>}(x^{<p>}) \phi^{<q>}(x^{<q>}) \right],$$

i.e. the inequality can be rewritten in the form:

$$\dot{V}(x) \leq -\frac{1}{2} \Phi^T (DS + S^T D) \Phi,$$

where

$$\Phi = [\phi^{<1>}, \phi^{<2>}, ..., \phi^{<m>}],$$

$$\mathbf{D} = \begin{bmatrix} d^{<1>} & 0 & . & . & . & 0 \\ 0 & d^{<2>} & 0 & . & . & 0 \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & 0 \\ 0 & . & . & . & 0 & d^{<m>} \end{bmatrix}$$

and S is an $m \times m$ matrix:

$$S_{pq} = \alpha^{<p>} - \beta^{<p>} \gamma_{pq}, \text{ if } p = q,$$

$$S_{pq} = -\beta^{<p>} \gamma_{pq}, \text{ if } p \neq q,$$

where

$$\gamma_{pq} = \frac{\|w_{qp}\| \operatorname{sech}(1) R^{<q>}}{C^{<p>}}.$$

As a result, if the matrix $\mathbf{DS} + \mathbf{S}^T\mathbf{D}$ is positive definite, we can conclude that $\dot{V}$ is negative definite. Hence the system will not oscillate in such a case, and moreover, it is asymptotically stable.

7. Examples and computer simulations

The following illustrative example of a 2-loop thalamocortical circuit configured as figure 37 will demonstrate the effectiveness of the obtained results. The circuit has three layers, and each layer can be treated as a CGNN module. The equilibrium point was found at 0. The parameters used are as in Table I.

Table I. Parameters

$< p >$	$R^{<p>}$	$C^{<p>}$	$A^{<p>}$
1	3	0.3	[0 0; 0 0]
2	3	0.6	[0 -0.2; -0.2 0]
3	3	0.3	[0 0.25; 0.25 0]

The interconnection matrix was set to:

$$\mathbf{W} = \begin{bmatrix} 0 & 0.1 & 0.1 \\ -0.1 & 0 & 0 \\ 0.3 & 0.2 & 0 \end{bmatrix}.$$

For the thalamocortical network, we have $m = 3$, and $n = 2$. The resulting matrix $\mathbf{S}$ is as follows:

$$\mathbf{S} = \begin{bmatrix} 3.3333 & -1.0000 & -1.0000 \\ -0.5000 & 1.6667 & 0 \\ -3.0000 & -2.0000 & 3.3333 \end{bmatrix}.$$

Because $\det(\mathbf{S}) = 10.8519 > 0$, there exists a positive diagonal matrix $\mathbf{D}$ such that the matrix $\mathbf{DS} + \mathbf{S}^T\mathbf{D}$ is positive definite (see [97] for the proof). Therefore, the equilibrium point 0 is asymptotic stable. This is demonstrated in figure 48.

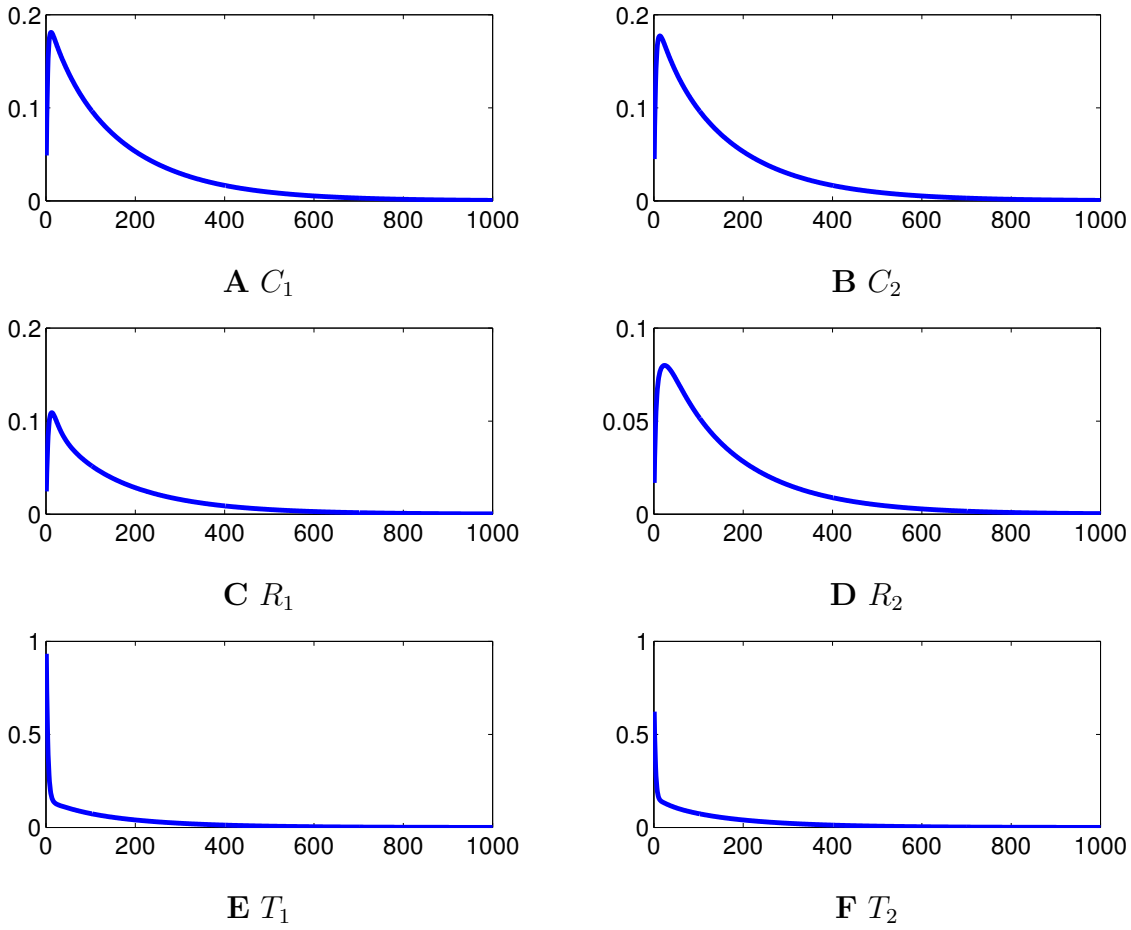


Fig. 48. **Simulation of thalamocortical circuits (two loops) showing convergence behavior.** The x-axis is the time step, while the y-axis the value of the state variable (membrane potential). The network is configured as in figure 37B. T_1 and T_2 are relay cells, R_1 and R_2 TRN neurons, and C_1 and C_2 cortical neurons. The initial value is 1.2 for T_1, 0.8 for T_2, and 0 for all the rest cells. All neurons reach a stable equilibrium by time step 800.

B. Controllability: Control resolution

Besides stability, controllability is also one of the important features of dynamic systems. From the controllability perspective, the pattern of disinhibition can play a role in improving the control accuracy of actions. This point can be best demonstrated

in the local circuit of the basal ganglia.

As shown in figure 49, there are multiple levels of inhibition (type I disinhibition) in the circuits of the basal ganglia. One question that arises here is why the system uses multiple feedforward inhibition instead of one simple excitation to control the action signal in basal ganglia. A possible answer could be that this feedforward disinhibition can reduce the noise in control and therefore increase the *accuracy*. Accuracy means how precisely we can direct an action to a specific state. For a system, if the i^{th} controlling variable is x_i, and the output (the variable being controlled) is y, the pointwise control accuracy $\chi(y_0)$ for output y_0 can be defined as:

$$\chi(y_0) = \max_{x_i} \left\| \log \left(\left. \frac{\partial x_i}{\partial y} \right|_{y=y_0} \right) \right\|. \tag{8.19}$$

The basic idea behind this definition is that if the input value x changes within a wide range, for high accuracy systems, the output y should only change in an arbitrarily small range. There might be many inputs that control the output y_0, so the max operator on all of these returns the maximum accuracy from these inputs. Note that for a meaningful accuracy we require $\chi(y_0)$ to be bounded, because if it is ∞, it means that the output over time is constant in which case it is not controllable (however, it can be seen as ∞-accurate at such value y_0).

In general, when controlling an action, accuracy and *efficiency* are both desired. Efficiency means how fast we can perform an action. For example, we may control a finger to point to an exact point in space, which may need fast movement to the rough neighborhood and then precisely reach a specific point within that local region. As another example, we may need to precisely control the eye muscle to focus on a certain object, in which accuracy is a very important issue.

Higher control accuracy also means higher stability under noise conditions. Con-

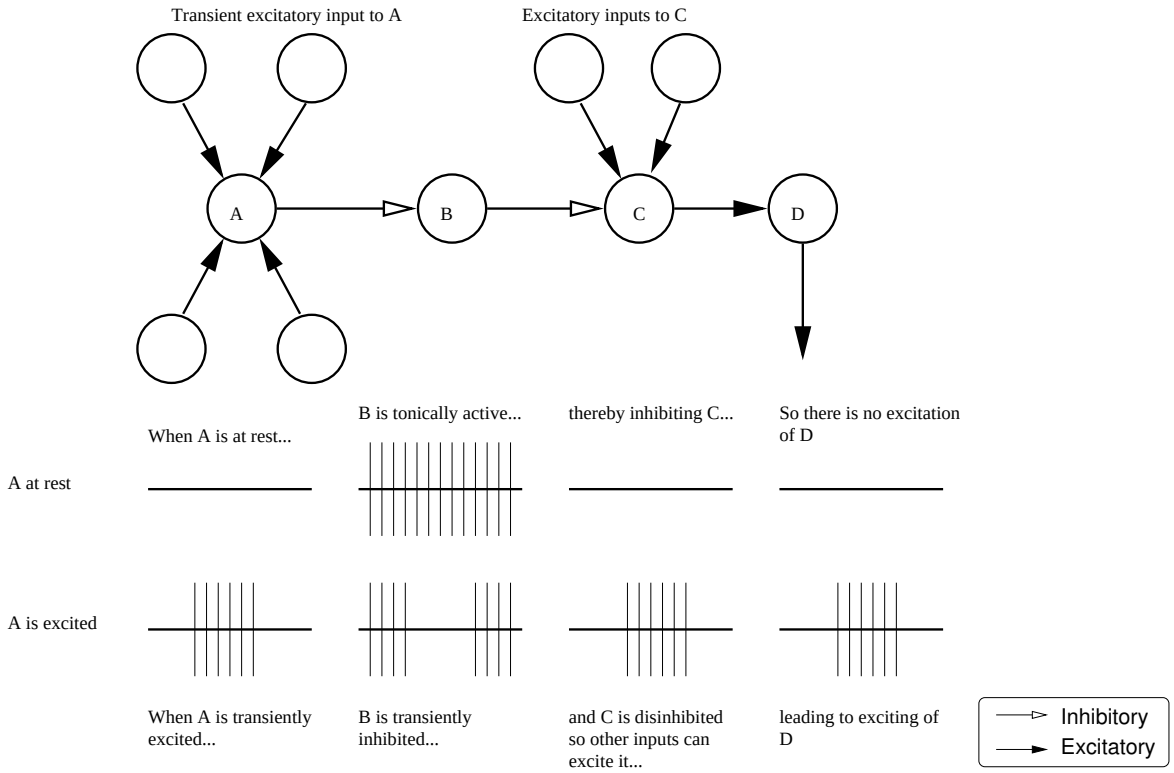


Fig. 49. **Disinhibition in basal ganglia.** Lines with filled arrows represent excitation, while those with unfilled arrows represent inhibition. The two rows below the neural network illustrate two scenarios: (1) neuron A is at rest; and (2) neuron A is excited. The other neurons' (B, C, and D) corresponding activities are shown. The text above and below the two cases provide a more detailed explanation. This figure demonstrates how the singals of actions in basal ganglia can be controlled through disinhibitory mechanisms. (Redrawn from [98].)

sider the first-order Taylor expansion for a system with noise ϵ:

$$y(x_0 + \epsilon) = y_0 + f'(y_0)\epsilon.$$

Thus, the effect of introducing noise in the output y is

$$\|\Delta y\| = \|f'(y_0)\epsilon\|,$$

and by substituting $f'(y_0)$ with χ from equation 8.19, we have

$$\|\Delta y\| = \|\epsilon e^{-\chi(y_0)}\|.$$

Therefore, with higher control accuracy, the noise in the output will decrease exponentially.

The basal ganglia model in figure 49 can be abstracted as figure 50. The inputs are represented by u_1, u_2, and u_3. Let u_2 be a constant input to represent the tonic excitation of the second cell, and the free variable u_1 and u_3 the other external inputs. The weights w_1 and w_2 are all inhibitory, so they have a value less than zero. The activity of each cell f_i is defined as

$$f_i = \sigma(v_i), \tag{8.20}$$

where v_i is the membrane potential, and $\sigma(\cdot)$ the sigmoid function.

The control accuracy of such a system can be derived as

$$\chi(f_3) = \left\|\log\left(\frac{du}{df_3}\right)\right\| = \left\|\log\left(\frac{1}{w_1 w_2 (\sigma')^3}\right)\right\| = -\log(-w_1) - \log(-w_2) - 3\log(\sigma') \tag{8.21}$$

In general, for n cells in feedforward disinhibition (e.g. $n = 3$ in the basal ganglia system), the control accuracy is

$$\chi(f_n) = -\sum_{i=1}^{n-1} \log(-w_i) - n\log(\sigma') \tag{8.22}$$

Therefore, the feedforward disinhibition structure has γ-times improvement in control accuracy, where

$$\gamma = \frac{\chi(f_n)}{\chi(f_1)} = \frac{-\sum_{i=1}^{n-1}\log(-w_i) - n\log(\sigma')}{-\log(\sigma')} = \frac{\sum_{i=1}^{n-1}\log(-w_i)}{\log(\sigma')} + n. \tag{8.23}$$

Note that when w_i and $\mid \sigma' \mid$ is less than 1, the $\log(-w_i)$ term will always have the

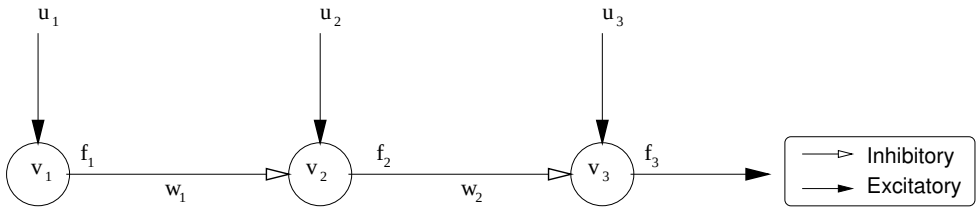


Fig. 50. **The abstract model of basal ganglia.** The lines with filled arrow are excitatory synapses, while those with unfilled arrows inhibitory synapses. This figure is an abstract version of the basal ganglia connectivity in figure 49. Through multiple levels of control, the accuracy is improved compared to direct activation. See text for details.

same sign as $\log(\sigma')$, so that γ is always greater than n. Figure 51 compares the control accuracy of the basal ganglia model ($n = 3$; solid line) with a direct control method (single sigmoid; dashed line). The control accuracy of the disinhibitory network is five time higher than a direct control network.

One problem with high accuracy systems is that, as the accuracy of a control device is improved, the efficiency has to be decreased. Thus, the system has to spend a relatively longer time to reach the specific state. However, note that there is a direct input at the cell C in figure 49, which allows a fast estimated input, and the disinhibition mechanism can be employed to refine the control at a much smaller scale. As shown in the multiple-input sites in the basal ganglia, both efficiency and accuracy can be achieved by such a disinhibitory structure.

Another question is why the synaptic weight has to be inhibitory to achieve the accuracy, because if both w_1 and w_2 are positive the same accuracy level can be achieved. The answer is two-fold. First, inhibitory synapses allow the modification of the input u_1 in both way (to add more input or reduce input) while purely excitatory ones cannot deal with overflow in input of u_1 by adjusting u_3. Second, the inhibitory mechanism also provides a logical function of action selection, which is an important

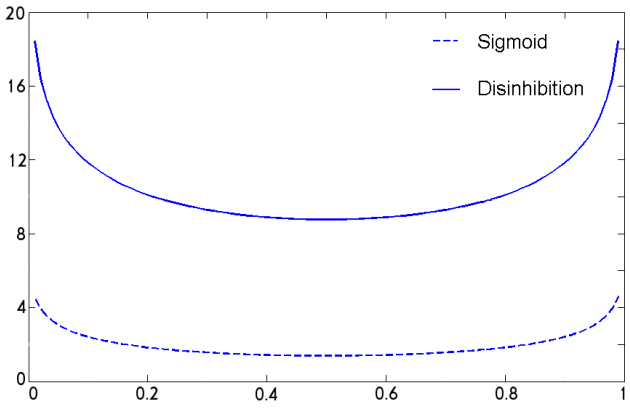


Fig. 51. **Control accuracy of the basal ganglia model with feedforward disinhibition.** The x-axis is the value of output (f_3 in figure 50), and the y-axis the control accuracy χ. The dashed line is the control accuracy of sigmoid function using direct control. The control accuracy of the feedforward disinhibition structure (solid line) is increased five times compared to the direct control method (dashed line). ($w_1 = 0.1$, $w_2 = 0.1$)

function of the basal ganglia. This observation will be analyzed in the next section.

In sum, feedforward disinhibition can improve the control accuracy by at least twice when the inhibition rate is less than 1.

C. Computability: The logic of control

In previous sections, we analyzed the stability and control accuracy of disinhibitory neural network. In this section, we will see how the disinhibition contribute to the the logic operation of the brain. In logic, "not" is an atomic operator, and every pair of "not" can be eliminated. The question is whether the neurons can implement the same logical operation through excitatory and inhibitory interactions. The answer is yes. If the inhibitory interaction is analogous to the "not" operator, then a chain of inhibitory interactions, which is type II disinhibition, can implement the same logic

as a chain of "not" logic operators. For example, as illustrated in figure 49, cell B tonically inhibit the excitation of cell C when cell B is not inhibited by cell A. Assuming there is a pattern of excitation $\Omega_C(t)$ to cell C, and cell A inhibits cell B at $t = t_0$, then only when $t > t_0$ the muscle controlled by cell C can exhibit a pattern of $\Omega_C(t)$. In this sense, the disinhibition mechanism behave as a switch in the system. With an analogy to logic, disinhibition realizes the logic of $\neg\neg\Omega_i = \Omega_i$ by neuronal connections.

With feedforward disinhibition, the system is able to jump its state from Ω_i to Ω_j, and this jump can be interpreted as a fast response (or action) to the external world. The mechanism can eliminate the setup time (or convergence time) of the internal neurons to get ready for any predicted action. For example, the system needs an action Ω at time t_0 but if the setup time is δt then the system can initiate the sequence at an earlier time $t_0 - \delta t$. The disinhibition mechanism can hold the action until time t_0, thus at time t_0 the system can generate the desired action pattern Ω without going through the unnecessary initiation behavior before the actual action.

Interestingly, signal routing by disinhibition mechanism can also be found in the thalamic circuit (see figure 52). The disinhibition mechanism can perform as a gating controller as proposed by Choe [99]. The input to the thalamic relay (T_2) cell is initially inhibited by the TRN neuron R_2, but when the cortical neuron (C_1) excites R_1, R_1 will inhibit R_2 and T_2 then gets disinhibited. Thus the disinhibition of $R_1 \rightarrow R_2$ acts as a switch for signal transferred from the relay cell T_2 to the cortical neuron C_2. Figure 53 shows the case the sensory input signal is transfered to the cortex, and figure 54 demonstrates that the sensory input is suppressed by disinhibition initiated by the cortex cell.

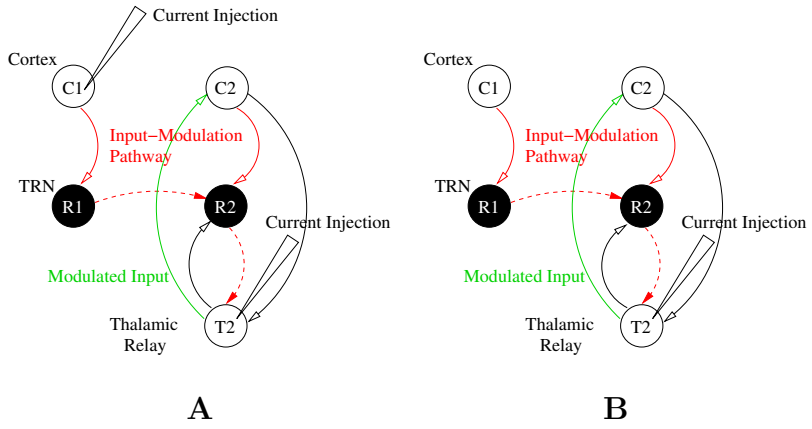


Fig. 52. **Signal routing by disinhibition.** The lines with filled arrows represent excitatory connections and those with unfilled arrows inhibitory connections. Cortical neurons: C_1 and C_2; TRN neurons: R_1 and R_2; relay cells: T_1 and T_2. **A** Current is injected to T_2 simulating tonic input, and to C_1 as the disinhibition signal. The disinhibition mechanism can perform as a gating controller as proposed by Choe [99]. The input to the thalamic relay (T_2) cell is initially inhibited by the TRN neuron R_2, but when the cortical neuron (C_1) excites R_1, R_1 will inhibit R_2 and T_2 then gets disinhibited. Thus the disinhibition of $R_1 \rightarrow R_2$ acts as a switch for signal transfer from the relay cell T_2 to the cortical neuron C_2. See figure 53 for the simulation results under this scenario. **B** There is no current injection to C_1. Activity at C_2 is significantly reduced by the inhibition from R_2. See figure 53 for the simulation result of this configuration.

D. Discussion

In this chapter, we explored the role of disinhibition from a system perspective. For some hybrid form of disinhibitory neural network (mixed of feedforward and recurrent types), Cohen-Grossberg (C-G) model cannot be directly applied due to its symmetry requirement. The interconnected Cohen-Grossberg model (iCGNN) proposed here solves this problem by treating each layer as a C-G module and allowing asymmetric connections between modules. The iCGNN model is motivated partly by the thala-

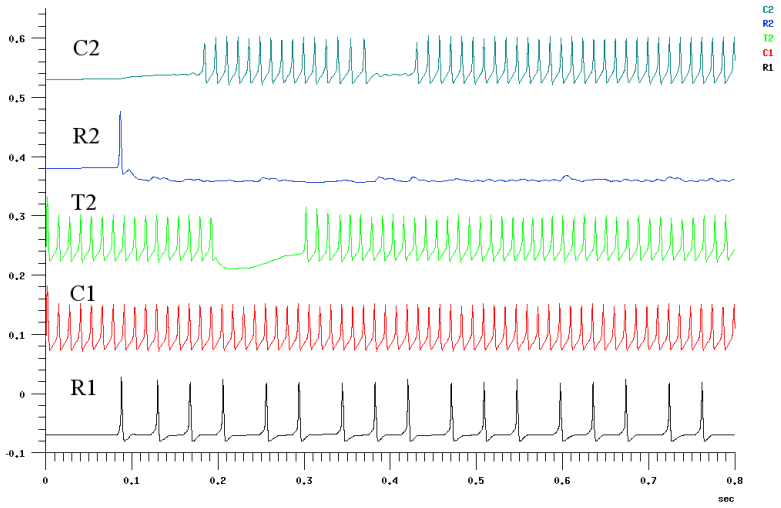


Fig. 53. **Experiment 1: Signal is allowed from the sensory input to the cortex cells.** The scenario illustrated in figure 52**A** is shown. Current is injected into T_2 to simulate tonic input. Current is also injected to C_1. Signal from T_2 directly goes to C_2. The neurons are Hodgkin-Huxley neurons, and they are simulated using GENESIS. The disinhibition mechanism can perform as a gating controller as proposed by Choe [99]. The input to the thalamic relay (T_2) cell is initially inhibited by the TRN neuron R_2, but when the cortical neuron (C_1) excites R_1, R_1 will inhibit R_2 and T_2 then gets disinhibited. Thus the disinhibition of $R_1 \rightarrow R_2$ acts as a switch for signal transferred from the relay cell T_2 to the cortical neuron C_2.

mocortical circuit proposed by Choe [20] and partly by the C-G model. The iCGNN model has general utility in describing neural networks with a layered structure, such as the three-layered thalamocortical circuit or six layered cortex.

One limitation of our current result is that we did not consider the delayed case. The existence of time delays can frequently cause oscillation, divergence, or instability in neural networks. For example, the temporal properties can give rise to several different behaviors in the thalamocortical model [20]. A network connection without delay can be asymptotic stable with conditions derived in this section, but

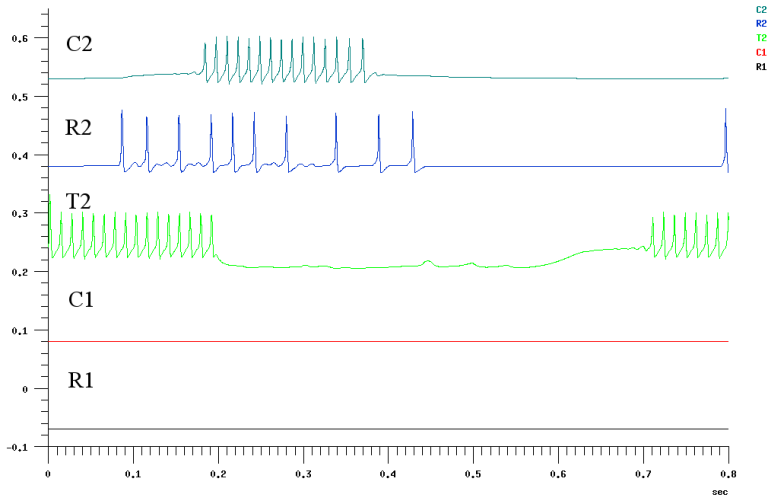


Fig. 54. **Experiment 2: Signal is blocked from the sensory input to the cortex cells.** The scenario illustrated in figure 52**B** is shown. There is no current injection to C_1. Activity is significantly reduced at C_2. Other details were the same as figure 53. The input to the thalamic relay (T_2) cell is inhibited by the TRN neuron R_2. Since the neuron C_1 is not activated, there is no activity for the neuron R_1. Therefore, R_1 does not inhibit R_2, and T_2's activity is reduced by the activation of R_2. As a result, the activity at C_2 gets reduced.

the same network connection can become unstable or oscillating when there is delay. Mathematically, to calculate the stability of neural networks with delay could be more complex. However, neural networks with delay is a topic of great theoretical interest and practical importance, and thus iCGNN with delay may be a promssing future research topic in studying the function of disinhibitory neural networks.

E. Summary

In this chapter, I proposed the iCGNN model for hybrid disinhibitory neural networks. I also derived a set of sufficient parameter conditions which can make the thalamocortical circuit model (as an iCGNN model) show asymptotically stable behavior. The

approach extends the Cohen-Grossberg network approach, and it can be applied to the stability analysis of multiple-layered brain network in general. We also discussed how the multiple level of inhibition can increase the accuracy of controllability and implement basic "not-not" logic function.

CHAPTER IX

DISCUSSION AND CONCLUSION

A. Summary

In searching for the computational role of disinhibition in brain function, I used the approach of building computational models, applying them in local circuits in the brain, and explaining psychological phenomena. The cycle in my approach is illustrated in figure 55. The computational role of disinhibition is discussed from three perspective: visual perception, attentional control, and system framework.

Furthermore, the research in this dissertation can be summarized from two dimensions (Table II). One dimension is along the study of disinhibition (i.e., computational model, brain organization, and brain function), and the other along the analysis under a system science framework. In the first dimension, the research covers disinhibition from abstract neuronal circuits and the disinhibition effect of various subsystems in the brain, such as the retina, the primary visual cortex, and the thalamus. In the second dimension, the system perspective, I first analyzed the equilibrium points by extending the Hartline-Ratliff equation to the IDoG model. The IDoG model with selfinhibition expresses dynamics in disinhibition. Furthermore, the asymptotic stability analysis of interconnected Cohen-Grossberg neural network (a case of type III disinhibition) was conducted to help us understand the contribution of disinhibition in the system's overall stability. Moreover, two other possible functions of disinhibition, to improve control accuracy and to implement the logic of control, were proposed.

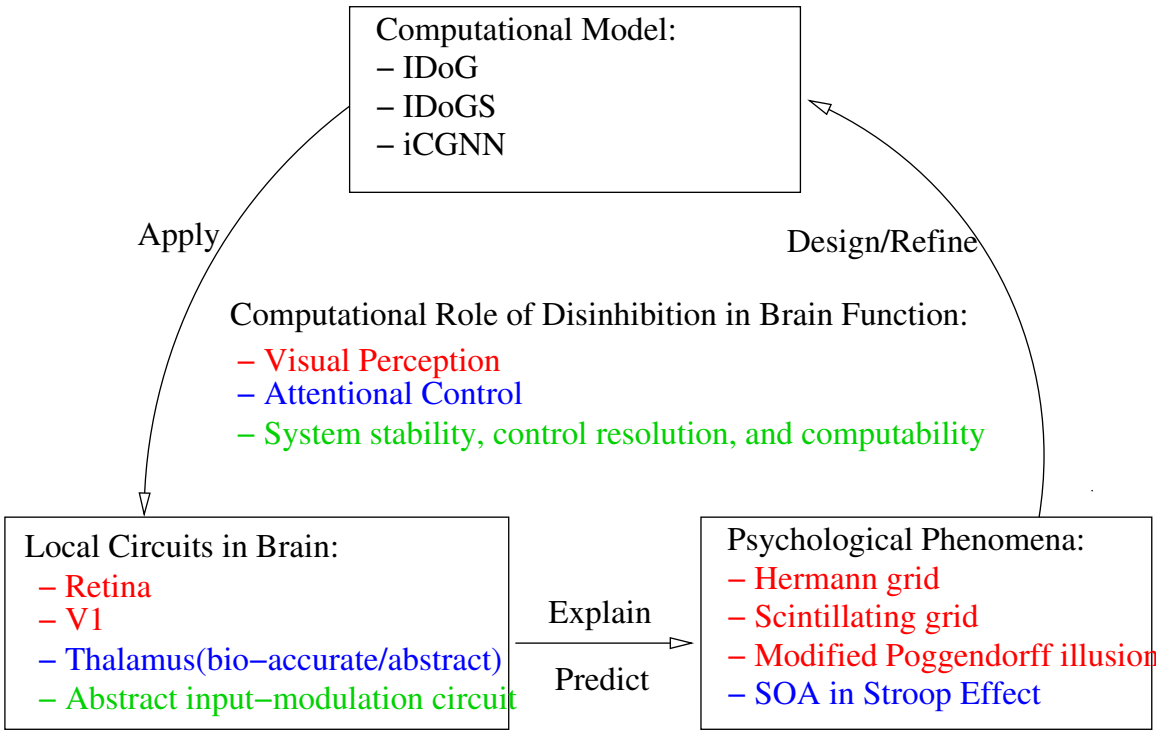


Fig. 55. **Main results in studying the computational role of disinhibition.** I studied the local circuit with disinhibition features in the brain, which includes the retina, the primary visual cortex (V1), and the thalamocortical circuit. Several psychological phenomena were explained, such as the periphery problem in the Hermann grid illusion, the White's effect, the scintillating grid illusion, the modified Poggendorff illusion, and the SOA effect in Stroop tasks. Three computational models have been proposed: IDoG, IDoGS, and iCGNN. In sum, the computational roles of disinhibition in brain function have been studied from three perspectives: (1) visual perception; (2) attentional control; and (3) system approach.

B. Discussion

In this section, I will first discuss the limitations of the computational models of disinhibition, and then summarize the predictions based on these models. Finally, I will conclude, with the contributions of my research.

Table II. Computational Roles of Disinhibition in Brain Function

	Equilibrium Point	**Single Layer Dynamics**	**Multiple Layer Dynamics**	**System**
Computational Model of Disinhibition	IDoG	IDoGS	iCGNN	Temporal input-modulation model
Brain Organization	Retina, LGN, primary visual cortex	Retina, LGN	Thalamocortical circuit	
Brain Function	Visual perception	Visual perception	Stability of neural networks	Attentional control, Input modulation

1. Limitations of the approach

The IDoG and IDoGS model (Chapter III) extended the Hartline-Ratliff equation to account for a population of neurons, but it requires matrix inverse to calculate the response. The inverse of the weight matrix, which is symmetric Toeplitz one, has the computational complexity of $O(n \log n)$ to calculate as reported by Heinig and Rost [24]. One method to alleviate the expensive calculation is to avoid doing the inverse operation for all the neurons (i.e., decrease the size of the matrix to be inversed), and rather use an impulse response matrix as a filter with a small receptive field size, and then convolve the filter with the initial input of the population of neurons. Another potential problem with the matrix inverse operation is that for some of the kernel functions, the resulting weight matrix may turn out to be singular, which means that the inverse operation cannot be applied. In such a case, parameters of the kernel function need to be changed in order to be accommodated by these models. Fortunately in all our experiments, applying the different of Gaussians profiles as the kernel function did not incur such a problem.

For the attentional input modulation model (Chapter VII), one limitation is that

it does not evaluate whether a stimulus is relevant or irrelevant. At the current stage, the relevancy of the input to the response was pregiven. However, this functional block of relevancy evaluation can be extended by checking if a stimulus is sufficient to invoke a correct response, and if so, it can be labeled as a relevant stimulus. For example, in the congruent case of color-naming task, a first appearing word "red" that is followed by a red block is sufficient to invoke a verbal response "red", therefore, it is a relevant stimulus and the association can be learned.

For the interconnected CGNN model, the limitation is that it did not consider the delayed case. The existence of time delays can frequently cause oscillation, divergence, or instability in neural networks. For example, the temporal properties can give rise to different behaviors in the thalamocortical model [20]. A network connection without delay can be asymptotic stable with conditions derived in section VIII.B, but the same network connection can also be unstable or oscillating when with delays. Mathematically, to calculate the stability of neural networks with delay could be more complex. Neural networks with delay is a topic of great theoretical interest and practical importance, and however, it is beyond the scope of this dissertation.

2. Predictions

The computational models of IDoG and IDoGS provide some interesting predictions.

In the dynamic brightness-contrast illusion experiments on the scintillating grid (see Chapter V), the IDoGS model gives a couple of interesting predictions (both of which were brought to our attention by Rufin VanRullen). The first prediction is that scintillating effect will occur only in an annular region in the visual field surrounding the fixation point where the size of the receptive field matches that of the grid element size. However, this does not seem to be the case under usual viewing conditions. Our explanation for this apparent shortcoming of the model is that the size of usual

scintillating grid images is not large enough to go beyond the outer boundary of the annular region. Our explanation can be tested in two ways: Test the strength of illusion with (1) a very large scintillating grid image where the grid-element size remains the same, or with (2) the usual sized image with a reduced grid-element size. We expect that the annular region will become visible in both cases, where no scintillating effect is observed beyond the outer boundary of the annular region.

The second prediction is that the scintillation would be synchronous, due to the same time course followed by the neurons responding to each scintillating grid element. Again, this is quite different from our perceived experience, which is more asynchronous. In our observation, the asynchronicity is largely due to the random nature of eye movement. If that is true, the scintillating effect will become synchronous if eye movement is suppressed. That is, if we fixate on one location of the scintillating grid while the stimulus is turned on and off periodically (or alternatively, we can blink our eyes to simulate this), all illusory dark spots would seem to appear all at the same time, in a synchronous manner. Then, why is our experience asynchronous? The reason why we perceive the scintillation to be asynchronous may be because when we move our gaze from point X to point Y in a long saccade, first the region surrounding X, and then at a later time the region surrounding Y scintillates. This will give us the impression that the scintillating effect is asynchronous.

In the orientation perception experiments on modified Poggendorff illusion (Chapter VI), the disinhibition model based on IDoG made two other novel predictions. First, the strength of inhibition can significantly affect the illusory effect (Figure 34A). The stronger the inhibition, the larger the magnitude of the curve in Figure 35A, while the locations of the peak and the valley of the curve are not affected. This observation cannot be verified through purely psychological means, however, if combined with physiological experiments, it may be possible to test. An animal can be trained

to select an apparently collinear lines over misaligned ones in a non-illusory task. In the first test, the stimulus can be present in the periphery of the animal's visual field, where the tuning curve has a standard deviation measured as σ_1. Record the animal's choices for the stimulus configured as in Figure 35A. Then, in the second test, apply bicuculline to block GABA receptors in the animal's primary visual cortex to reduce the strength of inhibition. The blocking of GABA receptor may also incur the change of the orientation tuning curve's standard deviation, and therefore we may need to move the stimulus towards the fovea region to increase the reduced standard deviation value. (In the fovea, its standard deviation is smaller, which will be explained in the next paragraph.) We can stop moving the stimulus right at the location where the receptive field has the same tuning width as in the first test, σ_1. Again, record the animal's choices for the stimulus configured as in Figure 35A. Comparing the two results obtained in those two tests, in which both of the standard deviations are the same but the inhibition strengths are different, the result of the first test is expected to have a larger magnitude than that of the second one due to the reduced inhibition strength in the second test. Furthermore, the peak and the valley position will not change as predicted in Figure 35A.

Another observation is that the illusory effect can heavily depend on the standard deviation σ of the orientation tuning curve. The value of σ can affect both the magnitude and locations of the peak and valleys in Figure 35B. Further psychological experiment could be conducted to verify this observation. The variation in perceived orientation can be compared under two conditions: one is to present the stimulus to the fovea, and the other is to present it to the periphery. In the fovea area, it is supposed to have smaller σ than that in the periphery. Therefore, according to our computational experiment based on the value of σ (figure 35B), we are expecting that in the periphery there is stronger illusion (larger magnitude) than in the fovea; and

also in the periphery, the second thick bar must have a higher degree of orientation than in the fovea to make the thin line's orientation to be maximally enhanced and while a lower degree of the second thick bar in the periphery can get the thin line's orientation maximally reduced (i.e., locations of the peak and valley will change).

In sum, the above predictions of the computational models are expected to be consistent with experiments under similar conditions. Further psychophysical experiments may have to be conducted to more rigorously test the model predictions.

3. Contributions

The main contribution of this research is that it used a unified framework of disinhibition to model various neural circuits in the brain. The framework explains different types of visual illusion in the context of disinhibition: brightness-contrast illusions [18] and geometric illusions [19]. As shown in the Hermann grid and Mach band experiments, the static disinhibitory model IDoG can preserve brightness and enhance contrast in multiple spatial frequency scales. The IDoGS model is the first computational model to explain the scintillating grid illusion [25; 100]. Furthermore, control accuracy, logic of control, and stability analysis of interconnected CGNN are novel ways of analyzing the role of disinhibition in the brain. The analysis framework for interconnected CGNN presented here eases the symmetry requirement of Cohen-Grossberg neural network [101], and has the potential to be applied in the analysis of large-scale multi-layer neural networks. Finally, the input modulation model combining recurrent and feedforward disinhibition (based on the gating controller theory of thalamocortical circuit [99]) was successfully applied in modeling SOA effects in the Stroop task. The proposed concept of "selection of When" extended selective attention into the temporal domain [60].

C. Future directions

Future research can be conducted in two directions. The first is to more broadly apply the disinhibition model to various local circuits in the brain. For example, analyzing the role of disinhibition in the cerebellum, the hippocampus, the basal ganglia, and their interconnections. Observation of various disinhibitory patterns can help us to have a deeper understanding of the role of disinhibition in the brain.

Second, I will further carry out systems analysis of disinhibitory neural circuits in the brain. The system-level analysis includes the studies of system dynamics and design methods. The system dynamic, which is characterized by sensitivity, phase portrait, and bifurcation, can provide insights on the change in behavior over time or in different parameter space. The design methods, such as controllability and stability studies, can guide the design of robust biologically inspired systems. Methods for increased controllability can direct us in designing controllable system components in a principled way. Systems with improved stability can deal with noisy inputs from the environment and ensure that the system is working properly under various unexpected circumstances.

Third, further study of disinhibition in higher-level perception and cognition could also be a worthwhile topic. For example in Stroop effect, for two dimensional stimulus, there are competitions between those two dimensions, or we say the two stimuli inhibit each other. It would be interesting to see what happens if we have stimulus from three or more dimensions (where disinhibition takes place), and to predict the perception and cognition of the multi-dimensional interference through the disinhibition model at a higher-level. Topics such as attentional selection and decision making in light of disinhibition may lead to new insights.

D. Conclusion

In sum, the computational role of disinhibition has been studied from three perspectives of visual perception, attentional control, and system framework. The study of the role of disinhibition in the brain can help us in understanding how the brain can realize intelligent and complex functions through simple neuronal excitation and inhibition. Further research on other basic circuits containing disinhibition can help us discover deeper principles of brain function.

REFERENCES

[1] H. K. Hartline and F. Ratliff, "Inhibitory interaction of receptor units in the eye of Limulus.," *Journal of General Physiology*, vol. 40, pp. 357–376, 1957.

[2] L. Spillmann, "The Hermann grid illusion: a tool for studying human perceptive field organization," *Perception*, vol. 23, pp. 691–708, 1994.

[3] M. Schrauf, B. Lingelbach, and E. R. Wist, "The scintillating grid illusion," *Vision Research*, vol. 37, pp. 1033–1038, 1997.

[4] H. K. Hartline, H. Wager, and F. Ratliff, "Inhibition in the eye of Limulus," *Journal of General Physiology*, vol. 39, pp. 651–673, 1956.

[5] H. K. Hartline and F. Ratliff, "Spatial summation of inhibitory influences in the eye of Limulus, and the mutual interaction of receptor units," *Journal of General Physiology*, vol. 41, pp. 1049–1066, 1958.

[6] G. Chevalier and JM. Deniau, "Disinhibition as a basic process in the expression of striatal functions," *Trends in Neurosciences*, vol. 137, pp. 277–280, 1990.

[7] David LaBerge and S. Jay Samuels, "Toward a theory of automatic information processing in reading," *Cognitive Psychology*, vol. 6, pp. 293–322, 1974.

[8] David LaBerge and S. Jay Samuels, "Toward a theory of automatic information processing in reading," in *Theoretical Models and the Processes of Reading*, Harry Singer and Robert B. Ruddell, Eds., pp. 293–322. International Reading Association, Newark, DE, 1985.

[9] D. A. McCormick and T. Bal, "Sleep and arousal: thalamocortical mechanisms," *Annual Review of Neuroscience*, vol. 20, pp. 185–215, 1997.

[10] G. T. Bartha and R.F. Thompson, "Cerebellum and conditioning," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed., pp. 169–172. MIT Press, Cambridge, MA, 1995.

[11] M. A. Arbib, P Erdi, and J. Szentagothai, *Neural Organization, Strucutre, Fuction, and Dynamics*, MIT Press, Cambridge, MA, 1997.

[12] C. Y. Li, Y. X. Zhou, X. Pei, F. T. Qiu, C. Q. Tang, and X. Z. Xu, "Extensive disinhibitory region beyond the classical receptive field of cat retinal ganglion cells," *Vision Research*, vol. 32, pp. 219–228, 1992.

[13] H. Kolb and R. Nelson, "Off-alpha and off-beta ganglion cells in the cat retina," *Journal of Comparative Neurology*, vol. 329, pp. 85–110, 1993.

[14] B. Roska, E. Nemeth, and F. Werblin, "Response to change is facilitated by a three-neuron disinhibitory pathway in the tiger salamander retina," *Journal of Neuroscience*, vol. 18, pp. 3451–3459, 1998.

[15] M. J. Frech, J. Perez-Leon, H. Wassle, and K. H. Backus, "Characterization of the spontaneous synaptic activity of amacrine cells in the mouse retina," *Journal of Neurophysiology*, vol. 86, pp. 1632–1643, 2001.

[16] M. Meister and M. J. Berry II, "The neural code of the retina," *Neuron*, vol. 22, pp. 435–450, 1999.

[17] Carpenter and Blakemore, "Interactions between orientations in human vision," *Experimental Brain Research*, vol. 18, pp. 287–303, 1973.

[18] Yingwei Yu, Takashi Yamauchi, and Yoonsuck Choe, "Explaining low level brightness-contrast visual illusion using disinhibition," in *Biologically Inspired Approaches to Advanced Information Technology, Lecture Notes in Computer Science 3141*, A. J. Ijspeert, M. Murata, and N. Wakamiya, Eds., pp. 166–175. Springer, Berlin, 2004.

[19] Yingwei Yu and Yoonsuck Choe, "Angular disinhibition effect in a modified Poggendorff illusion," in *Proc. the 26th Annual Conference of the Cognitive Science Society*, Kenneth D. Forbus, Dedre Gentner, and Terry Regier, Eds., 2004, pp. 1500–1505.

[20] Yoonsuck Choe, "The role of temporal parameters in a thalamocortical model of analogy," *IEEE Transactions on Neural Networks*, vol. 15, pp. 1071–1082, 2004.

[21] C. F. Stevens, "A quantitative theory of neural interactions: Theoretical and experimental investigations," Ph.D. dissertation, The Rockefeller University, New York, NY, 1964.

[22] S. Brodie, B. W. Knight, and F. Ratliff, "The spatiotemporal transfer function of the limulus lateral eye," *Journal of General Physiology*, vol. 72, pp. 167–202, 1978.

[23] D. Marr and E.C. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London B*, vol. 207, pp. 187–217, 1980.

[24] G. Heinig and K. Rost, "Hartley transform representations of symmetric toeplitz matrix inverses with apllication to fast matrix-vector multiplication," *SIAM Journal on Matrix Analysis and Applications*, vol. 77, pp. 86–105, 2000.

[25] Yingwei Yu and Yoonsuck Choe, "Explaining the scintillating grid illusion using disinhibition and self-inhibition in the early visual pathway," in *Program No. 301.10. 2004. Annual Conference of Society for Neuroscience*, San Diego, CA, 2004.

[26] E. H. Adelson, "Lightness perception and lightness illusions," in *The New Cognitive Neurosciences*, M. Gazzaniga, Ed., pp. 339–351. MIT Press, Cambridge, MA, 2000.

[27] A. Fiorentini, G. Baumgartner, S. Magnussen, P. H. Schiller, and J. P. Thomas, "The perception of brightness and darkness: Relations to neuronal receptive fields," in *Perception: The Neurophysiological Foundations*, L. Spillmann and J. S. Werner, Eds., pp. 129–161. Academic Press, San Diego, CA, 1990.

[28] F. Kelly and S. Grossberg, "Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception," *Perception and Psychophysics*, vol. 62, pp. 1596–1618, 2000.

[29] E. B. Goldstein, *Sensation and Perception*, Wadsworth-Thomson Learning, Pacific Grove, CA, 2000.

[30] M. White, "A new effect of pattern on perceived lightness," *Perception*, vol. 8, pp. 413–416, 1979.

[31] M. J. Frech, J. Perez-Leon, H. Wassle, and K. H. Backus, "Characterization of the spontaneous synaptic activity of amacrine cells in the mouse retina," *Journal of Neurophysiology*, vol. 86, pp. 1632–1643, 2001.

[32] F. Ratliff and H. K. Hartline, "The responses of limulus optic nerve fibers to

patterns of illumination on the receptor mosaic.," *Journal of General Physiology*, vol. 42, pp. 1241–1255, 1959.

[33] M. Schrauf, E. R. Wist, and W. H. Ehrenstein, "The scintillating grid illusion during smooth pursuit, stimulus motion, and brief exposure in humans," *Neuroscience Letters*, vol. 284, pp. 126–128, 2000.

[34] Akiyoshi Kitaoka, *Trick Eyes 2*, Kanzen, Tokyo, 2003.

[35] Espen Hartveit, "Reciprocal synaptic interactions between rod bipolar cells and amacrine cells in the rat retina," *Journal of Neurophysiology*, vol. 81, pp. 2932–2936, 1999.

[36] H. Neumann, L. Pessoa, and T. Hanse, "Interaction of on and off pathways for visual contrast measurement," *Biological Cybernetics*, vol. 81, pp. 515–532, 1999.

[37] J. J. McAnany and M. W. Levine, "The vanishing disk: a revealing quirk of the scintillating grid illusion," *Journal of Vision*, vol. 2(7), pp. 204a, 2002.

[38] Rufin VanRullen and Timothy Dong, "Attention and scintillation," *Vision Research*, vol. 43, pp. 2191–2196, 2003.

[39] C. Gilbert, M. Ito, M. Kapadia, and G. Westheimer, "Interactions between attention, context and learning in primary visual cortex," *Vision Research*, vol. 40, pp. 1217–1226, 2000.

[40] M. Schrauf and L. Spillmann, "The scintillating grid illusion in stereo-depth," *Vision Research*, vol. 40, pp. 717–721, 2000.

[41] S. Tolansky, *Optical Illusions*, Pergamon, London, 1964.

[42] M.J. Morgan, "The poggendorff illusion: a bias in the estimation of the orientation of virtual lines by second-stage filters," *Vision Research*, vol. 39, pp. 2361–2380, 1999.

[43] L. M. Martinez, J. Alonso, R. C. Reid, and J. A. Hirsch, "Laminar processing of stimulus orientation in cat visual cortex," *Journal of Physiology*, vol. 540.1, pp. 321–333, 2002.

[44] J. Alonso and L. M. Martinez, "Functional connectivity between simple cells and complex cells in cat striate cortex," *Nature Neuroscience*, vol. 1, pp. 395–403, 1998.

[45] R. H. Carpenter and C. Blakemore, "Interactions between orientations in human vision," *Experiment Brain Research*, vol. 18, pp. 287–303, 1973.

[46] Colin Blakemore and Elisabeth A. Tobin, "Lateral inhibition between orientation detectors in the cat's visual cortex," *Experiment Brain Research*, vol. 15, pp. 439–440, 1972.

[47] C. Y. Li, Y. X. Zhou, X. Pei, I. Y. Qiu, C. Q. Tang, and X. Z. Xu, "Extensive disinhibitory region beyond the classical receptive field of cat retinal ganglion cells," *Vision Research*, vol. 32, pp. 219–228, 1992.

[48] H. Kolb and R. Nelson, "Off-alpha and off-beta ganglion cells in the cat retina," *Journal of Comparative Neurology*, vol. 329, pp. 85–110, 1993.

[49] Colin Blakemore, Roger H.S. Carpenter, and Mark A. Georgeson, "Lateral inhibition between orientation detectors in the human visual system," *Nature*, vol. 228, pp. 37–39, 1970.

[50] B. Chapman, M. P. Stryker, and T. Bonhoeffer, "Development of orientation preference maps in ferret primary visual cortex," *Journal of Neuroscience*, vol. 16, pp. 6643–6653, 1996.

[51] Á. Zarándy, L. Orzó, E. Grawes, and F. Werblin, "CNN-based models for color vision and visual illusions," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 46, pp. 229–238, 1999.

[52] J. O. Robinson, *The Psychology of Visual Illusion*, Dover, Mineola, NY, 1998.

[53] C. Q. Howe, Z. Yang, and D. Purves, "The Poggendorff illusion explained by natural scene geometry," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 7707–7712, 2005.

[54] Barbara Gillam, "A depth processing theory of the Poggendorff illusion," *Perception and Psychophysics*, vol. 10, pp. 211–216, 1971.

[55] Barbara Gillam, "Geometric illusions," *Scientific American*, vol. 242, pp. 102–111, 1980.

[56] C. Fermüller and H. Malm, "Uncertainty in visual processes predicts geometrical optical illusions," *Vision Research*, vol. 44, pp. 727–749, 2004.

[57] David H. Hubel and Torsten N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology (London)*, vol. 160, pp. 106–154, 1962.

[58] B. Roska, E. Nemeth, and F. Werblin, "Response to change is facillitated by a three-neuron disinhibitory pathway in the tiger salamander retina," *Journal of Neuroscience*, vol. 18, pp. 3451–3459, 1998.

[59] M. J. Frech, J. Perez-Leon, H. Wassle, and K. H. Backus, "Characterization of the spontaneous synaptic activity of amacrine cells in the mouse retina.," *Journal of Neurophysiology*, vol. 86, pp. 1632–1643, 2001.

[60] Yingwei Yu and Yoonsuck Choe, "Selection in time: an extended model for simulus onset asynchrony (SOA) in Stroop task," in *IEEE Development and Learning Conference 2006 ICDL06*, Bloomington, IN, 2006.

[61] David LaBerge, *Attentional Processing: The Brain's Art of Mindfulness*, Harvard University Press, Cambridge, MA, 1995.

[62] Gene Johnson, "Pathways to consciousness: the thalamus as the brains's switching centre," *Science and Consciousness Review*, vol. 2004:2, April 2004.

[63] S.L. Feig R.W. Guillery and D.A. Lozsadi, "Paying attention to the thalamic reticular nucleus," *Trends in Neurosciences*, vol. 21, pp. 28–32, 1998.

[64] Mark A. Pinsk Daniel H. O'Connor, Miki M. Fukui and Sabine Kastner, "Attention modulates responses in the human lateral geniculate nucleus," *Nature Neuroscience*, vol. 5, no. 11, pp. 1203–1209, 2002.

[65] Yoonsuck Choe and Yingwei Yu, "Role of the thalamic reticular nucleus in selective propagation of the results of cortical computation," in *World Association of Modelers Biologically Accurate Modeling Meeting WAM-BAMM05*, San Antonio, TX, 2005.

[66] S. Yantis and John T Serencesy, "Cortical mechanisms of space-based and object-based attentional control," *Current Opinion in Neurobiology*, vol. 13, pp. 187–193, 2003.

[67] GC Baylis and J. Driver, "Visual parsing and response competition: the effect of grouping factors," *Perception and Psychophysics*, vol. 51, pp. 145–162, 1992.

[68] J. Duncan, "Selective attention and the organization of visual information," *Journal of Experimental Psychology: General*, vol. 113, pp. 501–517, 1984.

[69] R. Egly, J. Driver, and RD. Rafal, "Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects," *Journal of Experimental Psychology: General*, vol. 123, pp. 161–177, 1994.

[70] M. I. Posner, "Orienting of attention," *Quarterly Journal of Experimental Psychology*, vol. 32, pp. 3–25, 1980.

[71] S. Shomstein and S. Yantis, "The role of strategic scanning in object-based attention (abstract)," *Journal of Vision*, vol. 2, pp. 437a, http://journalofvision.org/2/7/437/, doi:10.1167/2.7.437., 2002.

[72] R. M. Klein and D. I. Shore, "Relationships among modes of visual orienting," in *Attention and Performance XVIII: Control of Cognitive Processes*, S. Monsell and J. Driver, Eds., pp. 195–208. MIT Press, Cambridge, MA, 2000.

[73] L. Chelazzi, E.K. Miller, J. Duncan, and R. Desimone, "A neural basis for visual search in inferior temporal cortex," *Nature*, vol. 363, pp. 345–347, 1993.

[74] D. LaBerge, M. Carter, and V. Brown, "A network simulation of thalamic circuit operations in selective attention," *Neural Computation*, vol. 4, pp. 318–331, 1992.

[75] Yuhong Jiang and Marvin M. Chun, "The influence of temporal selection on spatial selection and distractor interference: an attentional blink study," *Journal of Experimental Psychology*, vol. 27, pp. 664–679, 2001.

[76] J.R. Stroop, "Studies of interference in serial verbal reactions," *Journal of Experimental Psychology*, vol. 18, pp. 643–662, 1935.

[77] Colin M. MacLeod, "Half a century of research on the stroop effect: an integrative review," *Psychological Bulletin*, vol. 109, no. 2, pp. 163–203, 1991.

[78] F.N. Dyer, "The duration of word meaning responses: stroop interference for different preexposures of the word," *Psychonomic Science*, vol. 25, pp. 229–231, 1971.

[79] M. O. Glaser and W. R. Glaser, "Time course analysis of the stroop phenomenon," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 8, pp. 875–894, 1982.

[80] J. D. Cohen, K. Dunbar, and J. L. McClelland, "A parallel distributed processing account of the stroop effect," *Psychological Review*, vol. 97, pp. 332–361, 1990.

[81] J.D. Cohen and T.A. Huston, "Progress in the use of interactive models for understanding attention and performance," in *Attention and Performance XV: Conscious and Nonconscious Information Processing*, C. Umilta and M. Moscovitch, Eds., pp. 453–456. MIT Press, Cambridge, MA, 1998.

[82] Matthew Botvinick, T. S. Braver, D. M. Barch, Cameron S. Carter, and Jonathan D Cohen, "Conflict monitoring and cognitive control," *Psychological Review*, vol. 108, pp. 623–652, 2001.

[83] R. H. Phaf, A.H.C. Van Der Heijden, and P. T. W. Hudson, "SLAM: a connectionist model for attention in visual selection tasks," *Cognitive Psychology*, vol. 22, pp. 273–341, 1990.

[84] A. Roelofs, "Goal-referenced selection of verbal action: modeling attentional control in the stroop task," *Psychological Review*, vol. 110, pp. 88–125, 2003.

[85] Matthew Botvinick, Leigh E. Nystrom, Kate Fissell, Cameron S. Carter, and Jonathan D Cohen, "Conflict monitoring versus selection-for-action in anterior cingulate cortex," *Nature*, vol. 402, pp. 179–181, 1999.

[86] David LaBerge, "Attention, awareness, and the triangular circuit," *Consciousness and Cognition*, vol. 6, pp. 149–181, 1997.

[87] Randall C. O'Reilly and Michael J. Frank, "Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia," *Neural Computation*, vol. 18, pp. 283–328, 2006.

[88] K. L. Shapiro, J.E. Raymond, and K.M. Arnell, "Attention to visual pattern information produces the attentional blink in rapid serial visual representation," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 20, pp. 357–371, 1994.

[89] M.M. Chun, "Temporal binding errors are redistributed by attentional blink," *Perception and Psychophysics*, vol. 59, pp. 1191–1199, 1997.

[90] D. E. Broadbent, *Perception and communication*, Pergamon Press, London, 1958.

[91] D. A. Norman, "Towards a theory of memory and attention," *Psychological Review*, vol. 75, pp. 522–536, 1968.

[92] M. Cohen and S. Grossberg, "Absolute stability and global pattern formation and parallel memory storage by competitive neural networks," *IEEE Transactions on Systems Man Cybernetics*, vol. SMC-13, pp. 815–826, 1983.

[93] Hui Ye, A. N. Michel, and Kaining Wang, "Qualitative analysis of cohen-grossberg neural networks with multiple delays," *Physical Review E*, vol. 51, no. 3, pp. 2611–2618, 1995.

[94] Tianping Chen and Libin Rong, "Delay-independent stability analysis of cohen-grossberg neural networks," *Physics Letters*, vol. A 317, pp. 436–449, 2003.

[95] Hassan K. Khalil, *Nonlinear Systems*, Prentice Hall, Upper Saddle River, NJ, 2002.

[96] Lin Wang and Xingfu Zou, "Exponential stability of Cohen-Grossberg neural networks," *Neural Networks*, vol. 15, no. 3, pp. 415–422, 2002.

[97] M. Fiedler and V. Ptak, "On matrices with nonnegative off-diagonal elements and positive principal minors," *Czech. Math. J.*, vol. 12, pp. 382–400, 1962.

[98] Eric R. Kandel, James H. Schwartz, and Thomas M. Jessell, *Principles of Neural Science*, McGraw-Hill, Health Professions Division, New York, 2000.

[99] Yoonsuck Choe, "How neural is the neural blackboard architecture?," *Behavioral and Brain Sciences*, vol. 29, 2006.

[100] Yingwei Yu and Yoonsuck Choe, "A neural model of scintillating grid illusion: disinhibition and self-inhibition in early vision," *Neural Computation*, vol. 18, pp. 501–524, 2006.

[101] Yingwei Yu and Yoonsuck Choe, "Asymptotic stability analysis of the thalamocortical circuit," in *Society for Neuroscience Abstracts*. 2005, Washington, DC: Society for Neuroscience.

VITA

Yingwei Yu received his Bachelor of Engineering degree in computer science from Beihang University (formerly named Beijing University of Aeronautics and Astronautics) in 1997. He received his Doctor of Philosophy degree in computer science from Texas A&M University, College Station, in August 2006. His research interests include computer vision, neural networks, stability analysis of neural networks, machine learning, and pattern recognition.

Mr. Yingwei Yu may be reached at Department of Computer Science, Texas A&M University, TAMU-3112, Texas 77840. His email address is yingwei@tamu.edu.