Brown University

Honors Thesis in Cognitive Neuroscience

Computational Visual Modeling of the Hermann Grid Illusion

Charlie Holtz

Advisor: Thomas Serre

Submitted in partial fulfillment of requirements for the degree of Bachelors of Science with Honors in the Department of Cognitive, Linguistic, and Psychological Sciences.

April 15, 2018

Contents

List of Figures

0 Abstract

The Hermann Grid is a well known yet poorly understood optical illusion. The prominent explanation of the illusion is based on the interactions between ON-center OFF-surround retinal ganglion cells in the retina. Various studies have created different forms of the illusion that disprove this prominent theory as the sole source of the illusion. Many qualitative studies have found various other mechanisms that seem to contribute to the processes behind the illusion, but no quantitative and comprehensive study exists. Using human pyschophysics data as a baseline, this study shows that solely retinal ganglion cells nor simple V1 cells can account for the strength of the illusion. Rather, a recent computational model based on a contextual circuit of V1 orientation selective cells best accounts for illusory strength (Mély, Linsley, and Serre). The patterns of excitation and inhibition within the classical receptive field in V1 cells seems to be the primary mechanism underlying the illusion.

1 Introduction

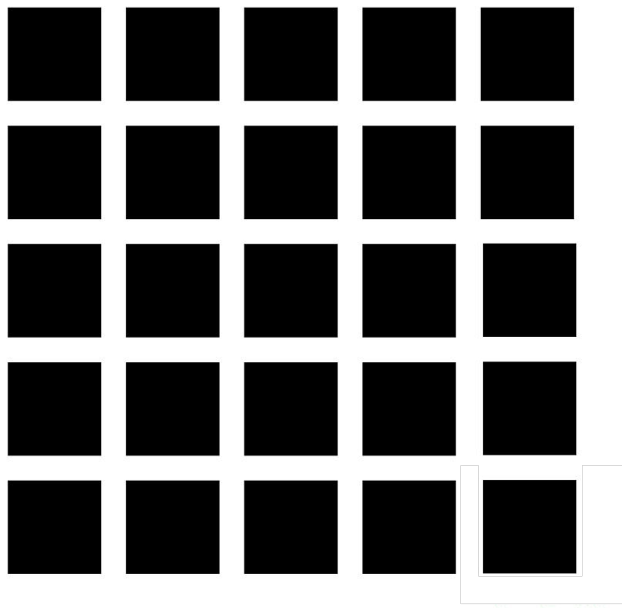

Figure 1: The Hermann Grid Illusion.

Source: Grid Illusion, Wikimedia Foundation

A series of equidistant black squares, when placed on a white background, leads to the perception of ghostly gray blobs at the intersections of those black squares (Figure 1). The blobs are seen at every intersection except for where the gaze is fixated. The eponymous illusion (henceforth referred to as the Hermann Grid) has puzzled scientists since its creation in 1870 by the German neuroscientist Ludimar Hermann (1870).

A much cited and debated illusion, the Hermann Grid appears in many introductory textbooks and neuroscience courses. The prominent explanation is based on the interactions between ON-Center OFF-Surround retinal ganglion cells in the retina (Hubel & Wiesel, 1968). But recent research has shown that this explanation is oversimplified.

While numerous theories have emerged in an attempt to explain the underlying processes, no complete explanation of the illusion exists. It is not clear which stage of the visual pathway is most involved in the perception of the illusion, nor why some variations of the Hermann Grid have stronger illusory power than others.

1.1 The Baumgartner Model

In (1960), Baumgartner devised the best known interpretation of the Hermann Grid Illusion. His explanation is based on the properties of retinal ganglion cells and is still the baseline explanation taught in introductory neuroscience courses today (Bear, Connors, & Paradiso, 2007).

To explain the Baumgartner Model, let's start with a brief reminder on the structure of retinal ganglion cells. Kuffler (1953) showed that ganglion cells in the retina can be broken down into two types: ON-center OFF-surround receptive fields and OFF-center ON-surround receptive fields. ON-center receptive fields are excited maximally by the presence of light in the center and darkness in the surround, and minimally to the presence of dark in the center and light in the surround of the receptive field (Figure 2).

The Baumgartner theory is based on the activities of these retinal ganglion cells.

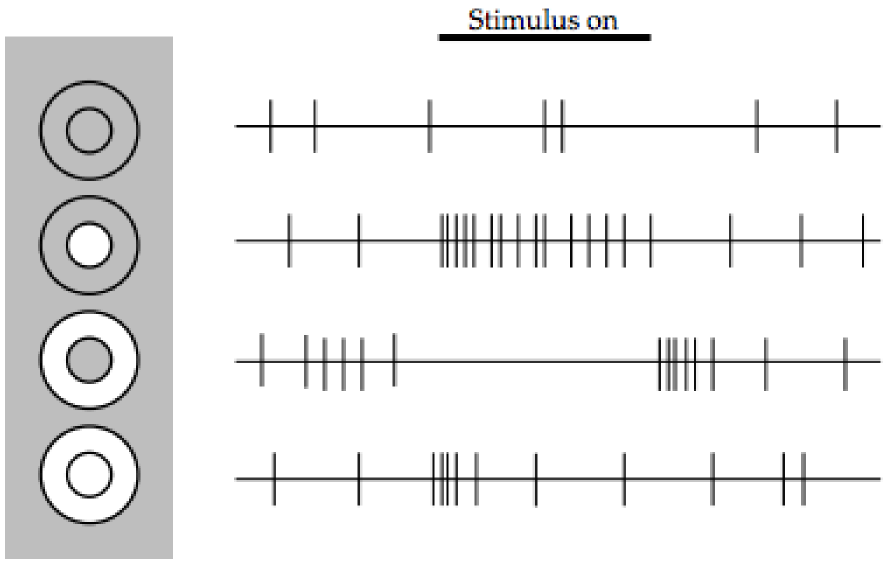

Figure 2: Center-ON Surround-OFF Retinal Ganglion Cell Receptive Field.

Source: Adapted From David Heeger, Perception Lecture Notes, NYU

When located at the alleyway between two boxes (position A), the ganglion cell's center field is excited by the presence of light (see Figure 3). The surround field of this ganglion cell is half darkness, half light, resulting in partial inhibition of center. As a result, we see lightness at position A. Now we move to position B, at the intersection of four squares. At position B, the ganglion cell's center field is equally as excited by the presence of light as at position A. However, position B's surround field is surrounded by twice as much white light as the ganglion cell at position A, resulting in a lower firing rate in the cell at position B. This leads to the perception that the white at the intersection is not as white as the white at the alleyway, and thus the perception of the grey blobs. When the illusion is presented with inverted colors, the Baumgartner explanation is inverted, and modeled on OFF-center ON-

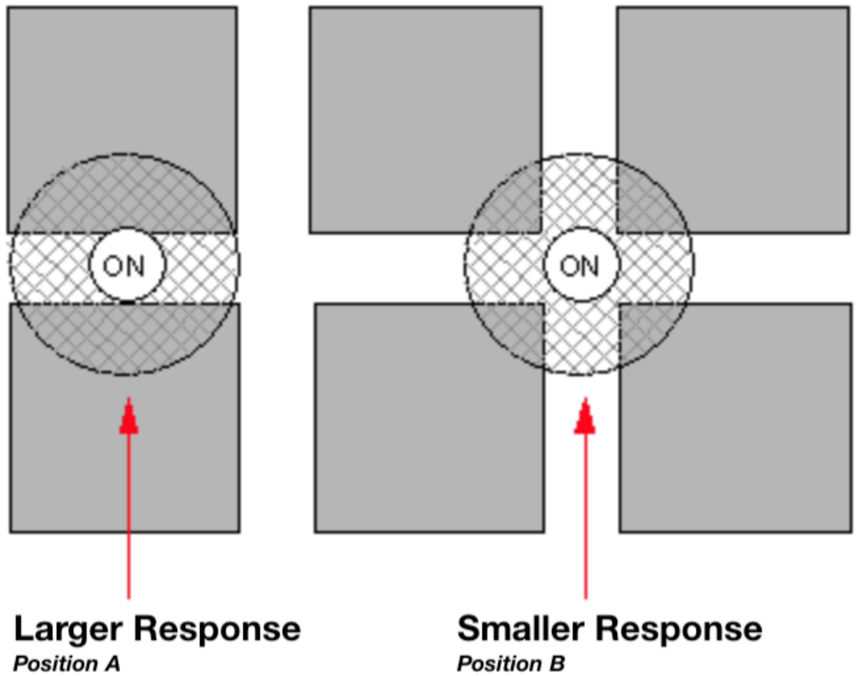

Figure 3: The Baumgartner Model.

Source: Adapted from the Neural Control of Vision, Schiller Lab, MIT

surround receptive fields. The reason that the blobs do not appear at the subject's fixation point is because the receptive fields within the fovea are small enough so that the entire range of stimulus occurs within the intersection of the squares.

1.2 Problems with the Baumgartner Model

The Baumgartner's simplicity and elegance has led to its ubiquity as a tool for understanding the Hermann Grid's illusory effects. However, there are three major problems with the Baumgartner explanation, as laid out by Schiller & Carvey (2005).

Firstly the illusory effects are diminished when the grid is rotated 45 degrees.

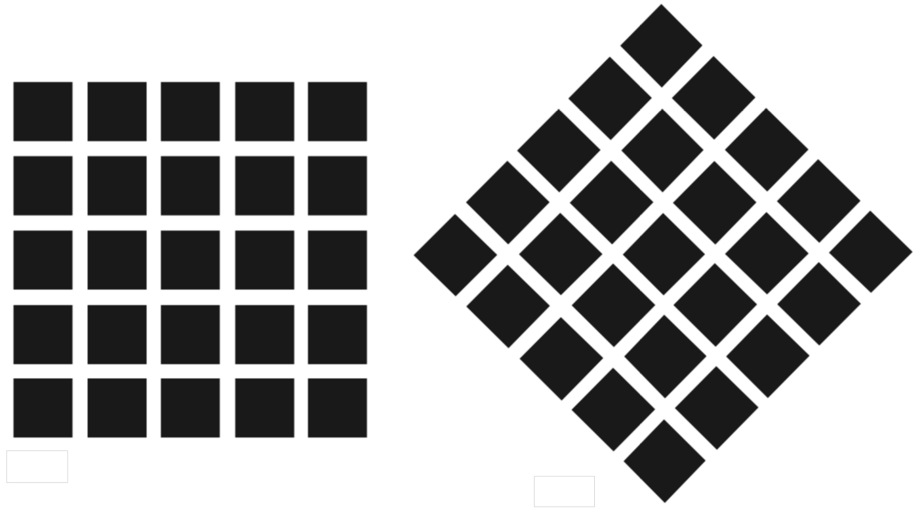

Figure 4: Classic Grid in comparison with Tilted Grid.

Source: (Schiller & Carvey, 2005)

Spillman and Jung (1970) showed that tilting the grid 45 degrees results in a less powerful illusion. One can see the difference most readily when both orientations are placed side-by-side (Figure 4). This "tilt effect" cannot be accounted for by the Baumgartner model because tilting the grid has no effect on receptive field size, location, or interactions.

The second problem is that enhancing the center/surround antagonism does not enhance the illusory effect. If the illusion was due to center/surround antagonism, the illusion on the right in Figure 5 (where center/surround antagonism is higher) should be stronger than the illusion on the left. Pyschophysics results show us that this is not the case (Schiller & Carvey, 2005). Variations of the illusion that should produce a high contrast response does not actually increase the strength of the illusion.

Thirdly, the illusory effects can be reduced or eliminated by manipulations that do

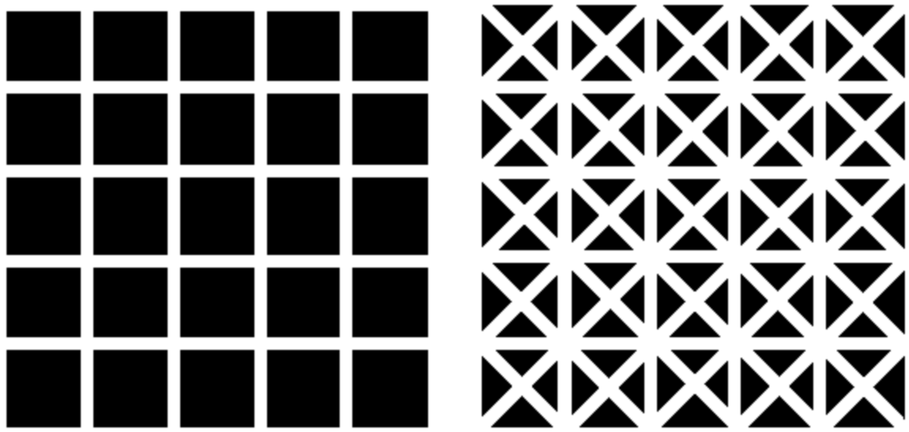

Figure 5: Increasing Center/Surround Antagonism.

Source: (Schiller & Carvey, 2005)

not alter the antagonistic center/surround activation in retinal ganglion cells. This is best demonstrated by the sinusoidal grid. Despite the same ratio of light to darkness in both the classic illusion and the sinusoidal variant, the illusory effect is diminished in a sinusoidal grid (Figure 6). According to the Baumgartner model, the illusory effects should be equal in Figure 6, but they are not.

Many other variations of the Hermann Grid have been created that provide further evidence that the Baumgartner Model is flawed (Wang, Hwang, & Lee, 2010), (Schiller & Carvey, 2005), (Vergeer & Lier, 2010). Among them are the sinusoidal grid, grids with curved edges, tilted grids, scintillating grids, and more (see Figure 6 for an overview of the variations). Each successive grid variation illuminates flaws with the Baumgartner model, but not without introducing new caveats.

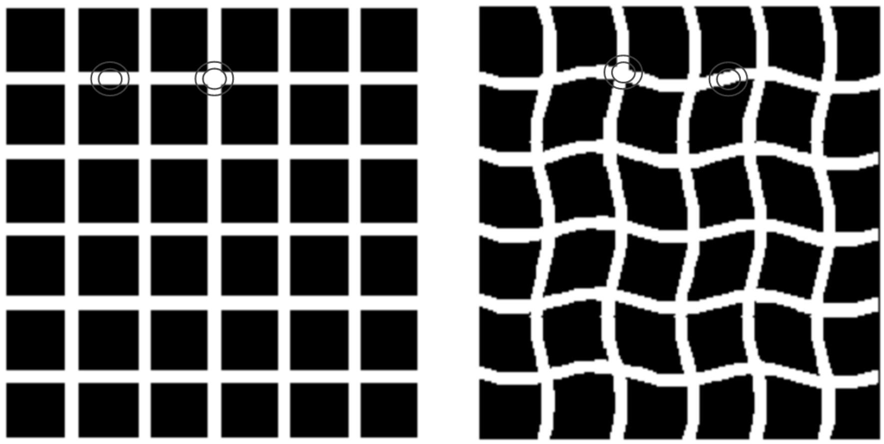

Figure 6: Classic Grid in comparison with Sinusoidal Grid.

Source: (Schiller & Carvey, 2005)

1.3 Alternative Models and Motivation

It is clear that the Baumgartner Model is not a satisfactory explanation for the illusory qualities of the Hermann Grid. Recent work has shown that the neural mechanisms involved may be occuring in the primary visual cortex, also known as V1 (Schiller & Carvey, 2005). Schiller and Carvey argue that interactions between orientation-selective V1 "simple" cells account for the illusory strength of the Hermann Grid. Geier et al. argue that the primary mechanism behind the illusory strength of the Grid is the straightness of the bars (Geier, Bernth, Hudk, & Sra, 2008). The lateral connections (both excitatory and inhibitory) between V1 simple cells may also be involved. Many more studies have been conducted, and many variations produced, each seeming to contradict another proposed solution. The field is cluttered. Although many qualitative studies have been performed, no satisfactory explanation exists.

Recent work in the Serre lab may yield another explanation to the mechanisms underlying the Hermann grid (Mély et al.). The proposed model accounts for the mechanisms of horizontal connections both within and across columns in receptive fields in V1. In doing so, their model takes into account the “context” surrounding the classical receptive field. It also models excitatory and inhibitory connectivity between receptive field columns. The contextual model is biologically constrained and accurately mirror psychophysics data in a variety of contextual illusions (see Discussion section). This begs the question of how accurate this generalized contextual model can replicate psychophysics results in the realm of the Hermann Grid, and brings us to the primary motivation behind this thesis: currently no comprehensive quantitative analysis of the mechanisms behind the Hermann Grid exists. Thus, by creating a biologically accurate model of V1, we can isolate which processes contribute to the illusion.

2 Methods

2.1 Psychophysics Experiment

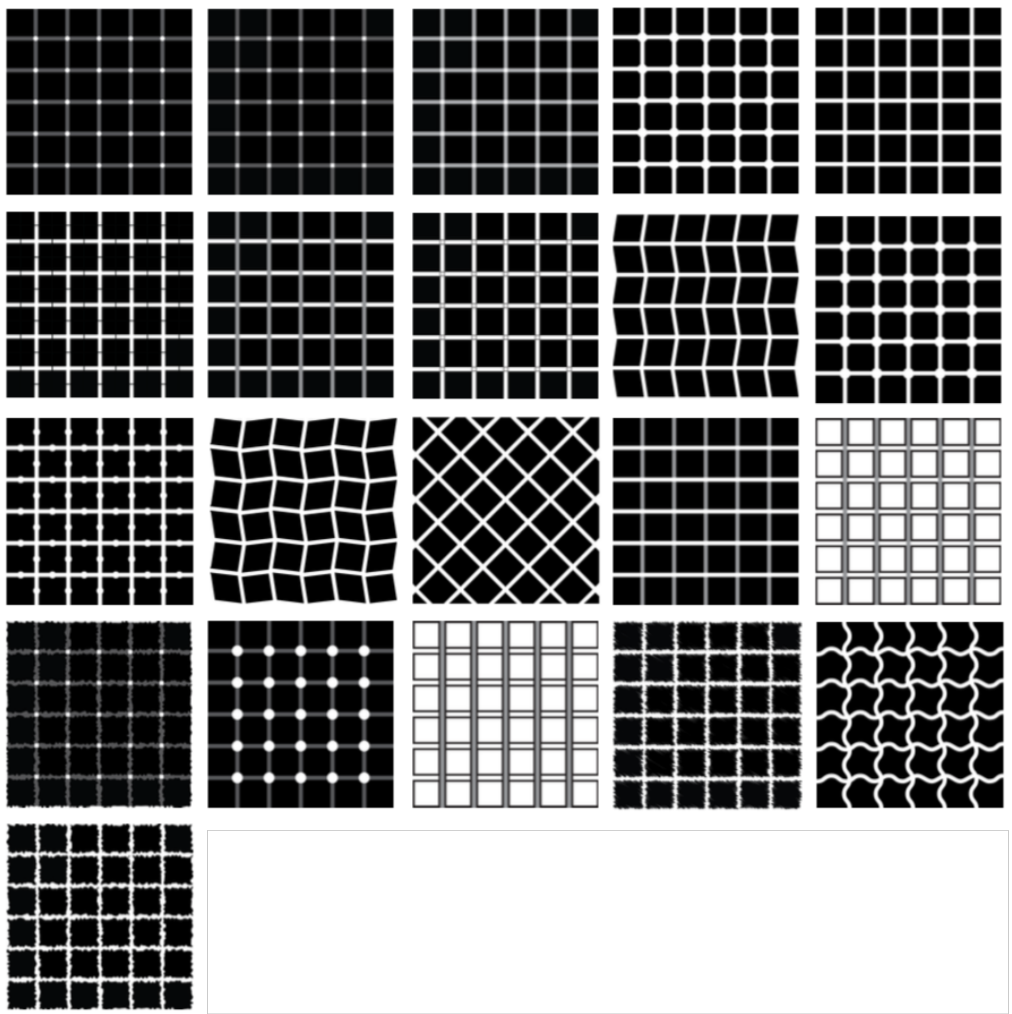

Figure 7: All Variations of the Grid.

Source: Computational Mechanisms Responsible for the Hermann Grid Illusion, R. Le, D. Mely, and T. Serre.

The psychophysics experiment was conducted by Rosemary Le '13. The experiment was conducted at Brown University in the Serre Lab. Twenty-one variations of the Hermann Grid were shown to subjects, two at a time. The twenty-one variations served as a representative for the many types of Hermann Grids. Figure 7 shows all the variations shown to subjects. The experiment consisted of showing each subject (n=20) two variations of the grid side-by-side. Then subjects were asked to rate from -5 to 5 (increments of 1) which illusion had a stronger illusory effect, with -5 meaning the illusion on the left was entirely stronger and 5 meaning the illusion on the right was entirely stronger (and 0 meaning that the illusory effects were equally as strong). Each subject was shown all possible pairwise combinations of illusions, for a total of 441 trials. For each of the twenty-one illusions, an overall average score was computed. This average served as illusion's overall illusory strength score.

The illusion subtended 16.47 degrees of visual angle. The width of the bars within the illusion subtended 0.4 degrees of visual angle.

2.2 Computational Models

All the computational models were written in python. Tensorflow, Google's machine learning library, was also used. The computations were run on lab computers fitted with Titan X GPUs. The only free parameter in the baseline mode was the Gabor filter sizes, which were consistent with biological findings at 17 x 17 pixels. We know this because we know the degrees of visual angle that the illusion subtended in the pyschophysics experiment, so we can estimate the Gabor filter size.

Alleyway and intersection coordinates were hard coded. Receptive field sizes were also fit to match primate V1 receptive field size ratios (Shushruth et al., 2013), with:

Classical Receptive Field (CRF) : 1x (16 pixels)

Near Extra-Classical Receptive Field (near eCRF) : 2x (30 pixels)

Far Extra-Classical Receptive Field (far eCRF) : 5.5x (88 pixels)

In the contextual model, free parameters were fit to generalized illusions. Model scores were calculated as the Pearson correlation between the model output for each image and the human baseline average score for that image.

Baseline Model Architecture:

Layer One Convolutional layer of Gabor filters in 8 orientations.

Layer Two Gather layer to compare brightness at intersection vs. alleyway.

Contextual Model Architecture:

Layer One Convolutional layer of Gabor filters in 8 orientations.

Layer Two Contextual normalization with center RF, near and far eCRFs.

Layer Three Gather layer to compare brightness at intersection vs. alleyway.

3 Results

The central goal of this study was to create a computational model that is both biologically constrained and that correlates with human psychophysics results. We started with a baseline model that modelled V1 orientation selective cells and then moved onto a contextual model that modelled classical and extra-classical V1 cells with connectivity.

3.1 Baseline Model

The baseline V1 model served as an approximation of V1 orientation selective cells. Using the average human scores from the pyschophysics experiment as a baseline, we found a 0.17 correlation between the baseline V1 model and the human scores. The Pearson correlation between subjects was 0.96 with a standard deviation of 0.019. The brightness activities in this model were generated simply by using a total of eight 17x17 pixel Gabor filters for eight different orientations.

Also included in this graph is a baseline retinal ganglion cell model. Previous work by David Mély and Rosemary Le in the Serre Lab found a correlation of 0.21 between a retinal ganglion cell model (Baumgartner Model) and human baseline scores.

3.2 Contextual Model

The contextual model included the same eight 17x17 pixel Gabor filters as in the baseline. Built on top of this layer was contextual circuit normalization (see Dis-

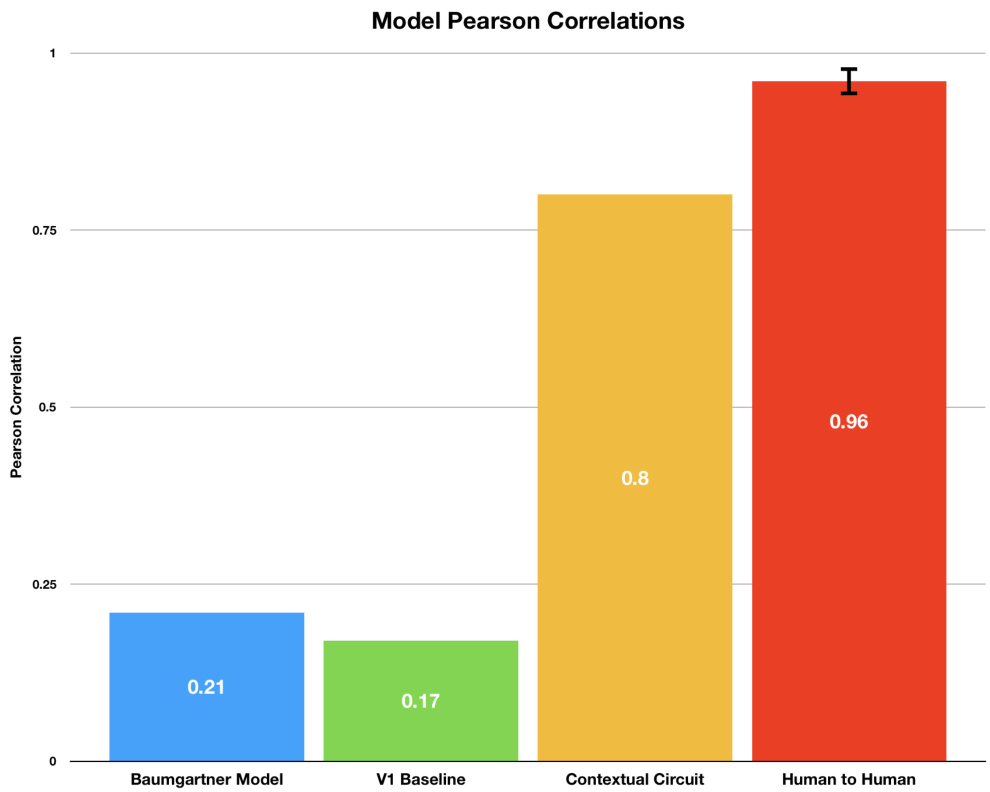

Figure 8: Contextual Model Results.

cussion section for further details). The center receptive field had a diameter of 16 pixels. With the near receptive field, the diameter extended to 30 pixels, and the far receptive field to 88 pixels. These receptive field sizes were kept in consistent proportion to the ratios shown in primate V1 areas (Shushruth et al., 2013).

With contextual circuit normalization, the model reached a correlation of 0.80 with the psychophysics results. This significantly beats the other models in terms of matching psychological results.

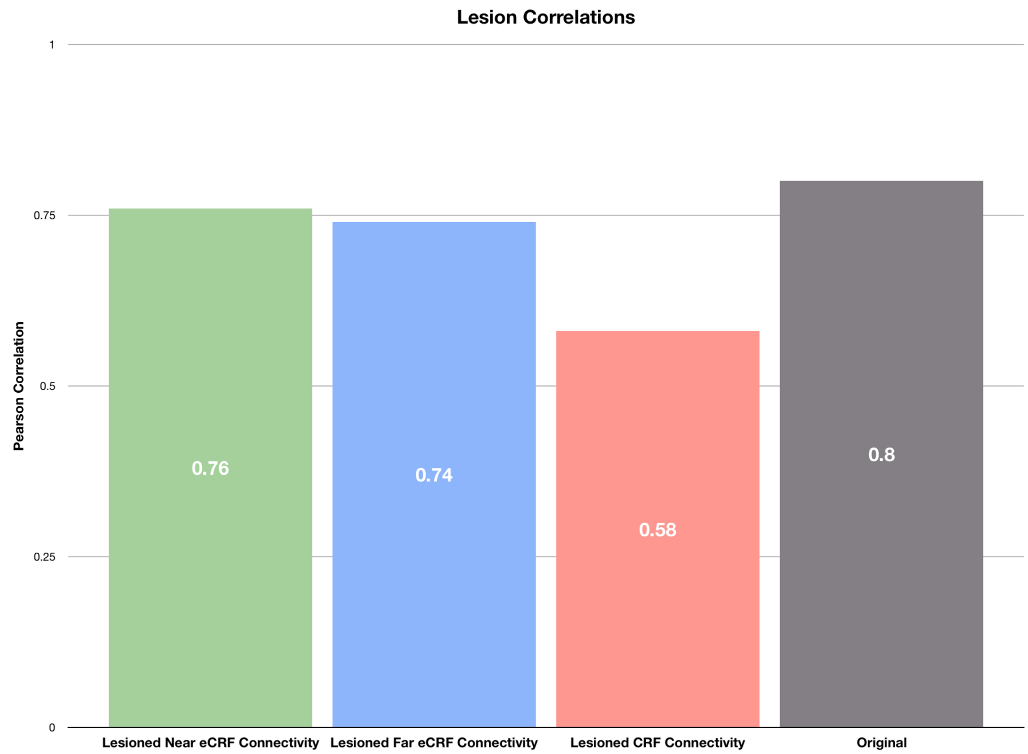

Figure 9: Lesioning Results.

3.3 Contextual Model with Lesions

In order to isolate which interactions contributed to the illusion, we lesioned specific processes within the contextual model. The contextual model with lesions was equivalent to the contextual model, but with artificial "lesions" to the connectivity within the CRF, excitation from the near eCRF, and inhibition from the far eCRF. Lesioning the inter-columnar connectivity from the near eCRF to the CRF resulted in a slight drop in correlation with psychophysics results, from 0.80 to 0.76 with lesion. Lesioning the inter-columnar connectivity from the far eCRF to the CRF had similar results—a drop from 0.80 to 0.74. Lesioning intra-columnar connectivity within the

CRF resulted in the most significant drop, from 0.80 to 0.58. This means that the intra-columnar connectivity within the CRF was qualitatively more responsible for explaining the processes underlying the illusion.

4 Discussion

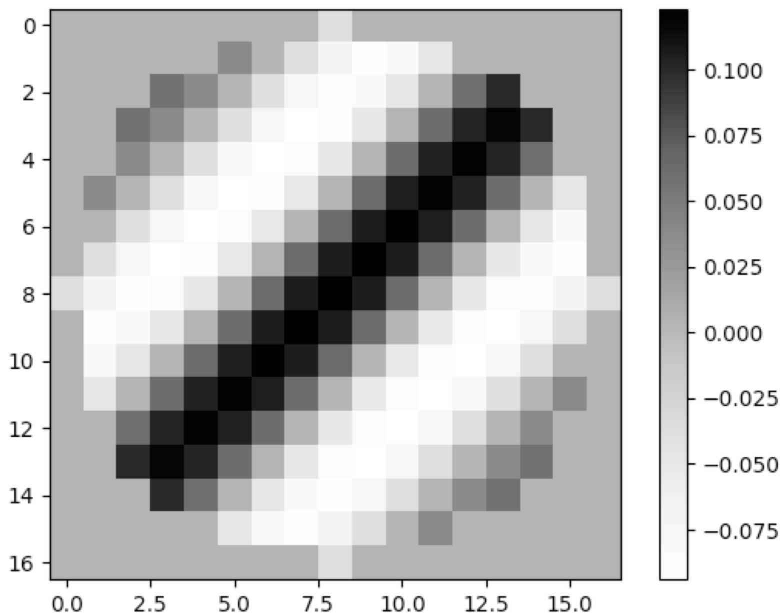

Figure 10: A Gabor Filter.

The power of computational models of vision is that we can isolate specific processes of biological vision. For example, Gabor filters serve as an accurate proxy for V1 receptive fields in mammalian vision systems (Cho et al., 2012). Figure 10 shows an example of the 17 x 17 pixel Gabor filters used in this experiment. The middle band and the outer bands are most tuned in response to light.

The baseline V1 model was created as a biological equivalent of isolating V1 simple cells. This model had a Pearson correlation with human results of only 0.17. This means that the mechanisms underlying the perception of illusory strength in humans must not be solely based on V1 simple cells with no interactions. Otherwise, this model would correlate highly with the human pyschophysics results. We can say the same about the retinal ganglion cell model (based on difference-of-gaussian receptive fields rather than Gabor filters). Low correlation with human results is evidence that the models are not biologically relevant.

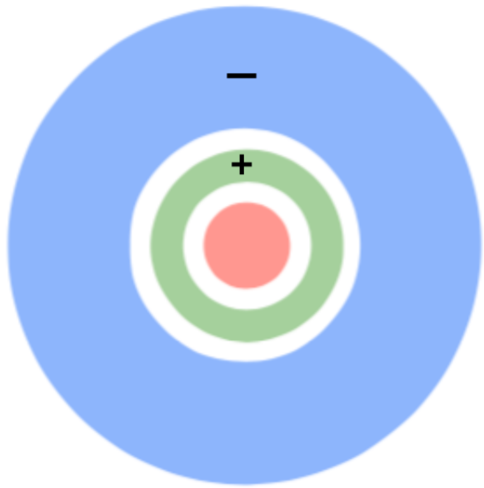

Figure 11: Classical (Red) and Extra-Classical Receptive Fields (Blue and Green).

Source: Opponent surrounds explain diversity of contextual phenomena across visual modalities, by D. Mely, D. Linsley, and T. Serre. In press, Psychological Review.

This brings us to the contextual model, which was mentioned briefly in the introduction (Mély et al.).The premise of the model is based on the existence of two “extra-classical receptive fields” (eCRFs) surrounding the classical receptive field (CRF). The near eCRF is primarily excitatory (green in Figure 11), and the far eCRF is primarily inhibitory (blue in Figure 11). In addition to this assumption, the contextual model makes three further assumptions about the processes underlying V1 cells. First, the pattern of connectivity from the surround to center is tuned. This means that the tuning to a particular orientation, color, movement pattern, etc., is kept constant across fields. Secondly, the facilitatory contribution of the near eCRF is linear, and the suppressive contribution of the far eCRF is non-linear. Finally, there is a pattern of excitation and inhibition within or between each receptive field's cortical columns. See Figure 12 for an illustration of this pattern.

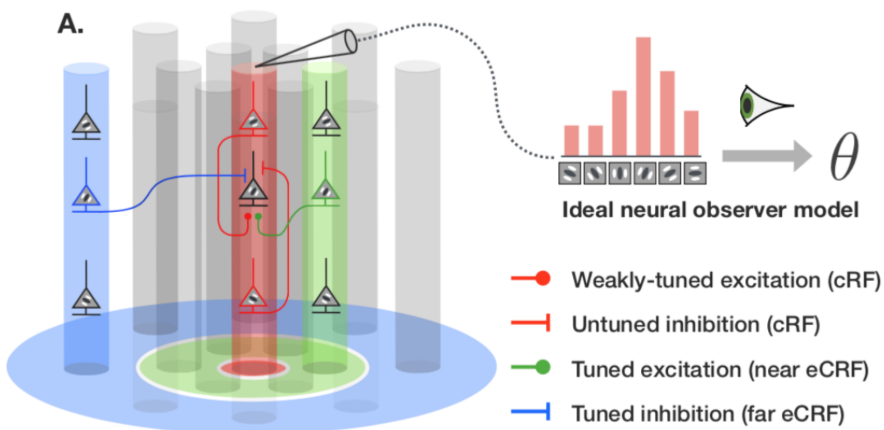

Figure 12: A Graphical Representation of the Contextual r Receptive Field Model.

Source: Opponent surrounds explain diversity of contextual phenomena across visual modalities, by D. Mely, D. Linsley, and T. Serre. In press, Psychological Review.

A computational implementation of this contextual circuit theory has shown to account for a variety of contextual illusions. In other words, the model is able to closely mirror human pyschophysics data across different types of illusions. For example, the tilt illusion refers to the phenomenon that the perceived orientation of a grating pattern is altered by the presence of a surround grating with a different orientation (Gibson, 1937). See Figure 13 for an illustration of both the tilt illusion (A) and also the comparison between psychophysics results and the models results (C).

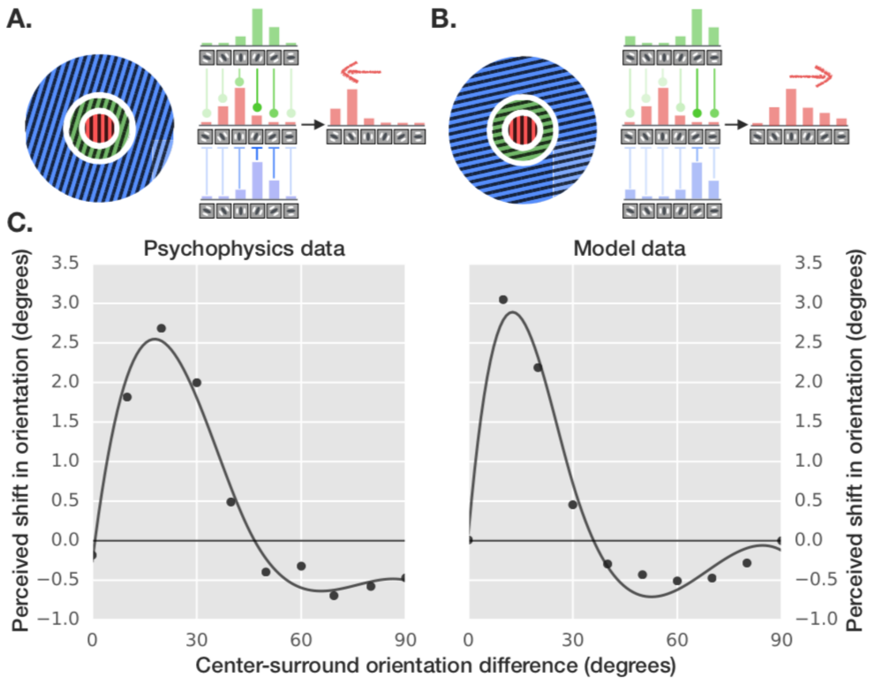

Figure 13: Pyschophysics vs. Model Results in the Contextual Circuit.

Source: Opponent surrounds explain diversity of Contextual Phenomena Across Visual Modalities, by D. Mely, D. Linsley, and T. Serre. In press, Psychological Review.

But the model is generalizable to more than just tilt illusions—it also accounts for illusions in depth, motion, and color domains. The model was not applied to the domain of the Hermann Grid Illusion until this thesis.

Because the contextual model's free parameters were tuned to a generalizable set of contextual illusions and not specifically the Hermann Grid, it is significant that the model was able to reach a correlation score of 0.80 with the psychophysics results. This means that some mechanism within the contextual model is accounting for the illusory strength far better than the simple V1 model could (0.80 vs. 0.17). In order to isolate these mechanisms, we can lesion the model. By lesioning different interactions within the contextual normalization layer, we can pinpoint which aspects of the model are most critical for the perception of the illusion.

For example, in Figure 12, the red cylinder emerging from the CRF represents the CRF's cortical column. Within this column there is a combination of both inhibitory and excitatory connections. All the connectivity within the crf is intra-columnar: the feature activation at a location is influenced by the activation of other features at the same location. The excitation within the CRF is weakly tuned to orientation, and the inhibition is untuned to orientation. The connectivity in the near eCRF and far eCRF columns are inter-columnar, extending from either the near or far eCRF and respectively exciting or inhibiting activity in the CRF (both tuned to orientation).

So, "lesioning" in the computational sense consists of reducing all connectivity within any given column to zero. For example, in order to lesion the excitation and inhibition within the near eCRF column, we replace all activities occurring in our tuned summation layer with zero.

We can see in Figure 9 that lesioning the connectivity emerging from the near eCRF and far eCRF resulted in little change in correlation with psychophsyics results. Lesioning the inhibitory connections emerging from the far eCRF resulted in correlation score of 0.76 (down from .80). Lesioning the excitatory connections emerging from the near eCRF resulted in a correlation score of 0.74.

Lesioning the intra-columnar connectivity within the CRF resulted in the most significant drop, from 0.80 to 0.58. CRF patterns of connections were included in the model for explaining neural responses to contextual stimuli. They were not considered crucial to explain many of the idiosyncratic perceptual phenomena elicited by contextual illusions. This makes the importance of CRF mechanisms in explaining the Hermann Grid illusion all the more surprising. In other words, our model predicts that the Hermann Grid illusory effects are from nonlinearities in neural responses that can be explained by intra-columnar rather than inter-columnar interactions.

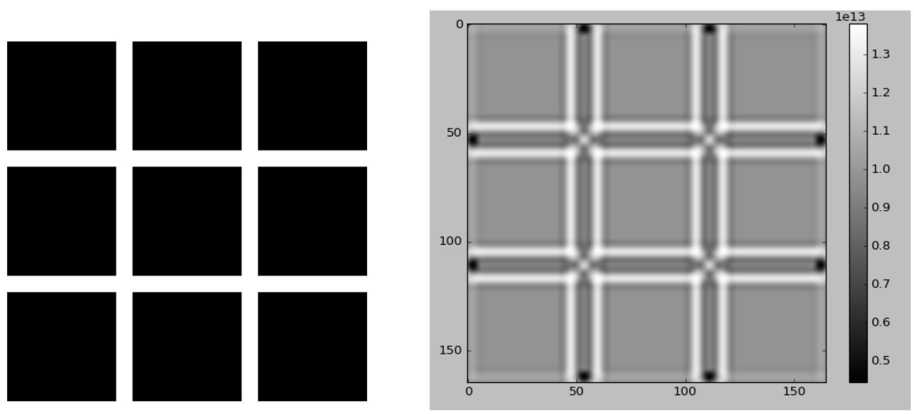

Figure 14: Brightness Activities Generated by the Contextual Model.

4.1 Further Study

The current model uses specified gather points at the intersections and alleyways of the images indicating which areas within the image to compare brightness. Further work will involve creating trainable models that may learn where on the Grid to compare brightness ratios. It may also be interesting to run further models with free parameters that are tuned to the Hermann Grid rather than the generalized set of contextual illusions that this model is currently tuned to. Finally, the sample size

(20) of this psychophysics experiment was limited and further study could elaborate on the number of subjects and variety of Grid Illusions represented in the study.

4.2 Concluding Thoughts

The processes behind the Hermann Grid Illusion cannot be accounted for solely by retinal ganglion cells or simple V1 cells. This study shows that the processes must be cortical in nature. Specifically, it seems as if the primary force behind the illusion is the pattern of excitation and inhibition within the center classical receptive field. A wide range of variations of the Hermann Grid can be accounted for by this quantitative contextual model. Further research should be done to better understand these patterns of activity occurring within the center classical receptive field.

This study also reminds us of one the clearest and most practical benefits of a computational model: we can precisely isolate mechanisms and perform infinite experiments on our in silico subjects. The intersection of neuroscience and computation provides invaluable abilities—How else could we gain the ability to lesion specific interactions in a neurological process? The only limits of computational models are the size of our GPUs and our understanding of biology.

Vision is an ill posed problem. At the lowest levels, our visual system measures the frequency and wavelength of light. But our visual systems go much further. We take into account context, adapt to changes in the position of light sources, predict depth, perceive dimensionality, track motion. Our visual systems perform incalculably complex computations in order to provide a consistent representation of reality. These calculations feel seamless. The only way our visual system can perform these calculations so swiftly is by making assumptions.

We can find out what these assumptions are by studying optical illusions. Optical illusions show us the edges of perception: they teach us where the limits lie, where the brain resorts to assumptions to keep our perception of the world constant. The function of perception is to decode reality into sensory information that the brain, and thus the mind, can understand. But we know through illusions that the translation between reality and our perception of that reality is not direct. Studying illusions like the Hermann Grid shows us where that information gets lost in translation.

5 Acknowledgements

My main contribution to this project was adapting the contextual circuit model to process the Hermann Grid. Fitting a baseline model was the initial step, followed by the contextual circuit model/lesioned models. Also involved in the research was parsing the data previously recorded and recreating Gabor filters from previous experiments in the lab (performed by Rosemary Le '13, with analysis done by David Mély.)

I'd like to thank Drew Linsley for all his help and patience, and Thomas Serre for his comments and guidance on this project.

References

Baumgartner, G. (1960). Indirekte grössenbestimmung der rezeptiven felder der retina beim menschen mittels der hermannschen gittertäuschung. *Pflugers Archiv für die Gesamte Physiologie des Menschen und der Tiere*, *272*(1), 2122. doi: 10.1007/bf00680926

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). *Neuroscience: exploring the brain.* Lippincott Williams & Wilkins.

Geier, J., Bernth, L., Hudk, M., & Sra, L. (2008). Straightness as the main factor of the hermann grid illusion. *Perception*, *37*(5), 651665. doi: 10.1068/p5622

Gibson, J. J. (1937). Adaptation, after-effect, and contrast in the perception of tilted lines. ii. simultaneous contrast and the areal restriction of the after-effect. *Journal of Experimental Psychology*, *20*(6), 553569. doi: 10.1037/h0057585

Hermann, L. (1870). Eine Erscheinung simultanen Contrastes. *Pflgers Archiv fr die gesamte Physiologie*, *3*, 13-15.

Hubel, D. H., & Wiesel, T. N. (1968, Jan). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215243. doi: 10.1113/jphysiol.1968.sp008455

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*(1), 3768. doi: 10.1152/jn.1953.16.1.37

Mély, D., Linsley, D., & Serre, T. (n.d.). Opponent surrounds explain diversity of contextual phenomena across visual modalities. *In press in Psychological*

Review.

Schiller, P. H., & Carvey, C. E. (2005). The hermann grid illusion revisited. *Perception*, *34*(11), 13751397. doi: 10.1068/p5447

Shushruth, S., Nurminen, L., Bijanzadeh, M., Ichida, J. M., Vanni, S., & Angelucci, A. (2013, Feb). Different orientation tuning of near- and far-surround suppression in macaque primary visual cortex mirrors their tuning in human perception. *Journal of Neuroscience*, *33*(1), 106119. doi: 10.1523/jneurosci.2518-12.2013

Spillman, L., & Junr, R. (1970). Receptive-field estimation and perceptual integration in human vision. *Early Experience and Visual Information Processing in Perceptual and Reading Disorders*, 181-197.

Vergeer, M., & Lier, R. V. (2010). Capturing lightness between contours. *Perception*, *39*(12), 15651578. doi: 10.1068/p6539

Wang, H.-W., Hwang, S.-H., & Lee, C. F. (2010). Effect of proportion on stopping hermann grid. *Second International Conference on Digital Image Processing*. doi: 10.1117/12.852362