

PERCEPTUAL GROUPING STRATEGIES IN VISUAL SEARCH TASKS

by

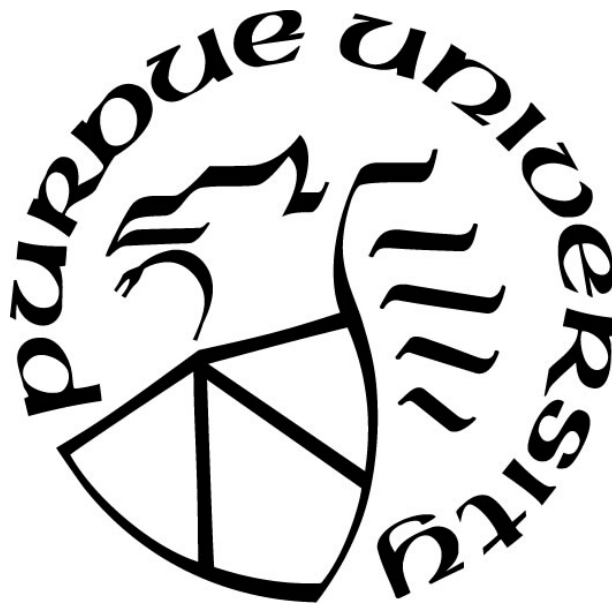
Maria Kon

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Psychological Sciences

West Lafayette, Indiana

May 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Gregory S. Francis, Chair

Department of Psychological Sciences

Dr. Sebastien Hélie

Department of Psychological Sciences

Dr. Darryl W. Schneider

Department of Psychological Sciences

Dr. Yu-Chin Chiu

Department of Psychological Sciences

Approved by:

Dr. Kimberly P. Kinzig

ACKNOWLEDGMENTS

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562, Jetstream at IU/TACC through allocation TG-SOC210003.

TABLE OF CONTENTS

LIST OF TABLES	8
LIST OF FIGURES	9
ABSTRACT	18
INTRODUCTION	19
PREVIOUS WORK: LAMINART	23
Proposed Developments	25
Conceptual Shift	25
Overview of Model Development	27
MODEL SUMMARY AND BOTTOM-UP CIRCUITS	29
Model Summary	29
Bottom-Up Circuits	31
TOP-DOWN CIRCUITS I: THE CONNECTION CIRCUIT	37
Spread Controller Circuit	37
Long and Short Controllers Overview	49
Long Controller Circuit	50
Short Controller Circuit	56
Combining Connection Controllers to Promote Gestalt Groupings	61
Proximity	65
Similarity of Orientation	65
Similarity of Size	65
Similarity of Shape	66
Closure	66
Symmetry	67
Summary	67
TOP-DOWN CIRCUITS II: THE SELECTION CIRCUIT	68
The Selection Circuit	70
Feature Filters	75
The Reset Signal	80
Selection Signal Size	82

Application to Other Stimulus Images.....	84
Summary	85
APPLICATION OF THE MODEL TO THREE EXPERIMENTS THAT USE VISUAL	
SEARCH TASKS TO INVESTIGATE GROUPING	88
General Experiment Methodology.....	88
Participants	88
Apparatus	89
Stimuli.....	89
Procedure	89
Simulation Methodology	89
PALMER AND BECK (2007)	92
Experiment 1: Replication of Palmer and Beck (2007)	95
Method	96
Participants	96
Apparatus	96
Stimuli	97
Procedure.....	97
Results and Discussion	99
Model Simulations of Experiment 1	101
Simulated Grouping Strategies	101
Connection strategies	101
Selection strategies	104
Target identification algorithm.....	106
Simulation Stimuli, Method, and Procedure.....	107
Model Results and Discussion.....	108
Experiment 2: Same Task, Different Stimulus Set	109
Method	111
Participants	111
Apparatus, stimuli, and procedure.....	111
Results and Discussion	111
Experiment 3: Controlled Row Width to Induce a Simpler Selection Strategy	114

Method.....	115
Participants	115
Apparatus, stimuli, and procedure.....	115
Results and Discussion	116
VICKERY (2008)	120
Experiment 4: Direct Replication of Vickery (2007)	122
Method.....	122
Participants	122
Apparatus	123
Stimuli	123
Procedure.....	123
Results and Discussion	124
Experiment 5: Nine Elements Per Row	125
Method.....	125
Participants	125
Apparatus, stimuli, and procedure.....	126
Results and Discussion	126
Model Simulations of Experiment 5	127
Simulated Grouping Strategies	127
Connection strategies	127
Selection strategies and target identification algorithm.	129
Simulation Stimuli, Method, and Procedure.....	131
Model Results and Discussion.....	131
TRICK AND ENNS (1997).....	135
Experiment 6: Replication of Trick and Enns (1997).....	138
Method.....	139
Participants	139
Apparatus	140
Stimuli	140
Procedure.....	141
Results and Discussion	141

Model Simulations of Experiment 6	147
Simulated Grouping Strategies	147
Connection strategies for dot figures	147
Connection strategies for line figures.....	151
Selection strategies	153
Target identification algorithm (counting).....	157
Simulation Stimuli, Method, and Procedure.....	157
Model Results and Discussion.....	159
GENERAL DISCUSSION AND CONCLUSIONS.....	165
REFERENCES	171

LIST OF TABLES

Table 1. Experiment 1 Descriptive Statistics and Correlations for Spacing Conditions	99
Table 2. Experiment 1 Descriptive Statistics and Correlations for Grouping Conditions.....	100
Table 3. Experiment 2 Descriptive Statistics and Correlations for Spacing Conditions	112
Table 4. Experiment 2 Descriptive Statistics and Correlations for Grouping Conditions.....	112
Table 5. Experiment 3 Descriptive Statistics and Correlations for Spacing Conditions	117
Table 6. Experiment 3 Descriptive Statistics and Correlations for Grouping Conditions.....	117
Table 7. Experiment 4 Descriptive Statistics and Correlations for Grouping Conditions.....	124
Table 8. Experiment 5 Descriptive Statistics and Correlations for Grouping Conditions.....	126
Table 9. Experiment 6 Descriptive Statistics.....	143
Table 10. Experiment 6 Slopes and Descriptive Statistics by Curve and Target Range	145
Table 11. Best Performing Grouping Strategy by Experiment.....	169

LIST OF FIGURES

Figure 1. A redrawing of Figure 1 in Köhler (1929, p.154).	19
Figure 2. Example of the bottom-up circuits created in LAMINART for the input of a white square line figure on a black background (panel a). Panels b1 and b2 are visualizations of these circuits sandwiched between the input image (bottom) and V1 output image (top). Each pixel in the input and output images is a grid cell. To provide a sense of its complexity, panel b1 shows the entire V1 network created for the input image shown in panel a. Panel b2 shows only a subset of this network, i.e., the circuits created for three pixels of the input, with the layers labelled. Panel c illustrates what the output represents: the spikes of orientation-tuned neurons in layer 2/3 of V1 (The output is an 8x8 image because an extra row and column are added in the simulation to avoid edge effects). In panel c, the cells elongated along the x-axis (y-axis) are sensitive to horizontal (vertical) edges. Note that this visualization of V1 is simplified: there are also cells tuned to diagonals at 45° and 135°, making the model have double the number of V1 cells that are represented here.	24
Figure 3. Example of the top-down segmentation process for the stimulus in the top row. The selection signal is represented by the translucent white circle in segmentation layer 0. Any boundaries directly under the selection signal or connected to boundaries under the selection signal are segmented out of layer 0. The sum of the activity over 50 milliseconds in each layer is represented in each non-stimulus image.	24
Figure 4. An example where a selection signal segments multiple unconnected stimulus elements and seems to thereby group the two far right lines.	26
Figure 5. An example to illustrate multiple ways in which the elements of a single image may be grouped.	27
Figure 6. An example to motivate the claim that groupings can be task-dependent.	28
Figure 7. A schematic overview of each component of the model.....	29
Figure 8. A schematization of the circuits representing V1 and V2. Left column: This illustrates cells and the synapses between them that are created for one pixel i in the input image. Here, there are cells tuned to two orientations, with the cells tuned to horizontal contours represented by the ovals elongated along the x-axis, and cells sensitive to vertical contours represented by ovals elongated along the y-axis. Right column: This shows the synapses that are between horizontally-tuned cells and their immediate neighbors, i.e., the bottom-up circuits created for neighboring pixels in the image. Excitatory synapses are represented by arrows, while an inhibitory connection is represented by a dashed line with a square head.	32
Figure 9. A demonstration of how the interneurons prevent runaway spread. The three green boxes on each stimulus image corresponds to three pixels, and the circuit diagrams for V2 horizontally-tuned cells are shown to their right.	33

Figure 10. An example to illustrate the AND-gate function of interneurons. The input image consists of three white squares on a black background. Each of the squares has a pixel-wide gap along the top edge. The three green squares on each white square indicate which pixel locations hat the circuits on the right correspond to..... 34

Figure 11. In the current version of the model, the dedicated pooling cells of the 2017 version have been replaced with grouping speed, which involves more synapses between nearby V2 bipole cells. Colors are used here to highlight the different synapse types that are added for each neighboring cell that is involved, the number of which is a function of grouping speed. 36

Figure 12. A schematized depiction of the entire Connection Circuit. As explained below, though, this is simplified for the sake of clarity: the Short Control Accumulator cell pools input from all V2, layer 4, cells tuned to the same orientation that are in positions $i \pm 5$ relative to the position of the simple cell at i . Similarly, the Long Control Accumulator cell pools activity from simple cells at positions $i \pm 24$. Each arrow from “Top-down connection control” represents the value of the connection parameters, i.e., onset and duration of the Spread Controller and a constant non-negative current into the Short and Long Controllers. The remainder of the arrows represent synapses: excitatory synapses are represented by arrows with triangular heads, while an inhibitory connection is represented by a dashed line with a square head. Color of synapses and cells indicate that they constitute a particular module of the Connection Circuit. 38

Figure 13. Horizontally-tuned V2 cells (ovals) and interneurons (gray circles) with the Spread Controller Circuit (green). The Spread Controller Circuit allows for top-down control of the spread of connections between bipole cells via the inhibition of the interneurons for a particular period of time. The longer the interneurons are inhibited by the Spread Controller cell, the farther the signal from active bipole cells can spread. 39

Figure 14. An illustration of the temporal dynamics of connections for a simulation where the Spread Controller cell is excited for a duration of 80 ms with onset at 0 ms. Each image shows the activity of V2 bipole cells for a 50 ms interval, e.g., the first image represents activity from the time period 50-100 ms after stimulus onset. 40

Figure 15. Altering the duration of excitation to the Spread Controller cell produces different connections between the detected edges for this stimulus. For each simulation, onset of top-down control excitation is at onset of the stimulus, i.e., at 0 ms. Here each image shows bipole cell activity summed over 50 ms at 1700 ms after stimulus onset. Each of these simulations was run for 1700 ms (model time) because this was the time at which the simulation with the longest Spread Controller duration (220 ms) settled down in to an equilibrium state. 42

Figure 16. Demonstration that the spread of V2 connections is orientation specific and, additionally, that each orientation has its own duration of top-down control. The images show bipole cell activity summed over 50 ms at 700 ms after stimulus onset. The top-down control excitation began at stimulus onset, i.e., at 0 ms. 44

Figure 17. An example of how different combinations of onset and duration of the top-down control excitation can produce the same patterns of connections. Each image shows bipole activity summed over 50 ms from 700-750 ms after stimulus onset. 45

Figure 18. An example of how varying the onset of the top-down control duration can produce different patterns of connections. The images in row 4 have different connection patterns. Each image shows bipole cell activity summed over 700-750 ms after stimulus onset. 46

Figure 19. Parameter maps describe the possible connection patterns for a small image of squares and circles (left) and H's and X's (right). A simulation was conducted for each pair of duration and onset plotted here, where duration and onset values ranged from 0 to 60 ms in increments of 1 ms. 47

Figure 20. The components of the Long Controller Circuit for a horizontally-tuned bipole cell at a particular pixel location are shown in orange. Although (for the sake of reducing clutter in the image) only two cells (bottom row) to the left and right of this bipole cell are shown here, the horizontal Long Control Accumulator cell pools activity from 24 horizontally-tuned complex cells to its left and right. 51

Figure 21. The connections resulting from simulations with different input to the Long Controller cell. The Long Controller Circuit is effectively off when the input to the Long Controller is very high, e.g., 1000, and has maximum restriction when the input is 0. For each simulation in the left column, Spread Controller onset was at 25 ms after stimulus onset and had a duration of 40 ms. For the simulations in the second column, Spread Controller onset was at 0 ms and the duration was 66 ms. 52

Figure 22. V2, layer 4, output from the grouping by orientation example to illustrate the difference between activity pooled by the Long Control Accumulator cell for a cell at the top of a vertical bar that is near the horizontal lines (right set of boxes) and that of a cell that is at the top of a vertical bar farther from the horizontal lines (left set of boxes). The white pair of boxes on the right overlaps with more active horizontally-tuned V2, layer 4 cells, and the Long Control Accumulator cell at the location indicated by the orange pixel at the center of these white boxes has stronger input since it pools from this set of cells. In contrast, the white pair of boxes on the left overlaps with far fewer and weaker horizontally-tuned cells. Since the Long Control Accumulator cell at the location indicated by the orange pixel between these white boxes is weakly excited, it does not excite the V2 interneurons as much. Thus, the Long Controller will more strongly excite the interneurons at the position on the right than the position on the left, and this prevents V2 connections forming between the far right vertical bars only given a Long Controller input of zero. 53

Figure 23. Parameter maps for the stimulus shown at the top of each plot. To provide a fine-grained picture of connections possible with different Long Controller inputs, Spread Controller duration ranged from 0 to 60 ms in increments of 1, and Long Controller input ranged from 0 to 3.50 in increments of 0.01. For all simulations here, the Spread Controller input was turned on 25 ms after stimulus onset. 55

Figure 24. The components of the Short Controller Circuit are shown in red for a horizontally-tuned bipole cell. Although, for the sake of simplicity, this schematization indicates that the Short Control Accumulator only pools the activity from two complex cells corresponding to locations neighboring that of the bipole cell (and the complex cell that feeds into the bipole cell), in the simulations each horizontal Short Control Accumulator cell pools activity from the five horizontally-tuned complex cells (bottom row) that are to the left and right of the bipole's pixel location. 57

Figure 25. Connections formed for different top-down inputs to the Short Controller cell. Each image shows the activity of bipole cells for 50 ms of model time after 700 ms of a simulation. For each simulation, the other connection parameters were fixed: Spread Controller duration was set to 40 ms (beginning at 25 ms), and the Long Controller Circuit was effectively turned off with an input of 1000. 58

Figure 26. The impact of different inputs to the Long Controller (left column) and Short Controller (right column) for the stimulus at the top. For each simulation, Spread Controller duration was 30 ms and Spread Controller onset was at 25 ms, i.e., Spread Controller parameters at which at least all similarly shaped pairs were connected. For the left column, which demonstrates the effect on connections of different Long Controller input values, the Short Controller (Long Controller) was turned off for each simulation by having the top-down Short Controller input set at 0. To demonstrate the effect of different Short Controller input values, the Long Controller was turned off for each simulation in the right column by having the top-down Long Controller input set at 1000. 59

Figure 27. Parameter maps of Short Controller input and Spread Controller duration for the stimulus shown at the top of each plot. To provide a fine-grained picture of connections possible with different Long Controller inputs, Spread Controller duration ranged from 0 to 60 ms in increments of 1, and Long Controller input ranged from 0 to 3.50 in increments of 0.01. For all simulations, Spread Controller onset was 25 ms after stimulus onset and, thus, the second column of Figure 26 provides some examples of parameter values and resulting connections that are shown in these maps. 61

Figure 28. Model output for two simulations that have a Spread Controller onset of 0 ms and a duration of 55 ms for a stimulus that can be grouped by similarity of size. For the model to form connections between all and only same-sized elements, the Short Controller must be on with an input of around 0.95 in order to prevent the connections that result between the pairs of same-sized circles shown in the middle image..... 62

Figure 29. Model output for four simulations with different combinations of connection parameter values. For each simulation, Spread Controller duration is 30 ms with an onset of 25 ms. The Long Controller input of 2.0 eliminates connections between the square and circle, while the Short Controller input of 0.8 eliminates connections between the H and X. By combining these connection parameters, the model can produce connections that effectively group the stimulus elements by shape as shown in the bottom image..... 63

Figure 30. Grouping rules that can be accounted for by tuning the top-down control parameters of the connection circuits. The first proximity stimulus (row 1) is similar to examples in Köhler (1925). The stimuli with filled shapes (rows 2, 6, and 8) and the orientation stimuli (rows 3 and 4) are adapted from examples of “classical principles of grouping” listed by Palmer (1999; 2002). The first closure stimulus (row 9) appears in Köhler (1925), and the second (row 10) is from Pomerantz and Kubovy (1986). The symmetry stimulus (row 11) is adapted from Palmer (1999; 2002). Spread Controller onset was at 0 ms for all stimuli. For the Long Controller input, “off” means that a large value (1000) is provided. For the Short Controller input, “off” means that a value of zero was provided. 64

Figure 31. A toy example stimulus, i.e., a row of letters (left image), and task, i.e., identify the location of the pair of repeated I's, to illustrate connection and selection strategies in this chapter. The simulation images show three possible ways to connect the elements of this stimulus. 68

Figure 32. A schematization of the Selection Circuit. Again, excitatory synapses are represented by arrows with solid shafts, while an inhibitory connection is represented by a dashed line capped with a square. Arrows with dotted shafts represent the flow of information, and the purple arrow to Reset represents a timed, positive current input to a reset cell. The bottom gray interneuron receives constant excitatory input, which can be inhibited when there is activity at the corresponding position in V2, layer 2/3. In turn, the bottom interneuron prevents runaway spread of the selection signal. The Boundary Segmentation Circuit box shows the circuit created for the V2 bipole cell that corresponds to position i . This circuit involves the boundary segmentation cell (labelled ' $i+1$ ') for its neighboring cell that corresponds to position $i+1$ 71

Figure 33. The activity, feature filters and selection signal map of two successive 50 ms timesteps in a simulated trial..... 72

Figure 34. Examples of different feature filters given the stimulus image at the top and the connection strategy where only adjacent letter I's connect. 77

Figure 35. An example of a horizontal feature filter given the toy example discussed at the beginning of this chapter where the connection parameters are tuned such that only the target pair of nearby letter I's connected. Left column shows the layer 0 activity and resulting feature filter distribution from which the selection signal's x-coordinate at the next time step was sampled. Right column shows the resulting location of the selection signal at the subsequent time step, which is superimposed on the layer 0 activity to show its position relative to layer 0 activity. The green arrow indicates that the sampled x-coordinate is used as the x-coordinate of the selection signal center on the subsequent time step. 80

Figure 36. Examples of how the selection process operates over time. In both examples the feature filter uses horizontal signal. Top row: An example of the selection process with the reset of segmentation layer 1, which occurs at the beginning of the 50 millisecond time step indicated by a red box. Bottom row: An example of the selection process without a reset. Even though the selection signal moves to a different location at 650 milliseconds, the pair of letters selected earlier remain in layer 1. 81

Figure 37. An illustration of how selection signal size impacts the speed of segmentation through a comparison of the segmentation layers of three simulations, each of which only differed in the size of selection signal used, which are indicated by the white circles. Each selection signal was centered at the same location (on the left target I). The smallest selection signal has a diameter that is half the width of a letter. The medium selection signal diameter is the width of a letter. And, the large selection signal diameter is twice the width of a letter. A comparison of segmentation layer 1 at 500-550 ms for each simulation shows that the segmentation process is faster, i.e., there is more boundary signal in layer 1 for the simulation with the large selection signal, than for the other simulations. 83

Figure 38. An example of the selection process for a matrix of shapes where a smoothed probability distribution is constructed from V2 signal. Here the feature filter used diagonal signal. Thereby, areas with diagonal signal are more likely to be the center coordinate of the selection signal. As in the previous figure, the red box indicates a reset of layer 1.	86
Figure 39. Screenshots of stimuli from Experiments 1-3. All spacing conditions are shown with the target pair in positions 5 and 6 (counting from the left) for ease of comparison.	93
Figure 40. Results by spacing condition. Top row: results from the Palmer and Beck (2007) experiment and Experiment 1 for each spacing condition. No error bars are shown for the original data since Palmer and Beck did not provide any measure of variability. Bottom row: results from simulations with three different grouping strategies. Error bars represent one standard error of the mean. Each spacing condition is numbered from 1-9 (see Figure 39), and each spacing condition falls under one of three grouping conditions, i.e., neutral, within-group, and between-group, indicated by gray, white, and black bars, respectively.	94
Figure 41. Results by grouping condition. Plot (a) shows mean response times for grouping conditions for the original data (which are an estimate given mean response times for spacing conditions reported by Palmer & Beck, 2007; means for grouping conditions were not provided) and Experiments 1-3. Plot (b) shows grouping condition mean response times for each simulated grouping strategy. Error bars represent one standard error of the mean.	95
Figure 42. The left column shows the time course of a correct trial with a between-group stimulus, and right column shows the time course of an incorrect trial with a within-group stimulus. This trial structure was used for Experiments 1-3. The black rectangle has been cropped to take up less room here.	98
Figure 43. (a) All of the stimuli for the Experiment 1 simulations. (b) A table summarizes the differences in grouping strategies, which references (c) and (d). Box (c) shows two connection strategies, and box (d) indicates two selection signal size strategies. The connection parameter sets numbered by i-iv in box (c) correspond to the following values for the horizontal connection circuits: (i) Spread Controller duration 30 ms, Long Controller input 2.0, Short Controller input 0.8; (ii) Spread Controller duration 60 ms, Long Controller input 1.7, Short Controller input 0.8; (iii) Spread Controller duration 95 ms, Long Controller input 1.0, Short Controller input 0.8; (iv) Spread Controller duration 0 ms, Long Controller input 1000, Short Controller input 0.8. Spread Controller onset was fixed at 0 ms for all simulations.	102
Figure 44. Horizontal feature filters for spacing conditions 4, 5, and 6. Top row provides those for Strategic Connections, and bottom row for No Connections.	105
Figure 45. Illustration of the timing of selection signals used for the simulations of Experiment 1. Two selections signals are used but are temporally staggered: the selection signal represented by a yellow circle begins 50 ms after the pink selection signal. After 150 ms at one location, i.e., after a 150 ms selection cycle (which was defined in Section 5.1), each selection signal shifts to a different location and its segmentation layer is reset.	106

Figure 46. Examples of V4 activity that are input to the target identification algorithm. Left column: The target identification algorithm identified a target in layer 1 because the V4 activity is generally symmetrical and of suitable width. In turn, the trial ended at 600 ms. Right column: Layer 1 has activity that is of suitable width yet is not symmetrical. Layer 2 has a single object. Although this activity is symmetrical, it is not of suitable width for a target pair. So, the target identification algorithm did not detect a target in either V4 layer, and the trial continued.	107
Figure 47. A plot comparing response times for Experiments 1 and 2 for shared spacing conditions. The number labels correspond to the conditions shown in rows 2 and 3 of Figure 39.	113
Figure 48. Mean response times for each spacing condition in Experiment 3. Error bars represent one standard error of the mean, and bar color indicates which grouping condition the spacing condition falls under.	116
Figure 49. Screenshots showing stimuli that exemplify the grouping conditions used in Experiments 4 and 5. For ease of comparison, here the target pair is always shown in positions 4 and 5, counting from the left.	120
Figure 50. Plot (a) provides mean response times for grouping conditions for the original experiment and Experiments 4 and 5. Plot (b) shows the results of simulations running a version of Experiment 5 with the model implementing different grouping strategies. Error bars represent one standard error of the mean, and bar color indicates the grouping condition.	121
Figure 51. (a) All stimuli input to the model. (b) Grouping strategies implemented by the model, which reference boxes (c) and (d). Box (c) shows connections formed by two connection strategies. For Strategic Connections, one set of connection parameters was used for the horizontal connection circuit: Spread Controller duration was 57 ms, Long Controller input was 1000, and Short Controller input was 0.8. For No Connections, the set of connection parameters was: 0 ms duration, 1000 Long Controller input, and 0 Short Controller input. Spread Controller onset was fixed at 0 ms for all simulations. Box (d) shows the selection signal map for three selection signal size strategies. The map is overlaid with a stimulus to give a sense of a signal's size relative to stimulus elements.	128
Figure 52. Example of the selection strategy used for the Vickery simulations. The output of each segmentation layer from the final 300 milliseconds (simulation time) from a particular trial, with each row showing segmentation layer activity during 50 ms. The y-coordinate of each selection signal was fixed such that it was centered on the row of circles yet did not overlap with the crosses. Shown here is the Large selection signal size strategy.	130
Figure 53. Examples of the distributions used in each condition of the Vickery simulation for both connection strategies implemented by the model.	131
Figure 54. Screenshots of stimuli from Experiment 6 for the three distractor conditions and two figure conditions. Each stimulus shown here has three target diamonds.	136

Figure 55. Top row shows results for the dot figure condition from the original experiment (with a single group of 10 participants seeing each figure condition in separate blocks), the replication, and the simulation when the model implements Grouping Strategy 1S. Bottom row provides results for line figure conditions. Distractor condition is indicated by the color, and target condition is indicated by the x-axis. Bars represent percent incorrect, and points represent mean response times, which were calculated from correct trials only. Following Trick and Enns, error bars indicate standard deviation. For the simulations, 150 trials were simulated for each target-distractor condition. 137

Figure 56. The time course for a trial of Experiment 6. For incorrect trials, the green ‘Correct’ was replaced with a red ‘Incorrect’ 142

Figure 57. Grouping strategies for the dot figure condition. (a) shows an example stimulus image generated on a particular trial and input to the model. (b) provides a summary of two grouping strategies with reference to boxes (c) and (d). Only Grouping Strategy 1S was simulated for dot figures since pilot simulations indicated that they led to similar performance. Box (c) shows two connection strategies. For the Intra-Target Only strategy, the connection parameters for diagonally-tuned cells were: 1.5 ms Spread Controller duration, 50 ms Spread Controller onset, 1.0 Long Controller input, and 0.0 Short Controller input. The horizontal and vertical connection circuits were turned off. For the Intra-Target and Intra-Distractor strategy, the parameters for the diagonal connection circuit were the same as for the Intra-Target Only strategy. For the horizontal and vertical connect circuits, the parameters were: 1.0 ms Spread Controller duration, 50 ms Spread Controller onset, 1000.0 Long Controller input, and 0.0 Short Controller input. Box (d) illustrates the selection signal size strategy by overlaying a selection signal map with the stimulus to give some sense of the signal size relative to that of the stimulus elements. 148

Figure 58. Grouping strategies for the line figure condition. (a) shows an example stimulus image generated on a particular trial and input to the model. (b) provides a summary of two grouping strategies with reference to boxes (c) and (d). Box (c) shows two connection strategies. For the No Connections strategy, all connection circuits were turned off with Spread Controller duration at 0 ms, Long Controller input at 1000.0, and Short Controller input at 0.0. For the Inter-Target strategy, the parameters for the diagonal connection circuit were: 50 ms Spread Controller duration with onset at 50 ms, 1000.0 Long Controller input, and 0.8 Short Controller input. For the horizontal and vertical connection circuits, the parameters were: 50 ms Spread Controller duration with onset at 50 ms, 0.1 Long Controller input, and 0.0 Short Controller input. Box (d) illustrates the selection signal size strategies by overlaying a selection signal map with the stimulus to give some sense of selection signal size relative to the stimulus elements. 149

Figure 59. V2 output from a simulation using the stimulus image on the far left in which Spread Controller duration was 1.4143 ms, Spread Controller onset was 50 ms, and the Long Controller and Short Controller Circuits were both turned off. In contrast, for a simulation with a slightly shorter Spread Controller duration of 1.4142 ms (not pictured), there was no noticeable spread of diagonal signal from 50-100 ms and none of the dots connected. 150

Figure 60. V2 activity at the beginning (middle image) and at a later time (right image) in a simulated trial with the set of connection parameters used in the Inter-Target connection strategy (Figure 58c, bottom image)..... 152

Figure 61. Examples of diagonal feature filters that result from each of the connection strategies simulated. These feature filters are shown over time since their temporal dynamics are important in my discussion of simulation results. 154

Figure 62. Illustration of the segmentation process for the Trick and Enns simulations. At 150 ms, the first pair of selection signals (pink circles) begin at some location and shift to a new location every 150 ms, e.g., at 300 ms. At 200 ms, the second pair of selection signals begin (yellow circles), and the final pair of selection signals (green circles) begins at 250 ms. Each pair of selection signals shifts to a new location after a 150 ms selection cycle. If a target is successfully segmented, a selection signal (red circles) at that location is added to the selection signal map for the rest of the trial. The reset signal occurs at the timestep at which the first pair of selection signals shift, e.g., at 300 ms indicated by the box with a red border. 156

Figure 63. V4 output for an input image with each type of line figure and dot figure used in the simulation. Green lines indicate how this output is divided into a 2x2 grid consisting of four cells, where each cell may contain a single figure. Similarly, for the stimuli used for the Experiment 6 simulation, each stimulus was divided but into a 5x5 grid. Measurements in pixels are shown: the width and height of line squares and dot squares (12x12 pixels) are smaller than line diamonds (17x17 pixels) and dot diamonds (16x16 pixels). In each cell, the identification algorithm determines the row numbers in which there was V4 activity. To illustrate this, the top right cell (grid cell 1,2) is divided into rows by brown lines. (In the figure, there are two pixels per row to the make the example visibly clearer, but the algorithm divides the V4 output into rows that are each one pixel tall.) The algorithm finds the most extreme row numbers that have activity in them, e.g., in grid cell 1,2, the row with the lowest row number with activity is the row that contains the top two dots, and the row with the greatest row number with activity is the row that contains the bottom two dots. The algorithm then checks whether the distance between these two rows is greater than 12 pixels. This process is also applied to columns. If the distance between the extreme rows with activity is greater than 12 pixels and if the distance between extreme columns with activity is greater than 12 pixels, then the grid cell must contain a diamond and the algorithm registers that there is a hit in this grid cell. Otherwise, as is the case in grid cell 1,2, the algorithm determines that the cell does not contain a diamond..... 158

Figure 64. Plots of simulation results for all grouping strategies implemented for the line figure condition. Mean response times shown in each plot were calculated from the reaction times in trials in which the correct number of targets was reported. Error bars represent standard deviation..... 161

ABSTRACT

A fundamental characteristic of human visual perception is the ability to group together disparate elements in a scene and treat them as a single unit. The mechanisms by which humans create such groupings remain unknown, but grouping seems to play an important role in a wide variety of visual phenomena. I propose a neural model of grouping; through top-down control of its circuits, the model implements a grouping strategy that involves both a connection strategy (which elements to connect) and a selection strategy (spatiotemporal properties of a selection signal that segments target elements to facilitate identification). With computer simulations I explain how the circuits work and show how they can account for a wide variety of Gestalt principles of perceptual grouping. Additionally, I extend the model so that it can simulate visual search tasks. I show that when the model uses particular grouping strategies, simulated results closely match empirical results from replication experiments of three visual search tasks. In these experiments, perceptual grouping was induced by proximity and shape similarity (Palmer & Beck, 2007), by the spacing of irrelevant distractors and size similarity (Vickery, 2008), or by the proximity of dots and the proximity and shape similarity of line figures (Trick & Enns, 1997). Thus, I show that the model accounts for a variety of grouping effects and indicates which grouping strategies were likely used to promote performance in three visual search tasks.

INTRODUCTION

Spatially disconnected visual elements can appear to form a perceptual group, and much effort over the last century investigated how groups form (Wagemans et al., 2012). For example, Köhler (1929) proposed that any observer who “looks passively” at Figure 1 will see two groups of patches.

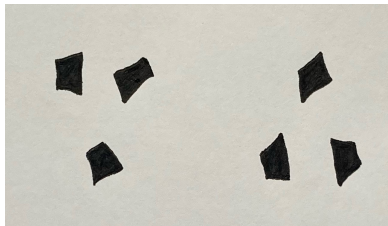


Figure 1. A redrawing of Figure 1 in Köhler (1929, p.154).

The example was used to argue against the possibility that observers group elements when they have previous experience of these elements behaving as a unit, e.g., we perceive a pencil as a unit because it behaves as a single unit when we use it. Since an observer looking at the image has not seen the left three patches behaving as a unit, Köhler argued that grouping was not learnt from experience. As a member of the Gestalt School, Köhler proposed that these groupings can be accounted for by some rule, e.g., grouping by proximity, that generalizes to other cases.

Köhler’s argument includes the following assumption: for a given stimulus, perceptual groups are, in some sense, primitives. In other words, it is assumed that there is one way to group the six patches, and everyone will perceive them as forming a group of three patches on the left and another group of three patches on the right. A similar assumption is relied on for the demonstrative examples throughout Wertheimer’s (1923/1950) seminal paper on perceptual organization (Wertheimer did identify some possible roles of experience; for discussion see Wagemans, 2018), and it continues to be presupposed in modern experimental work on grouping (Palmer & Beck, 2007; Trick & Enns, 1997; Vickery, 2008) and in examples of grouping rules given in textbooks (e.g., Palmer, 1999).

However, Köhler’s argument involves a subtle qualification: everyone *who looks passively* at Figure 1 perceives two groups of patches. This statement does not rule out the possibility that

perceived groups can be task-dependent, i.e., the way in which an observer groups stimulus elements may depend on the particular task at hand. Other Gestalt psychologists were less flexible about the role of experience in perceptual grouping, which was vigorously debated (Brady, 1933; Gottschaldt, 1926/1950; Koffka, 1935/1963; Moore, 1930; Wertheimer, 1923/1950). For example, experiments by Gottschaldt led him to conclude that experience has a negligible effect on perceived organization, e.g., seeing a figure repeatedly has little impact on whether an observer reports seeing this figure when it is embedded in a larger figure. His conclusion continued to be contested through the 1950s (for overviews, see Bevan, 1961; Bevan & Zener, 1952). This issue of whether past experience influences perceived groupings continues to be experimentally investigated (e.g., Kimchi & Hadad, 2002; Vickery & Jiang, 2009; Zemel et al., 2002; for a review see Peterson & Kimchi, 2013).

The present project explores the implications of rejecting the implicit assumption that for a given stimulus there is only one way to group its elements. If this assumption is rejected, then we are in a better position to investigate what grouping strategy(s) observers may use to promote performance on a particular task and stimulus set. However, rejection of this assumption does not entail ignoring or reducing the impact of bottom-up information from the stimulus. Indeed, for a given stimulus, an observer may use bottom-up information about, e.g., size, proximity, shape, and position, to group its elements in several possible ways. Moreover, it seems that such groupings can be tuned to promote performance for a given task, e.g., if asked to find the pair of nearby squares in a row of shapes, an observer may group the items by proximity and shape such as to allow for easier selection and segmentation of the target pair. In effect, bottom-up information in conjunction with top-down processes can be used to promote a perceived grouping of particular elements in accord with a learned strategy for performing some task efficiently, e.g., finding a target among distractors as quickly as possible.

The properties and mechanisms of the formation of perceptual groups have been rigorously investigated (Elder & Goldberg, 2002; Thórisson, 1994; Wagemans et al., 2012), and low-level, neuroscience-based algorithms have been proposed (Roelfsema, 2006; Roelfsema & Houtkamp, 2011; Ward & Chun, 2016). Although it is clear that the computations to form groups are influenced by both bottom-up and top-down information (Vecera & Behrmann, 2001), accounts of perceptual grouping tend not to integrate bottom-up and top-down processes in detail.

This project aims to address these gaps, i.e., the lack of an account of grouping that both integrates bottom-up and top-down information and the lack of an account of grouping strategies given particular stimuli and tasks, by: (1) developing a computational model that both integrates bottom-up and top-down processes and can run simulations of previous experiments investigating grouping, (2) making predictions from the simulated results regarding, e.g., possible grouping strategies that would augment or inhibit performance given the task and stimuli, (3) conducting experiments on human participants to test these predictions, and, ultimately, (4) offering a mechanistic explanation of such phenomena. Specifically, this project aims to study perceptual grouping by modelling experiments that use visual search tasks.

This project focuses on visual search tasks, where an observer must locate a target item in a static display, because such tasks have been used to investigate different aspects of perceptual grouping (Wagemans et al., 2012). Additionally, under the approach to grouping explored here, i.e., observers use learned grouping strategies that are dependent on the particular task and stimulus set, apparently simple tasks and stimulus sets turn out to be more complex than was intended. For example, *prima facie* one experiment simulated below only varies the proximity of stimulus elements along a row. However, the optimal strategy appears to be more complex than just grouping nearby target elements; in view of the model, an optimal grouping strategy seems to depend upon proximity, shape, and total width of the row of elements. In turn, it is necessary to conduct experiments that, e.g., control for width of the row of stimulus elements, in order to tease apart different features that are, according to the model, likely to play a role in beneficial grouping strategies.

The computational model used and developed in this project is the version of LAMINART implemented in Francis et al. (2017). Francis et al. (2017) showed how a neural circuit for grouping and segmentation explains several visual uncrowding effects. When a target vernier does not group with flankers, the circuit can segment out the flankers and, thereby, free the target from crowding effects. However, this model is limited in two ways that prevent it from being applied to visual search tasks. First, whether elements in a scene formed a perceptual group was strictly determined by fixed parameter values. In other words, if two stimulus objects were within a particular distance apart and appropriately aligned, they would always group. Otherwise, they would not group. Although this may be sufficient if we accept the assumption that there is only one possible perceived grouping of the elements of a given stimulus, it lacks both the flexibility

needed if there are alternative groupings and a top-down grouping control process. Second, the model's segmentation circuit was not designed to simulate the search process involved in visual search tasks. This circuit has a selection signal, which, as explained further below, acts as an attentional spotlight that falls on a particular area and selects stimulus elements that fall under it. On each trial in the simulations conducted by Francis et al. (2017), each selection signal had a fixed location for the duration of the trial. Although reasonable for vernier judgement tasks, this segmentation circuit lacks the dynamic selection process characterizing visual search tasks.

In order to run simulations of experiments that investigate grouping, this project develops modified grouping and selection circuits that address the two limitations of LAMINART mentioned above. In particular, developing LAMINART's selection process provides a way of investigating segmentation strategies used in visual search tasks. Further, altering LAMINART's grouping and selection circuits to enable different possible groupings of visual elements allows the model to simulate experiments that investigate grouping and to make testable predictions about grouping strategies given a specific task and stimulus set. After introducing the model, describing modifications made to it, and elaborating on what a task-dependent grouping strategy is, I apply the model to three experiments designed to investigate grouping with a visual search task, namely Experiment 1 of Palmer and Beck (2007), Experiment 2a of Vickery (2008), and Experiment 2 of Trick and Enns (1997). Since these experiments used small sample sizes, I conducted replication experiments of these studies to provide better estimates of performance. Similarities between the results of the replication experiments and those of simulations when the model implements particular grouping strategies suggest that human observers may be using similar grouping strategies to promote performance on the tasks.

PREVIOUS WORK: LAMINART

The LAMINART neural network model (Cao & Grossberg, 2005; Francis et al., 2017; Grossberg & Raizada, 2000) consists of bottom-up circuits modelling areas of the human visual cortex and novel grouping and segmentation circuits. Input to the model consists of black-and-white images that could be shown to a participant (e.g., left panel of Figure 2, and top panels of Figures 3 and 4). For each pixel in the image, a network of neurons is created with a layered structure similar to that found in brain areas V1 and V2 (Figure 2, panel b2). Output of the model is presented as an image (Figure 2, panel c), in which the color of each pixel indicates the spike counts of orientation-tuned cells in the final layer of V1 or V2 at that location. A bright green pixel indicates that the cell sensitive to horizontal edges at that location was firing rapidly, a red pixel indicates vertically-tuned cell activity, blue indicates diagonally-tuned cell activity, and black indicates that no neurons are firing at that pixel location.

The version of LAMINART in Francis et al. (2017) featured a grouping circuit and a novel top-down segmentation mechanism. Following Grossberg and Mingolla (1985a, b), grouping occurs in the model when illusory contours connect spatially separated elements, e.g., the four flankers on either side of the vernier in Figure 3. The grouping circuit causes illusory contours to form between two elements in a stimulus if their edges are of a particular alignment and distance apart. In this context, the term “illusory contours” refers to the boundary signal in V2 that does not correspond to bottom-up luminance signals from the stimulus. In other words, illusory contours are contours in a V2 output image that do not reflect contours in the stimulus image. These illusory contours are often amodal, meaning that an observer would be aware of them yet not explicitly see them as producing a modal experience (e.g., color or brightness). Although Grossberg’s illusory contours have been used to (at least partially) explain illusory contours exemplified by the Kanizsa triangle, i.e., the well-defined luminance edges connecting the pac-men that most observers report seeing in this illusion (Grossberg & Mingolla 1985b; Kogo et al., 2010), Grossberg’s illusory contours refer to a particular pattern of model V2 cell activity and are more general than the illusory contours experienced in typical visual illusions, e.g., nearby flankers in Figure 3 may be perceived as grouped/connected even where there are no visible lines connecting them.

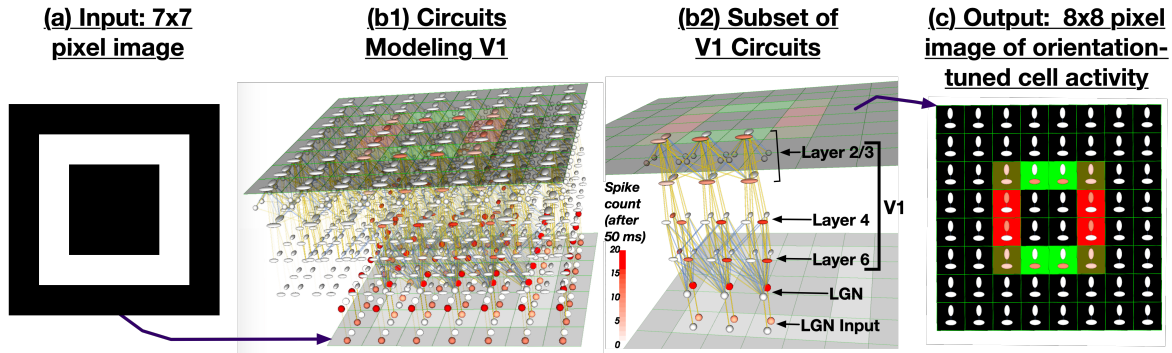


Figure 2. Example of the bottom-up circuits created in LAMINART for the input of a white square line figure on a black background (panel a). Panels b1 and b2 are visualizations of these circuits sandwiched between the input image (bottom) and V1 output image (top). Each pixel in the input and output images is a grid cell. To provide a sense of its complexity, panel b1 shows the entire V1 network created for the input image shown in panel a. Panel b2 shows only a subset of this network, i.e., the circuits created for three pixels of the input, with the layers labelled. Panel c illustrates what the output represents: the spikes of orientation-tuned neurons in layer 2/3 of V1 (The output is an 8x8 image because an extra row and column are added in the simulation to avoid edge effects). In panel c, the cells elongated along the x-axis (y-axis) are sensitive to horizontal (vertical) edges. Note that this visualization of V1 is simplified: there are also cells tuned to diagonals at 45° and 135° , making the model have double the number of V1 cells that are represented here.

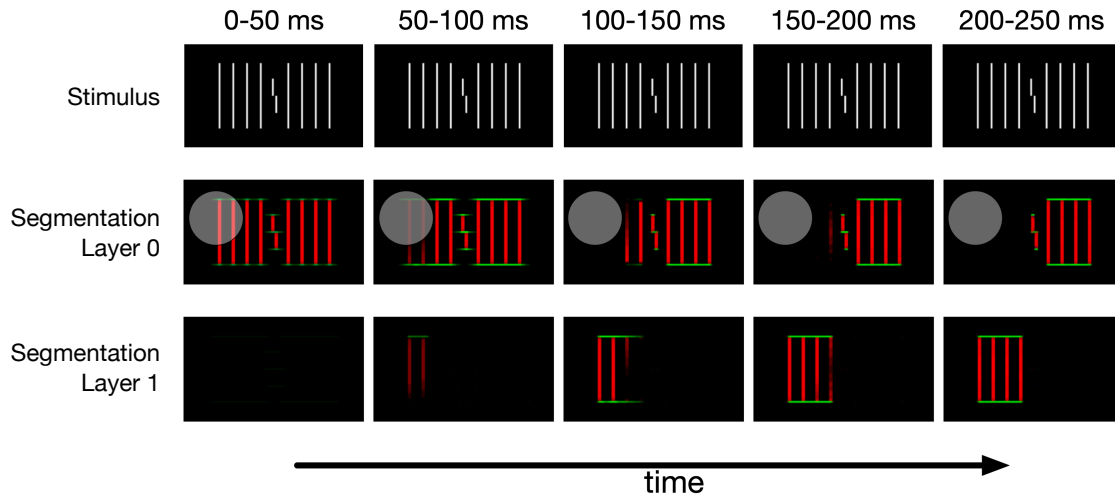


Figure 3. Example of the top-down segmentation process for the stimulus in the top row. The selection signal is represented by the translucent white circle in segmentation layer 0. Any boundaries directly under the selection signal or connected to boundaries under the selection signal are segmented out of layer 0. The sum of the activity over 50 milliseconds in each layer is represented in each non-stimulus image.

The segmentation mechanism uses a selection signal and segmentation layers (Figure 3) to perform a kind of figure-ground processing.

The selection signal is represented by the gray circle in segmentation layer 0. The segmentation mechanism segments V2 output into different layers depending upon the location of the selection signal. In effect, a selection signal acts as an attentional spotlight with location-based and object-based features: if an element in the stimulus falls under the selection signal or is connected to an element that falls under the selection signal, then over time the element is segmented out and put into a different layer, e.g., the left flankers connected by horizontal illusory contours in Figure 3.

Proposed Developments

Conceptual Shift

This project further develops the model of Francis et al. (2017). These model developments are outlined in the next subsection. However, this project involves a conceptual shift that includes a change in terminology for clarificatory purposes. Francis et al. (2017) proposed that illusory contours group stimulus elements, and that these groups of elements could be segmented out by a selection signal. So, in their model there are two mechanisms: one that creates groups via illusory contours, and the other that allows for the selection and segmentation of these groups.

Rather than assuming that the illusory contour mechanism alone explains grouping, I propose that grouping involves both of these mechanisms, namely a connection mechanism that can join stimulus elements together via illusory contours, and a selection mechanism that selects and segments out stimulus elements. This proposition is motivated by the possibility of multiple stimulus elements not being joined by illusory contours yet being selected and segmented, e.g., Figure 4.

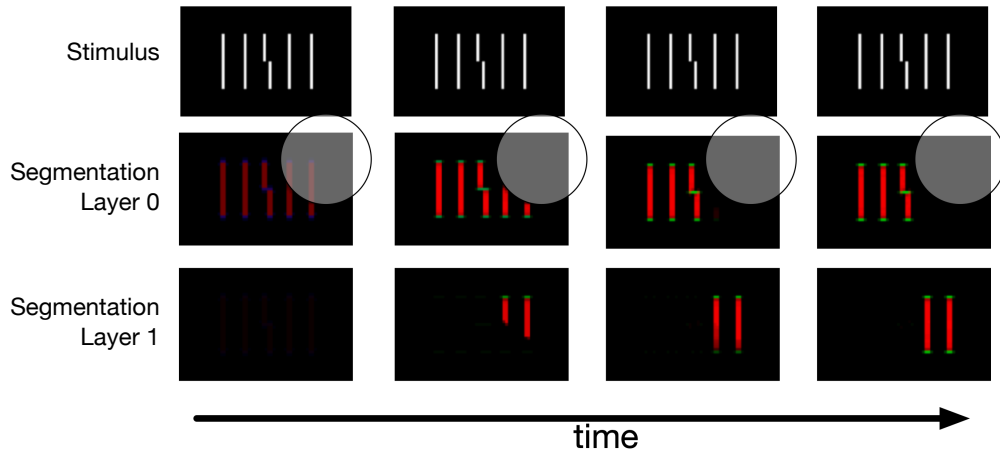


Figure 4. An example where a selection signal segments multiple unconnected stimulus elements and seems to thereby group the two far right lines.

In this example, two unconnected flankers happen to fall under the selection signal and, thus, are segmented out and isolated from the other elements. With such selection and isolation, it seems reasonable to conclude that these two flankers form a group, which suggests that separate stimulus elements can be grouped by a segmentation process and that connections between stimulus elements are not necessary for groupings of nearby elements. Thus, the segmentation process alone can result in groupings of elements.

So, for the sake of clarity, I will henceforth refer to illusory contours as ‘V2 connections’ or simply as ‘connections’ (depending on the context) and the circuit called ‘the grouping circuit’ above as ‘the connection circuit’. This is because I propose that this circuit is only a part of a grouping circuit that includes both the connection circuit and the selection circuit. In turn, this project explores the relations between the connection mechanism (i.e., the grouping mechanism of Francis et al., 2017) and the selection mechanism, both of which are subject to top-down control, and the roles they play in perceptual grouping. I contend that, given a particular stimulus set and task, an observer uses a grouping strategy, which is constituted by a connection strategy, i.e., a particular way of joining stimulus elements by V2 connections, and a selection strategy, which involves determining where and when to place selection signals.

Overview of Model Development

In this project I develop the model of Francis et al. (2017) in two fundamental ways.

First, (1) I introduce a trial-wise dynamic selection process that enables the model to impose a real-time search strategy for visual search tasks. (1) is necessary in order to simulate tasks that require scanning a display for a target. To simulate experiments with vernier discrimination tasks, the simulations in Francis et al. (2017) used a single pair of selection signals on each trial. Each selection remained fixed in one position for the entire trial. They used pilot simulations to determine the best pixel coordinates at which to place the center of each selection signal, and noise was added to each pair of coordinates. In contrast, visual search tasks involve a more dynamic and strategic selection process. Thus, modification of the segmentation circuit is required.

Second, (2) I introduce top-down control over parameter values that determine whether stimulus elements connect. The connection circuit introduced in Francis et al. (2017) only allows elements that are within a certain distance and alignment to connect, which is too rigid to allow for alternative connections of a particular set of elements in a scene. To motivate the need for this kind of circuit, consider the row of squares in Figure 5.



Figure 5. An example to illustrate multiple ways in which the elements of a single image may be grouped.

There are a number of ways that an observer may perceive the squares as forming groups, e.g., as three pairs of nearby squares, a group of four squares on the left and a pair on the right; a single group of six squares; or, they may be regarded as six individual squares. A fixed set of receptive field parameters in the Francis et al. (2017) connection circuit only allows for one of these grouping options. In turn, the top-down control is necessary to allow the model to flexibly capture different ways in which an observer may use, e.g., shape, size and proximity, to group the elements of a given stimulus.

As indicated in the introduction, the approach taken here is to reject the assumption that there is only one way in which an observer groups the elements of a given stimulus. So, to explore the implications of rejecting this assumption, the model must be flexible enough to connect stimulus elements in a number of different ways. Additionally, I propose that an observer uses a grouping strategy to group stimulus elements to promote performance on a particular task and given stimulus set. For example, consider the row of shapes in Figure 6.

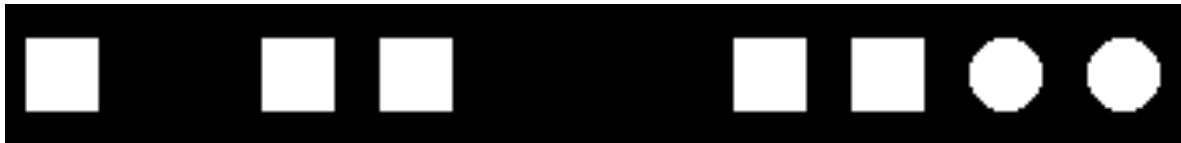


Figure 6. An example to motivate the claim that groupings can be task-dependent.

Suppose the task is to count the number of elements that are in the row. In this example, there are seven shapes. Some researchers have argued that one way to speed up performance on this task is to group nearby elements and add up the elements in each group (Starkey & McCandliss, 2014; van Oeffelen & Vos, 1982). Recall that I propose that a grouping strategy consists of a connection strategy and a selection strategy. One connection strategy that would promote performance on this task is to connect the three squares on the left and connect the four shapes on the right. These connections will allow the observer to more quickly select and segment out the shapes that are to be counted, add up the elements in each group and, thereby, count the shapes faster. If, in contrast, the task is to count the number of squares only, this connection strategy would be detrimental to performance since it would cause the distractor circles to be selected along with the target squares on the right. Instead, it seems likely that an observer would adopt a different connection strategy, e.g., group by proximity and shape so that only nearby squares are connected. In turn, the model must have the flexibility to promote different connections among stimulus elements, and the model should incorporate a connection circuit that is subject to top-down control in order to simulate the top-down control of observers over which groupings are perceived.

MODEL SUMMARY AND BOTTOM-UP CIRCUITS

Model Summary

In this project, I developed the LAMINART model such that it can be used to run simulations of existent experiments that use visual search tasks to investigate perceptual grouping. The main modifications necessary to simulate such experiments are provided in the next two chapters. Additionally, the bottom-up circuits are presented in the next section because they have been modified from those presented in Francis et al. (2017) mainly to simplify the network. But, before getting into the details of the model, consider the Circuit Overview in Figure 7 to get a sense of the structure and operations of the main components.

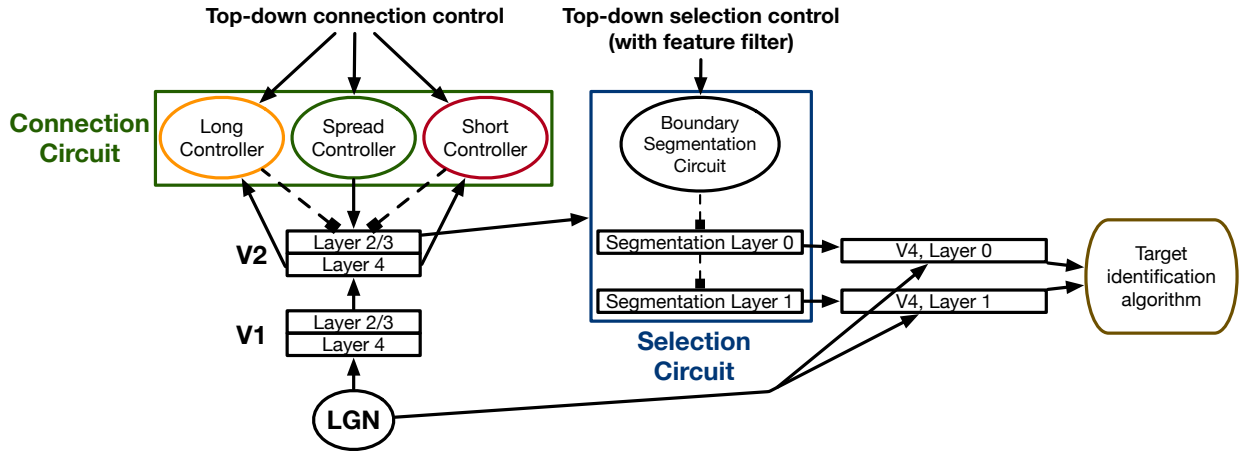


Figure 7. A schematic overview of each component of the model.

Starting at the bottom left and working upwards, LGN input, which represents brightness and darkness information from the input stimulus image indexed by each pixel coordinate, is fed in through an oriented convolutional filter to V1, which consists of two layers: layer 4 and layer 2/3. The activity of V1, layer 2/3, passes to V2, which also consists of two layers. V1 and V2 comprise the bottom-up circuit of the model.

The green Connection Circuit in Figure 7 consists of three modules: Spread Controller, Long Controller, and Short Controller. The latter two modules use activity from V2, layer 4, plus an observer-specified input current, to prevent the spread of long or short connections in V2, layer

2/3, while the former encourages the spread of these connections for a specified time. I propose that an observer on a given task with a particular set of stimuli can learn to tune the parameters of each module, i.e., the onset and duration of the Spread Controller and input current into the Short and Long Controllers, such as to group particular stimulus elements via V2 connections in a way that allows them to efficiently perform the task. This ability of an observer to tune these parameters is what is labelled as ‘Top-down connection control’ in Figure 7.

V2, layer 2/3, feeds into the blue Selection Circuit consisting of segmentation layers and the Boundary Segmentation Circuit. This circuit models the ability of an observer to strategically search for a target using particular features of the target, e.g., if the target is a diamond and the distractors are squares, an observer may use diagonal signals to guide placement of the selection signals. The location of these selection signals is fed into the Boundary Segmentation Circuit. The Boundary Segmentation Circuit functions to separate boundary signals in segmentation layer 0, which is essentially the output from V2, layer 2/3, that either falls under the selection signal or is connected to boundaries that fall under the selection signal, into another segmentation layer, as exemplified in Figure 3 above.

Finally, each segmentation layer feeds into a V4 layer. In each V4 layer, activity represents the results of a filling-in process and the resulting segmentation that occurs via surfaces. Since I did not develop these circuits, I do not describe the V4 circuits (for more details, see the surface segmentation process explained in Francis et al., 2017). However, I use the signal in V4 layers to determine whether a target has been selected or not using a simple algorithm tailored to the targets in each experiment.

To ensure biological plausibility, I developed and simulated the model within the NEST 2.14.0 simulator program (Gewaltig & Diesmann, 2007; Peyser et al., 2017), which provides models of various neurons and synapses and a framework to simulate networks constructed by these elements. All of my simulations used leaky integrate-and-fire neurons with alpha-function shaped synaptic currents and static synapses with manually set weights. Inputs to the network were provided by simulated direct current within the NEST program. Grossberg and Raizada (2000) previously provided links to known neurophysiology for much of the bottom-up circuit, so I do not repeat that discussion here. As detailed in the next section, the bottom-up circuit is simpler than the version described in Grossberg and Raizada (2000) and Raizada and Grossberg (2001). So, many of their links should apply here as well. The proposed top-down control circuits are novel,

and I am not aware of neurophysiological investigations that reveal these precise circuits. However, there is overwhelming neurophysiological evidence (at least in mice) that top-down control influences local circuits for visual perception (Zhang et al., 2014) and modulates figure-ground perception (Kirchberger et al., 2021).

Bottom-Up Circuits

The V1 and V2 circuits used in this model are simplified versions of the circuits used in Francis et al. (2017), which are based on earlier work that modelled the computations performed by components of the human visual cortex (Raizada & Grossberg, 2001). Unlike the circuits of Francis et al. (2017), the current circuits lack cells representing layer 6, feedback connections from layer 2/3 cells to layer 6 cells, and pooling neurons in V2, layer 2/3. These pooling neurons have been replaced by more synapses between nearby cells and by locally defined weights between each V2, layer 2/3 cell and its immediate neighbors, which will be explained in more detail below.

V2 detects edges and connections, and the output of its bipole cells, i.e., cells at V2, layer 2/3, represents edges at particular retinotopic locations. To explain how this circuit operates, I first present the structure of the circuit created for a single pixel location where competition occurs between cells tuned to opposite orientations, which highlights the functions of intra-position synapses. Then, I explain how these cells are connected to cells tuned to the same orientation at neighboring V1/V2 circuits, highlighting the functions of inter-position synapses.

Consider the V1/V2 circuit created for one pixel in a network that only has cells tuned to horizontal and vertical edges. This circuit is depicted in Figure 8, left column, where horizontally-tuned cells are represented by ovals elongated along the x-axis and vertically-tuned cells by ovals extended along the y-axis. This is a dipole circuit (Grossberg, 1980) in which the activity in V2, layer 4, is the result of competition between V1 complex cells (i.e., cells in V1, layer 2/3) that are tuned to edges of opposite orientations. So, for example, if the input is a horizontal edge, the corresponding horizontally-tuned V1 complex cell will have more activation than the vertically-tuned V1 complex cell. In turn, the vertically-tuned V2, layer 4 cell becomes inhibited while the horizontally-tuned V2, layer 4, cell will be quite active.

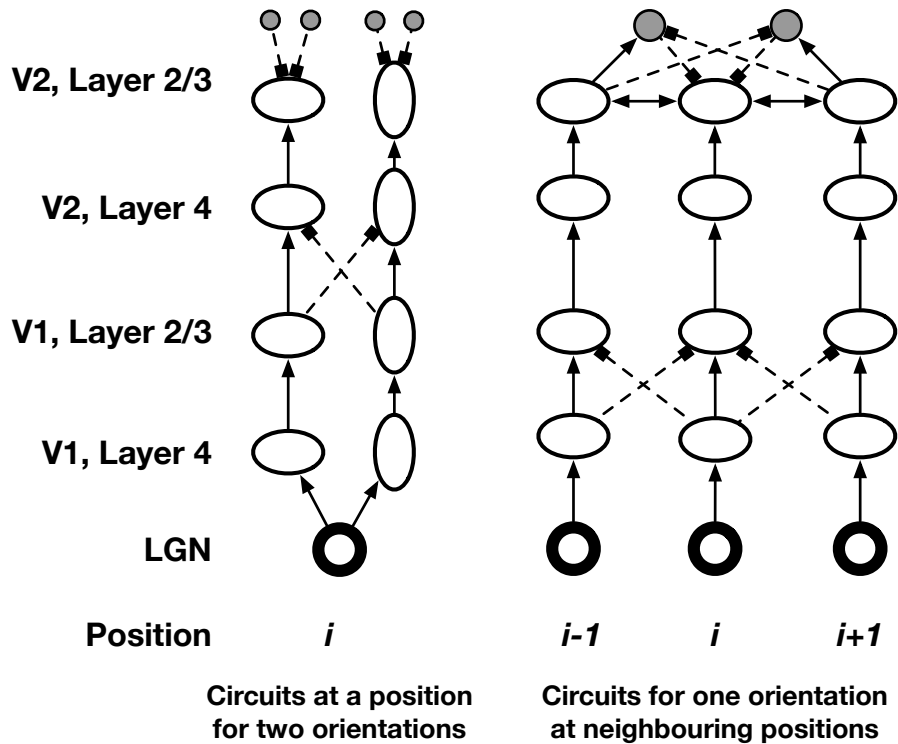


Figure 8. A schematization of the circuits representing V1 and V2. Left column: This illustrates cells and the synapses between them that are created for one pixel i in the input image. Here, there are cells tuned to two orientations, with the cells tuned to horizontal contours represented by the ovals elongated along the x-axis, and cells sensitive to vertical contours represented by ovals elongated along the y-axis. Right column: This shows the synapses that are between horizontally-tuned cells and their immediate neighbors, i.e., the bottom-up circuits created for neighboring pixels in the image. Excitatory synapses are represented by arrows, while an inhibitory connection is represented by a dashed line with a square head.

Next, consider the synapses between this circuit and neighboring cells. Figure 8, right column shows connections between this circuit (located at position i) and neighboring circuits (located at positions $i-1$ and $i+1$) for horizontally-tuned cells. The inhibitory synapses from the neighboring V1 simple cells (i.e., cells in V1, layer 4) to the V1 complex cell corresponding to the pixel at location i induces spatial inhibition within orientation. This functions to clean up edges and promote end-cutting. For example, if there is a rather weak input into the horizontally-tuned simple cell at location i but a strong input into the horizontally-tuned simple cell at location $i-1$ and/or $i+1$, they will inhibit the horizontally-tuned cell at i in V1, layer 2/3 and, thus, promote a clean edge or gap at location i .

After the activity of V1 complex cells is pruned, they feed into the V2, layer 4, cells, which feed into the V2 bipole cells (i.e., the cells depicted as ovals or circles with white fill in V2, layer 2/3). The interneurons (i.e., the cells depicted as gray filled circles in Figure 8) and bipole cells

neighboring the V2 bipole cell at position i are connected by synapses. The interneurons play a key role in the model due to their two interrelated functions. First, the interneurons prevent runaway spread. Figure 9 provides an illustration of this process given a simple stimulus with two squares.

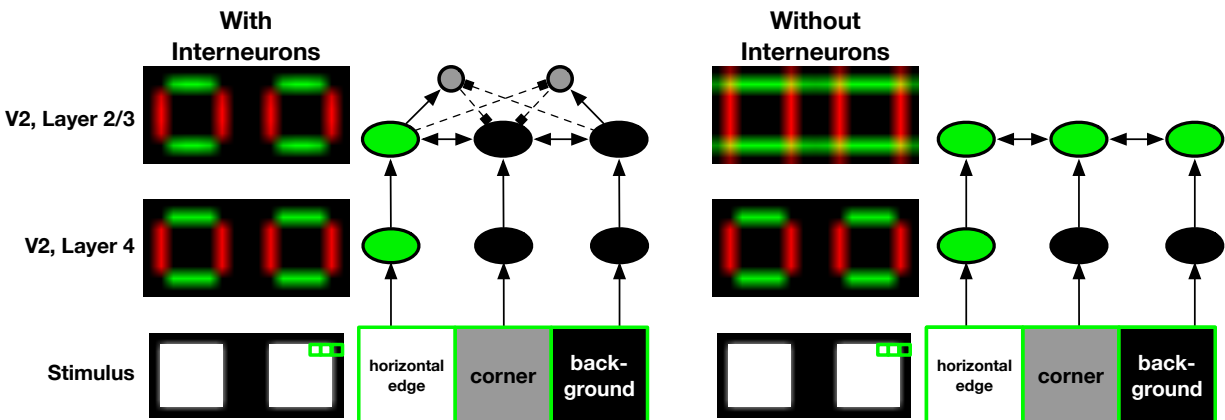


Figure 9. A demonstration of how the interneurons prevent runaway spread. The three green boxes on each stimulus image corresponds to three pixels, and the circuit diagrams for V2 horizontally-tuned cells are shown to their right.

Figure 9 shows V2, layer 4, cell activity (middle row) and bipole cell activity (top row) from two simulations with the same stimulus image (bottom row). To simplify the example, the simulations were run with only cells sensitive to horizontal and vertical edges. The left column illustrates how the intact bottom-up circuits of the model (i.e., when the model includes interneurons) operate, while the right column depicts what happens when the interneurons are removed and the model is run with the same stimulus. V2, layer 4, cell activity is the same, but the V2 bipole cell output in the top row is different because, without the interneurons, the activity from every cell spreads to neighboring cells tuned to the same orientation via lateral connections between them. The right circuit diagram provides an example of this behavior. To explain this example, I must first present how these circuit diagrams relate to the stimulus image and V2 output. Due to the pruning of the V1 complex cells described earlier, there is little horizontal or vertical signal for cells corresponding to the corners of the squares. In turn, there is no signal propagated to V2, layer 4, cells at the corners. The circuit diagrams show the activation of the layer 4 and bipole cells in V2 for two pixels of the top edge of the right square plus a cell to the right of the

square, indicated by the three green boxes. In Figure 9, the color of the horizontally-tuned cells indicates its activation, where black indicates no activation and bright green indicates that the horizontally-tuned cell at that location is firing rapidly. In the circuit diagram on the right of Figure 9, although the V2 bipole cell at the corner position has no input from its V2 simple cell, it becomes active due to the lateral connection between it and the V2 bipole cell on the right. The net effect of this unchecked V2 bipole cell activity is the extended horizontal activation across each row where there is an active horizontally-tuned bipole V2 cell. To explain how the interneurons prevent this runaway spread of activation and to explain the circuit diagram on the left of Figure 9, it is necessary to explain the second key function of the interneurons.

Their second function is to form an AND-gate, which is established by how interneurons and V2 bipole cells are connected by synapses. This is illustrated by Figure 10.

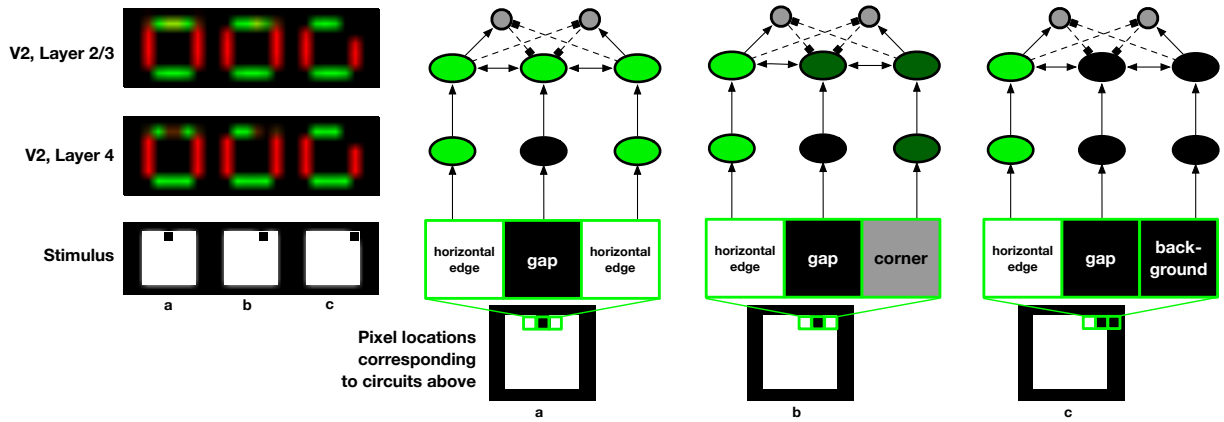


Figure 10. An example to illustrate the AND-gate function of interneurons. The input image consists of three white squares on a black background. Each of the squares has a pixel-wide gap along the top edge. The three green squares on each white square indicate which pixel locations that the circuits on the right correspond to.

Figure 10, left column, provides output from the model given a stimulus consisting of three squares. *Square a* has a gap in the center of its top edge, and the corresponding V2 simple cell output image also has a gap, i.e., although the vertically-tuned cells in the gap are firing a bit, there is little to no horizontal signal. However, the corresponding V2 bipole cell output indicates that the horizontal cells at the gap location are active. The left-most circuit diagram provides an idealization of how the AND-gate function of the interneurons facilitates this behavior: if both of

the neighboring V2 bipole cells are similarly active, then both interneurons will be inhibited and, thus, the interneurons do not inhibit the bipole cell at the gap location. Thus, the gap is bridged by horizontal activity in this case.

Squares b and *c* provide cases in which the interneuron AND-gate greatly reduces or shuts down activity of V2 bipole cells. Unlike the example in Figure 9 above, the V2 simple cell activity in the corners and gaps is not entirely negligible when gaps are added to the squares. To simplify this illustration of the properties of the interneurons, suppose that the corners have weak but non-negligible horizontal V2 simple cell activity, which is represented in the Figure 10 circuit diagrams by the dark green fill, while the gaps have no activity. The middle circuit diagram illustrates cell activity for three pixels at the top of *square b*: a pixel to the left of the gap, the gap, and the pixel corresponding to the top right corner of *square b*. The right-most circuit diagram illustrates cell activity for three pixels at the top of *square c*: a pixel to the left of the gap, the gap (which was created by removing the pixel at the top right corner of the square), and a pixel to the right of the square corresponding to the background. In both of these cases, one of the neighboring bipole cells, i.e., the left V2 horizontally-tuned bipole cell, is much more active than the other. In turn, the interneuron on the left that it feeds into it will be excited and, thus, strongly inhibit the bipole cell at the gap. These examples illustrate that if the V2 bipole cell activity to the left and right of the gap are not similar in strength, the interneuron AND-gate will greatly reduce or eliminate the activation of the V2 bipole cell corresponding to the position of the gap.

So, given the bottom-up circuit presented thus far, neighboring cells must both be active and have around the same strength of activation for V2 bipole cell activity to spread across neighboring cells. Otherwise, the interneuron AND-gate will prevent the spread of activity to neighboring cells that are tuned to the same orientation. Returning back to Figure 9, the left circuit diagram (which is quite similar to *square c*), it is the AND-gate function of interneurons that prevents the runaway spread of V2 bipole cell activity: the strong activation of the left interneuron from the left V2 horizontally-tuned bipole cell causes the interneuron to strongly inhibit the V2 bipole cell at the corner. In turn, this strong inhibition offsets any activation sent from the left bipole cell to the bipole cell at the corner.

Although somewhat simplified, this circuit functions very similarly to the bottom-up circuits of Francis et al. (2017). However, in the version used here, the dedicated long-range pooling cells of the Francis et al. version have been replaced by: locally defined weights, and

interneurons and V2 bipole cells that now effectively pool input from neighboring V2 bipole cells as a function of a parameter for grouping speed. For example, when grouping speed is set to 1, which is schematically depicted in Figure 11, bipole cells at $i-1$ and $i-2$ both have excitatory feedforward connections to the left interneuron, inhibitory feedforward connections to the right interneuron, and bidirectional excitatory connections to the bipole cell at location i .

At higher grouping speed values, the number of pooled neighboring neurons is higher, but the same basic synapses are used, i.e., bidirectional lateral synapses from the central cell to its neighbors (brown arrows, Figure 11), an excitatory synapse from each neighboring bipole cell to one of the interneurons (magenta arrows, Figure 11), and an inhibitory synapse from each neighboring bipole cell to the other interneuron (blue synapses, Figure 11). Note, however, that this parameter is not considered to be subject to top-down control. Rather, it is regarded as set by the neural architecture and, in each simulation below, its value was fixed.

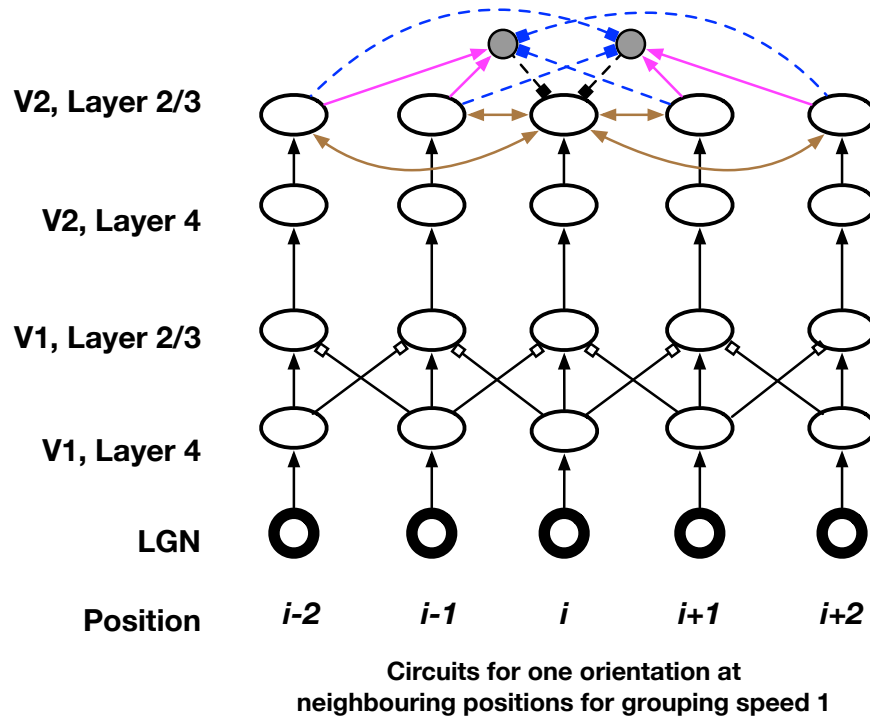


Figure 11. In the current version of the model, the dedicated pooling cells of the 2017 version have been replaced with grouping speed, which involves more synapses between nearby V2 bipole cells. Colors are used here to highlight the different synapse types that are added for each neighboring cell that is involved, the number of which is a function of grouping speed.

TOP-DOWN CIRCUITS I: THE CONNECTION CIRCUIT

To introduce top-down control of V2 connections into the model, we created a connection circuit with three component circuits: Spread Controller, Long Controller, and Short Controller (see Figure 12 for an overview, but each component will be presented and explained in detail below). These circuits allow for top-down control of the range of V2 connections by adjusting two timing parameters, which allows boundary signals to spread, and two size parameters, which prevent the spread of either long or short boundaries. A version of the Spread Controller Circuit was originally implemented by Francis and Bornet (2019). (Although the structure of this version is the same in the present model, I modified some weights and filters due to its interaction with the other components of the connection circuit.) I implemented the Long and Short Controller circuits.

In the next two sections, I present each component of the Connection Circuit in turn.

Spread Controller Circuit

Spread is encouraged by the Spread Controller Circuit, shown in green in Figure 13. Top-down control of the Spread Controller cell consists of tuning two timing parameters: the onset of top-down input, and the duration of this input. To introduce this circuit, I initially focus on the impact of top-down control duration on connection formation. Onset is here fixed at 0 milliseconds (ms), i.e., top-down control begins at the same time as stimulus onset. Later in this section I explain how varying the onset can impact connection formation.

When the Spread Controller cell is excited, it inhibits the interneurons and thereby allows for uninhibited spread of activity among bipole cells. When the Spread Controller is no longer active the interneurons return to their default function, i.e., they halt the spreading of connections and, eventually, eliminate any created connections that have not formed a closed set.

To understand the dynamics of the circuit, consider its behavior in Figure 14.

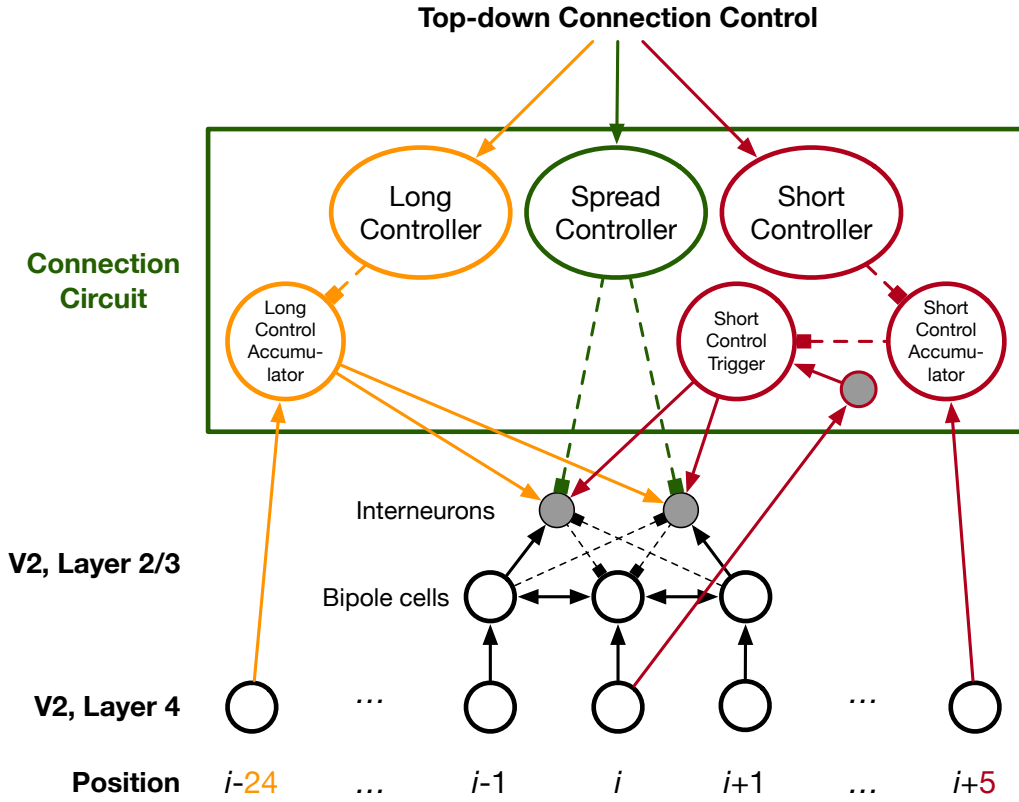


Figure 12. A schematized depiction of the entire Connection Circuit. As explained below, though, this is simplified for the sake of clarity: the Short Control Accumulator cell pools input from all V2, layer 4, cells tuned to the same orientation that are in positions $i \pm 5$ relative to the position of the simple cell at i . Similarly, the Long Control Accumulator cell pools activity from simple cells at positions $i \pm 24$. Each arrow from “Top-down connection control” represents the value of the connection parameters, i.e., onset and duration of the Spread Controller and a constant non-negative current into the Short and Long Controllers. The remainder of the arrows represent synapses: excitatory synapses are represented by arrows with triangular heads, while an inhibitory connection is represented by a dashed line with a square head. Color of synapses and cells indicate that they constitute a particular module of the Connection Circuit.

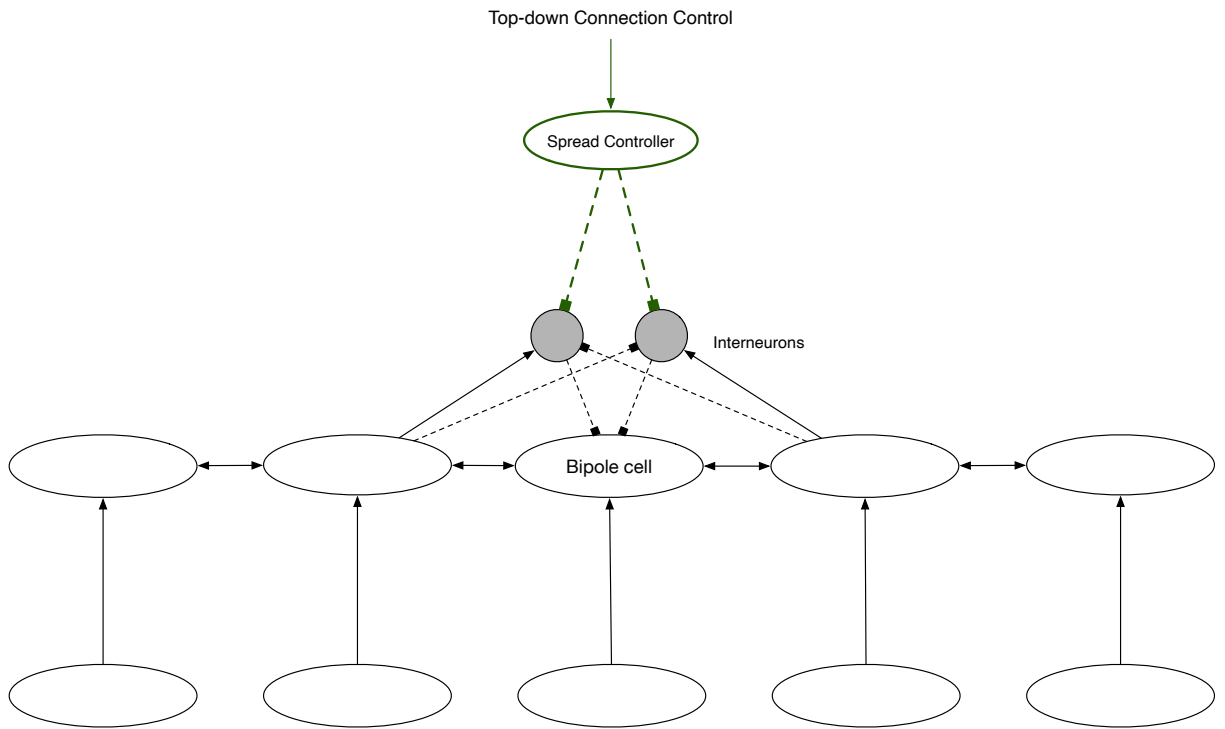


Figure 13. Horizontally-tuned V2 cells (ovals) and interneurons (gray circles) with the Spread Controller Circuit (green). The Spread Controller Circuit allows for top-down control of the spread of connections between bipole cells via the inhibition of the interneurons for a particular period of time. The longer the interneurons are inhibited by the Spread Controller cell, the farther the signal from active bipole cells can spread.

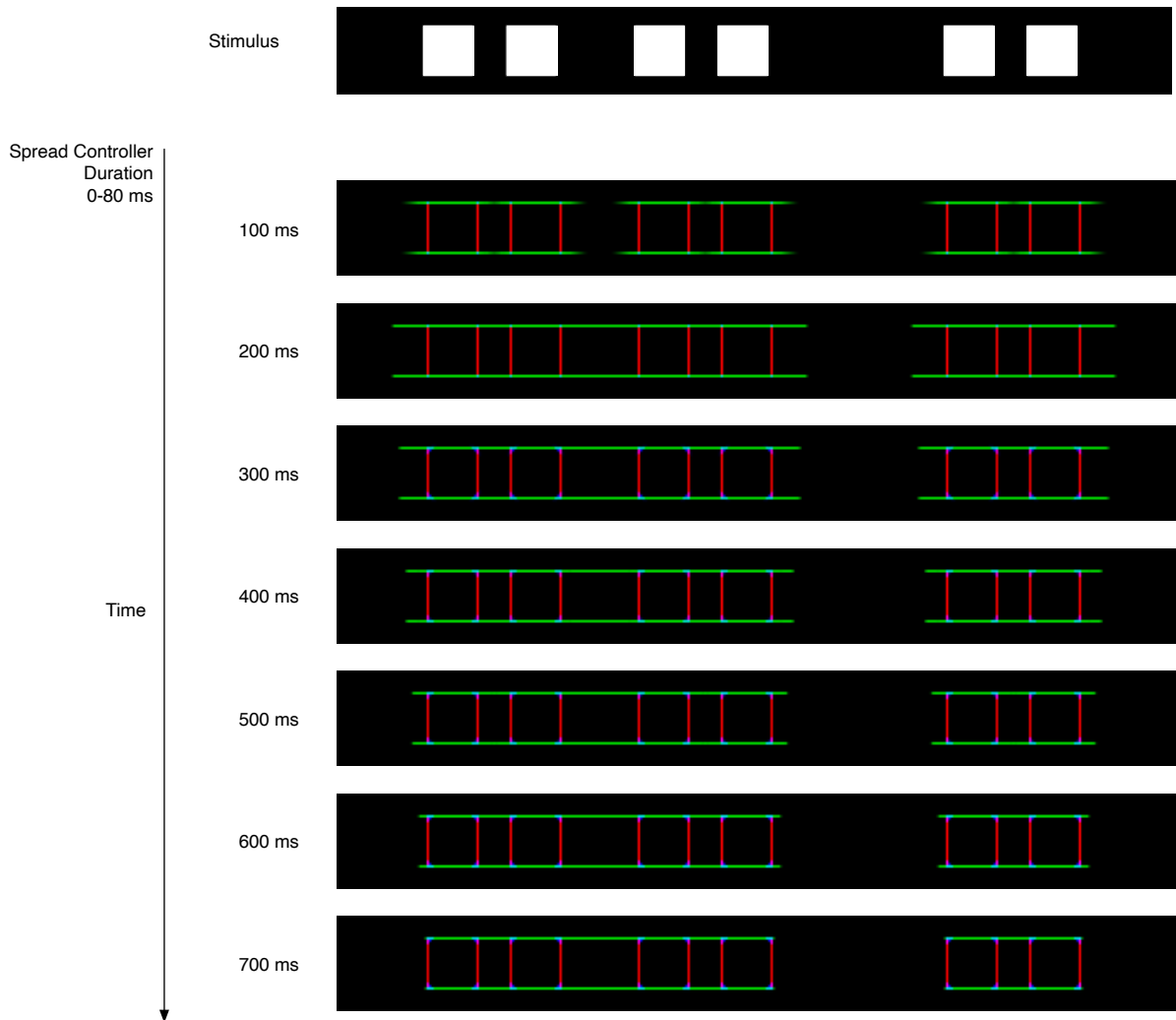


Figure 14. An illustration of the temporal dynamics of connections for a simulation where the Spread Controller cell is excited for a duration of 80 ms with onset at 0 ms. Each image shows the activity of V2 bipole cells for a 50 ms interval, e.g., the first image represents activity from the time period 50-100 ms after stimulus onset.

Here, the stimulus is the set of squares from Figure 5. Top-down control is imposed for a duration of 80 milliseconds (ms) beginning at the appearance of the stimulus. Because it takes time for neurons to respond to changes in stimulation (in particular for the interneurons in Figure 13 to recover from the inhibition provided by the top-down control), connections are able to spread for around 200 ms. With these parameters, the horizontal bipole cells aligned with the top and bottom edges of the four leftmost squares connect and so do the two rightmost squares. However, not enough time has elapsed for the two groups to connect. Once the effects of top-down control have dissipated, the interneurons start to block spreading of signals. This is due to the interneurons resuming their AND-gate function described above: asymmetric contours are inhibited, and symmetric contours are maintained. In more detail, wherever there is sufficient imbalance between the inputs to the left and right interneurons of a given bipole cell, one of the interneurons will be activated and will inhibit its associated bipole cell. In essence, this means that the end of a line of connections erodes back to wherever there is bottom-up stimulation (Francis et al., 1994). This erosion can be seen in Figure 14 by observing how the green (horizontal) connections on the far right shrink back to the edges of the rightmost square. In contrast, connections that link stimulus driven edges remain active because each adjacent bipole cell along a connection provides nearly equivalent excitation and inhibition to each interneuron, so the interneurons hardly inhibit their bipole. For example, the connections between the middle two squares on the left remain active.

Figure 15 shows equilibrium results, i.e., the contours that are maintained after the system settles down, from four simulations, each with a different duration of top-down control excitation of the Spread Controller cell.

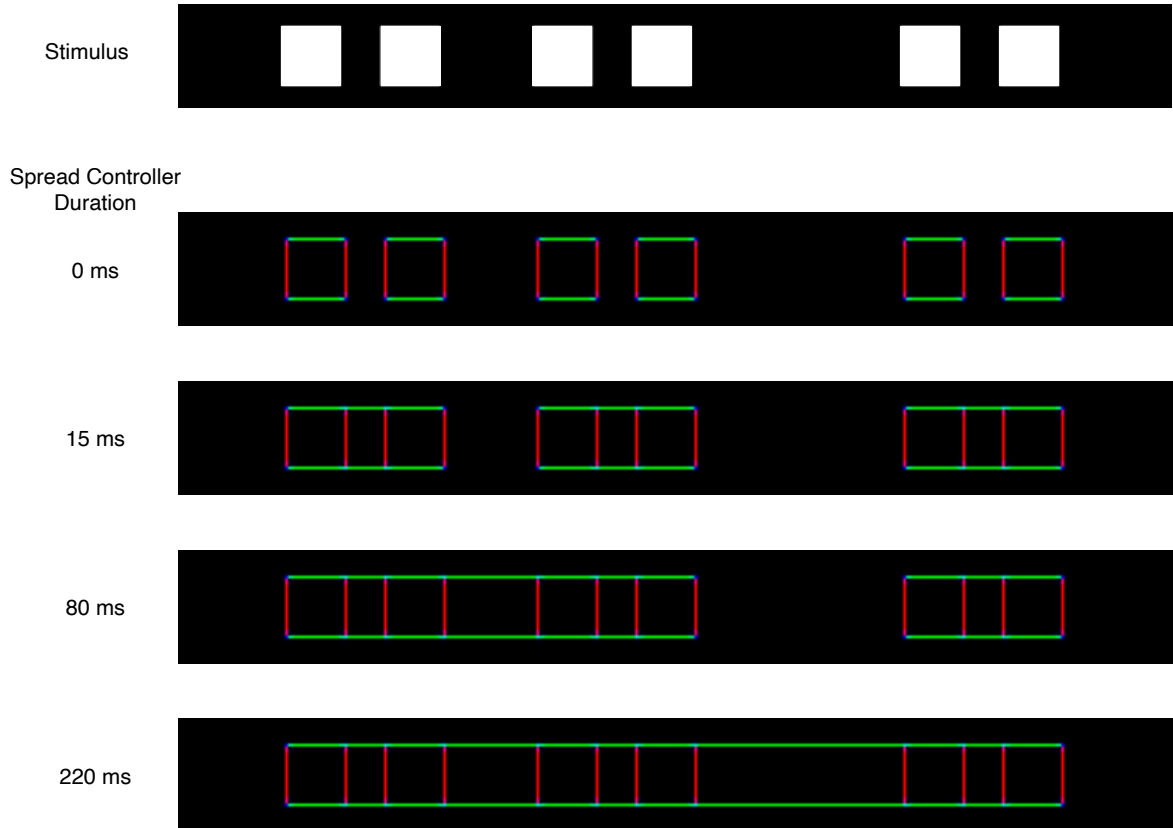


Figure 15. Altering the duration of excitation to the Spread Controller cell produces different connections between the detected edges for this stimulus. For each simulation, onset of top-down control excitation is at onset of the stimulus, i.e., at 0 ms. Here each image shows bipole cell activity summed over 50 ms at 1700 ms after stimulus onset. Each of these simulations was run for 1700 ms (model time) because this was the time at which the simulation with the longest Spread Controller duration (220 ms) settled down in to an equilibrium state.

For a 0 ms duration, there is no top-down control at all, and no connections form between the elements; the only boundaries are those generated by the stimulus edges. In this case, no inhibitory signal is sent to the interneurons, which results in V2, layer 2/3, operating as described in the previous section: V2 connections do not spread because they are kept in check by excitatory input from neighboring bipole cells to the interneurons. For a positive duration, the connections spread unchecked for a period of time before the interneurons resume their AND-gate function. The longer this duration, the farther the connections can spread. For a duration of 15 ms, each square connects to its nearest neighbor. For a duration of 80 ms, connections form among the four leftmost squares and between the two rightmost squares. For 220 ms, all six squares connect with each other. Thus, by altering a single timing parameter, the Spread Controller Circuit in Figure 13 allows for flexible groupings that emulate how observers can perceive Figure 5 in different ways.

So far we have seen how stimulus elements can be linked using horizontal connections. The Spread Controller Circuit is orientation specific, i.e., it can encourage the spread of active bipole cells that are tuned to a specific orientation independently of active bipole cells with different preferred orientations. Figure 16 shows the connections resulting in five simulations where top-down control duration was varied for horizontal and vertical orientations.

With different combinations of durations for these orientations, the circuit can promote a wide variety of connections.

Recall that the Spread Controller Circuit is influenced by two top-down control parameters: duration and onset. So far in this discussion, onset has been kept fixed at 0 ms. Depending on the stimulus, varying Spread Controller onset may or may not alter what connection patterns are possible. For some stimuli, shifting the onset of the top-down control excitation to the Spread Controller cell can produce the same set of connections provided that the excitation duration is also changed. For example, Figure 17 shows that different connections can be formed among pairs of horizontal and vertical lines.

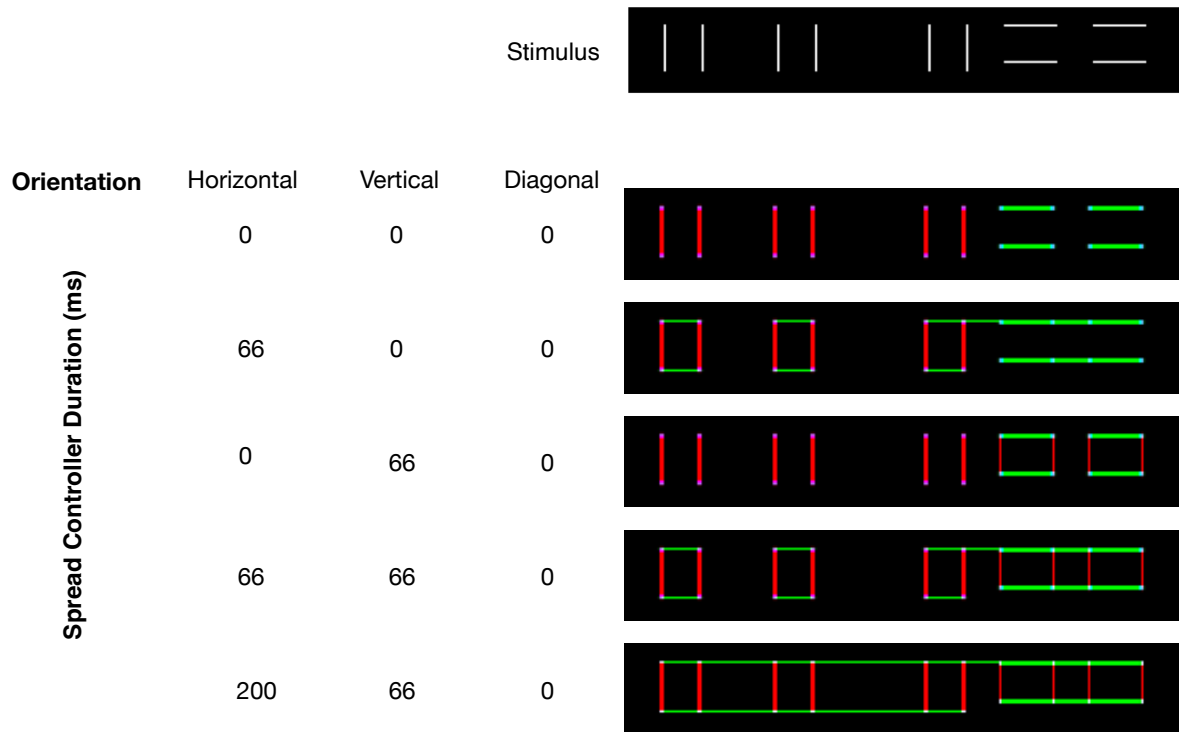


Figure 16. Demonstration that the spread of V2 connections is orientation specific and, additionally, that each orientation has its own duration of top-down control. The images show bipole cell activity summed over 50 ms at 700 ms after stimulus onset. The top-down control excitation began at stimulus onset, i.e., at 0 ms.

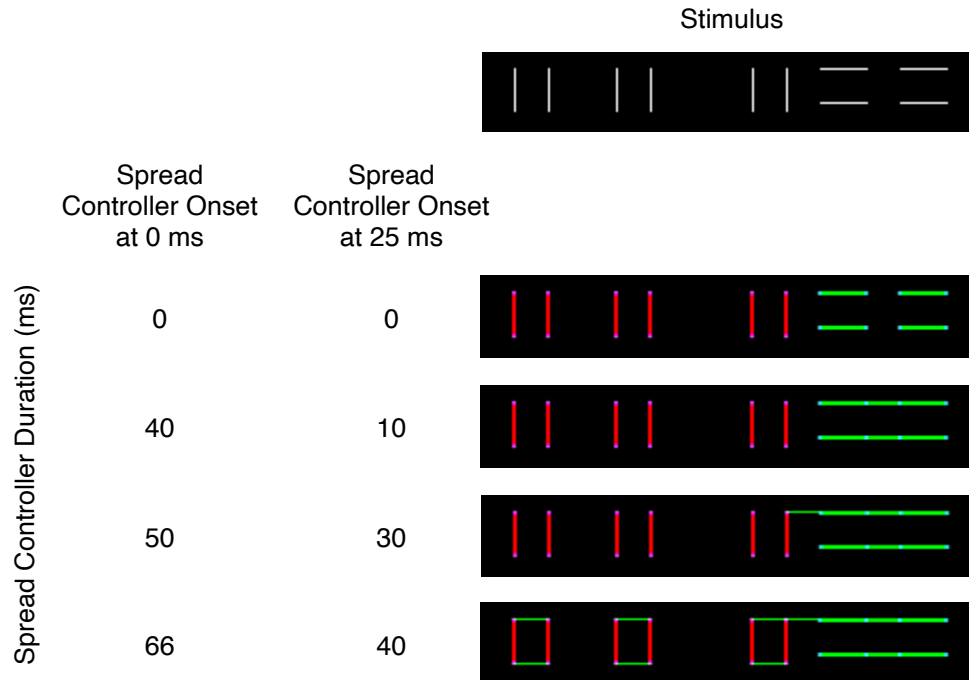


Figure 17. An example of how different combinations of onset and duration of the top-down control excitation can produce the same patterns of connections. Each image shows bipole activity summed over 50 ms from 700-750 ms after stimulus onset.

The same connections can be produced with an early onset of Spread Control excitation (0 ms, stimulus onset) or with later onset of Spread Control excitation (25 ms after stimulus onset). To produce equivalent connections, the later onset requires a shorter duration of Spread Control excitation. This is because it takes time for activity to propagate through the system after the stimulus is shown and, in turn, bipole cell activity is relatively weak during the first 50 ms of stimulation. Due to this initially weak bipole signal, a longer duration of Spread Control input is needed to produce the same connections when control input onset is at 0 ms compared to a simulation where input onset occurs after stimulus onset.

For other stimuli, a different onset will sometimes produce different connection possibilities. For example, consider the stimulus consisting of equidistant shapes in Figure 18.

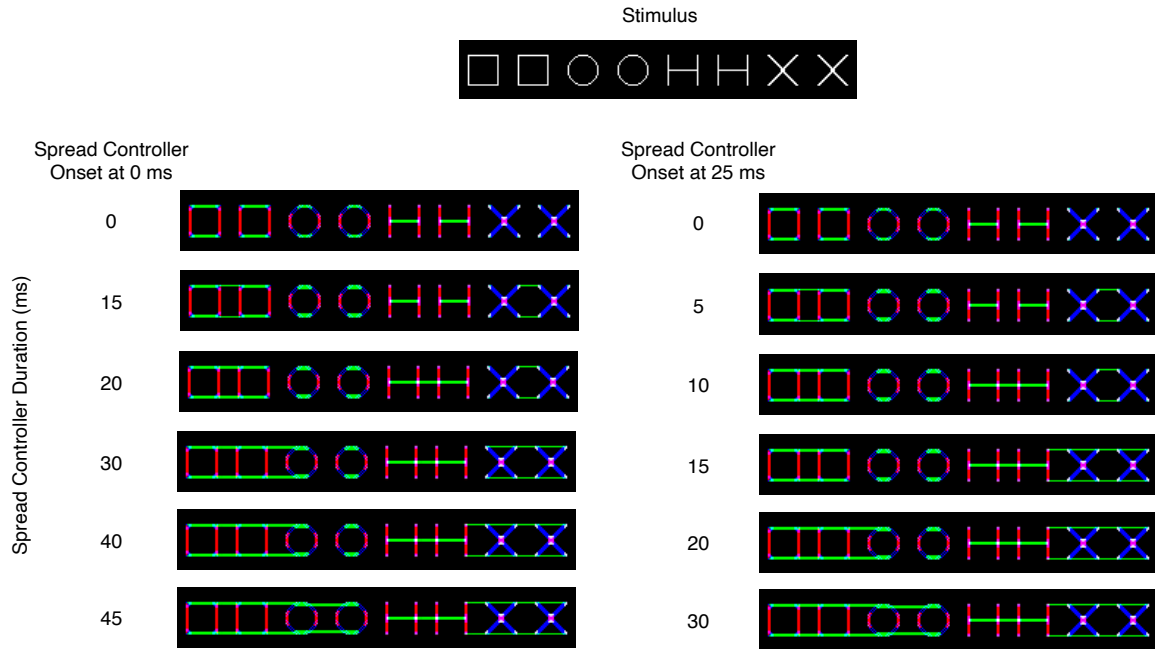


Figure 18. An example of how varying the onset of the top-down control duration can produce different patterns of connections. The images in row 4 have different connection patterns. Each image shows bipole cell activity summed over 700-750 ms after stimulus onset.

Each column shows the possible connections for a particular Spread Controller onset. Compare the images in each row. The connections in all but the fourth row are identical, which is similar to the results in Figure 17, where different combinations of onset and duration produce the same connections. In row 4 of Figure 18, the left image has connections between the square and circle, while the right image does not. Moreover, the left image lacks connections between the H and X, while there are such connections in the right image. In fact, there is no duration with onset at 0 ms that produces the connections shown in row 4 of the right column, and similarly for an onset of 25 ms and the connections in row 4 of the left column. This is due to the boundary signal being relatively weak at the beginning of a simulation. The single pixel horizontal edge of the H top/bottom is quite weak for a longer period of time than the top/bottom of the circle. Thus, a longer top-down control duration is required to connect the (weak) H and X (e.g., Figure 18, column 1, row 5), but this long duration also causes the top and bottom of the (strong) circle to connect with the squares. In contrast, with a 25 ms onset of the Spread Controller, the horizontal top/bottom of the H is strong enough to spread to the X before the top/bottom of the circle spreads to the square (because the connections from the circle must travel a longer distance). Thus, with a

0 ms onset, the relative strength of the circle signal allows it to overcome the handicap imposed by distance since here the H top/bottom is handicapped by being very weak. Similarly, differences in the activity of neurons near stimulus onset compared to later in the simulations explain why settings with the same top-down control duration but different onset do not produce identical connections. For later onsets, even with different durations, the circuits would produce nearly identical connection patterns.

Figure 19 contains parameter maps that demonstrate the abilities and limitations of the Spread Controller for the images shown at the top of each map.

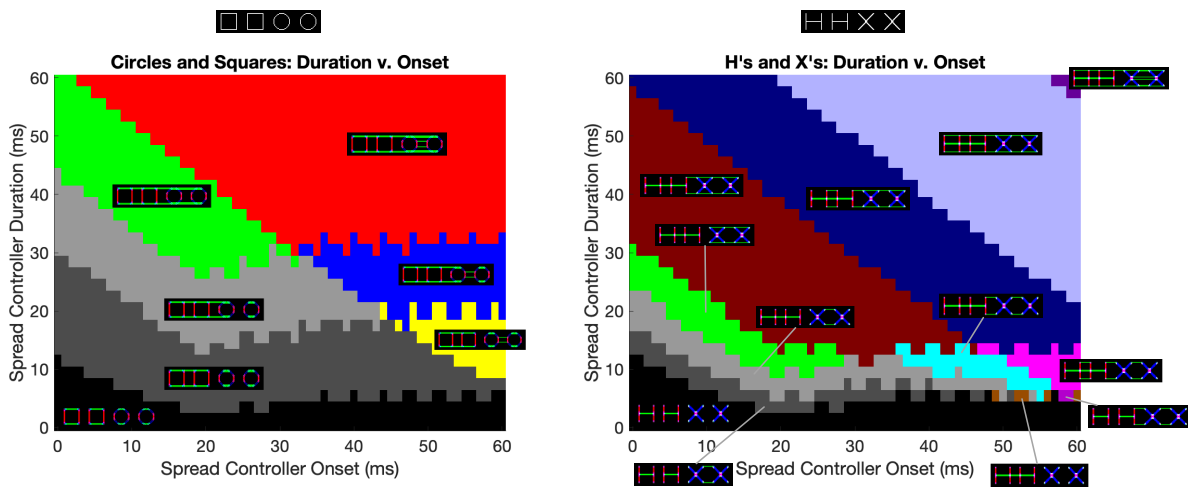


Figure 19. Parameter maps describe the possible connection patterns for a small image of squares and circles (left) and H's and X's (right). A simulation was conducted for each pair of duration and onset plotted here, where duration and onset values ranged from 0 to 60 ms in increments of 1 ms.

Each parameter map was created by running a simulation for each pair of values in the map. For the Figure 19 maps, the values for onsets and durations ranged from 0 to 60 ms in increments of 1 ms. Each simulation was run for 750 ms because the system reached an equilibrium state by this time. I identified sets of parameter values that produced the same connection pattern and used these to produce the parameter maps. Simulations for all parameter maps were run in parallel on three systems: a virtual machine running Ubuntu 20.04.2 (24 cores, 60 GB RAM) on XSEDE Jetstream (Stewart et al., 2015; Towns et al., 2014), a 2021 M1 Max MacBook Pro (10 cores, 64 GB RAM), and a 2019 Intel MacBook Pro (8 cores, 32 GB RAM). Checks of the different computing systems indicate that they give the same results. All simulations were programmed

using Python scripts with the package NEST 2.14.0 (Peyser et al., 2017) for creating the cells and synapses and for managing network dynamics.

Each pixel in a parameter map corresponds to a pair of onset and duration values. The color of a pixel indicates the resulting connections produced by these parameter values, and the image on/near each colored area shows the resulting bipole activity (700-750 ms after stimulus onset) and connection pattern at equilibrium. Consider the parameter map on the left. Consistent with the description of the Spread Controller, when Spread Controller duration is small, no elements connect with any other element (the black region on the bottom of the parameter map), regardless of the Spread Controller onset. When the Spread Controller duration is long (the red and green regions), every element connects with its nearest neighbor (in slightly different ways for the red and green regions, respectively). The other colored regions indicate parameter pairs that produce slightly different connections involving how the elements connect (or not) with each other.

These Spread Controller parameters play a similar role in the connection patterns for the stimulus image on the right of Figure 19. With a short Spread Controller duration, elements do not connect with each other (black region on the bottom of the parameter map). With a long Spread Controller duration all elements connect with their neighbors (burgundy, dark blue, purple, light purple, cyan, and magenta regions) in different ways. As discussed above, delayed onset of the Spread Controller signal promotes more connections even for short Spread Controller durations (as indicated by the negatively sloped boundary between different colored regions). The other colored regions (intermediate Spread Controller durations) indicate connection patterns where some elements do not connect with their neighbors.

At least for these stimuli, a given connection pattern is generally produced by a relatively large set of similar Spread Controller parameter values. This property suggests that such connections are not “fragile” and that top-down control could reasonably be defined, even with some noise, in such a way as to promote a desired set of connections. However, there are a few esoteric connection patterns, such as the pattern corresponding to the brown pixels near the lower right of the H’s and X’s parameter map, that should be difficult to produce because they depend on a precise set of parameters.

Unsurprisingly, the Spread Controller seems to implement the Gestalt principle of grouping by proximity, where “close” elements are defined in various ways by the onset and duration of the Spread Controller, e.g., it can form groups based on proximity for the row of squares in Figure 15.

However, the variability provided by this circuit is not enough to account for how humans seem to group some sets of visual elements. In the example stimulus in Figure 17, manipulation of the Spread Controller cannot lead to “grouping by orientation,” where there are connections only between lines of the same orientation, with no connections between any horizontal and vertical lines. The problem is that horizontal Spread Controller durations that allow connections to form between the tops/bottoms of the vertical lines on the left also lead to connections between the upper horizontal lines with the top of the rightmost vertical line. The stimulus in Figure 18 provides another example of how the Spread Controller Circuit is insufficient to capture some perceived groupings, e.g., connect only each pair of same-shaped elements. This ability is oftentimes labeled an instance of the Gestalt “similarity principle.” Proximity control afforded by the Spread Controller Circuit does not support such an ability for this stimulus because there seems to be no top-down Spread Controller onset and duration that leads to connections only between same-shaped pairs.

To deal with these kinds of limitations, I speculate that there are two additional circuits: a Long Controller and a Short Controller. For particular kinds of elements, they each reduce the spread of connections induced by the Spread Controller and, in turn, can be used to control which elements connect. I describe both of these connection circuits below and then explain how all three circuits work together to support different kinds of connection patterns.

Long and Short Controllers Overview

To provide even greater flexibility in the formation and control of connections between elements, I introduce two additional circuits that reduce the contribution to connections of bottom-up edges of specified lengths. Both of these circuits interact with the Spread Controller by exciting the interneurons that shut down connection formation. For the Long Controller, such excitation prevents boundaries of “sufficiently long” edges from spreading connections while allowing boundaries of shorter edges to spread. For the Short Controller, such excitation prevents boundaries of “sufficiently short” edges from spreading while allowing boundaries of longer edges to spread. In each circuit, the term “sufficiently” is established by the strength of top-down control input. Like the Spread Controller, the Long and Short Controllers are both orientation specific.

Typically, the top-down input for the Long or Short Controller is set to a fixed value for the duration of a stimulus, making the Short and Long Controller Circuits active for an entire

simulation. However, the impact of these circuits depends on the timing of the Spread Controller, because neither the Long nor Short Controller has much influence when the Spread Controller fully inhibits the interneurons (then connections spread from edges of all lengths). Thus, the Long and Short Controllers typically influence the formation of connections by affecting the interneurons before and shortly after onset and after offset of top-down excitation to the Spread Controller. If Spread Controller onset occurs with stimulus onset, the Long and Short Controllers typically influence connection formation only after Spread Controller offset, which, as is discussed in the next two subsections, can alter what set of connections form.

Long Controller Circuit

The Long Controller prevents a long edge (consisting of many pixels) in a stimulus from generating the outward spread of connections. Figure 20 shows how the Long Controller integrates into the connection circuit described in Figure 13.

The new components are emphasized with thick lines and orange color. A Long Control Accumulator at each pixel for each orientation has a large receptive field that gathers inputs from many oriented cells along the direction of the preferred orientation. The size of the large receptive field only needs to be big enough to cover long enough edges. The simulations shown here accumulated inputs from a receptive field of plus/minus 24 pixels around a given bipole cell pixel. If the Long Control Accumulator receives sufficient excitatory input, then it activates the bipole interneurons, and their activity shuts down the connection spreading process at that pixel. Without any top-down control signal, the Long Control Circuit prevents any sufficiently strong boundaries from spreading connections.

Top-down control, i.e., Long Controller input, enables the spread of connections from a long edge by exciting the Long Controller Cell, which inhibits the Long Control Accumulator. If this inhibition is sufficiently strong, then the Long Control Accumulator does not drive the interneurons and connections can spread from a long edge of input signals. For moderate values of Long Controller input, connections will spread for short edges but not for long edges. The effect of different Long Controller input values can be seen in Figure 21, left column.

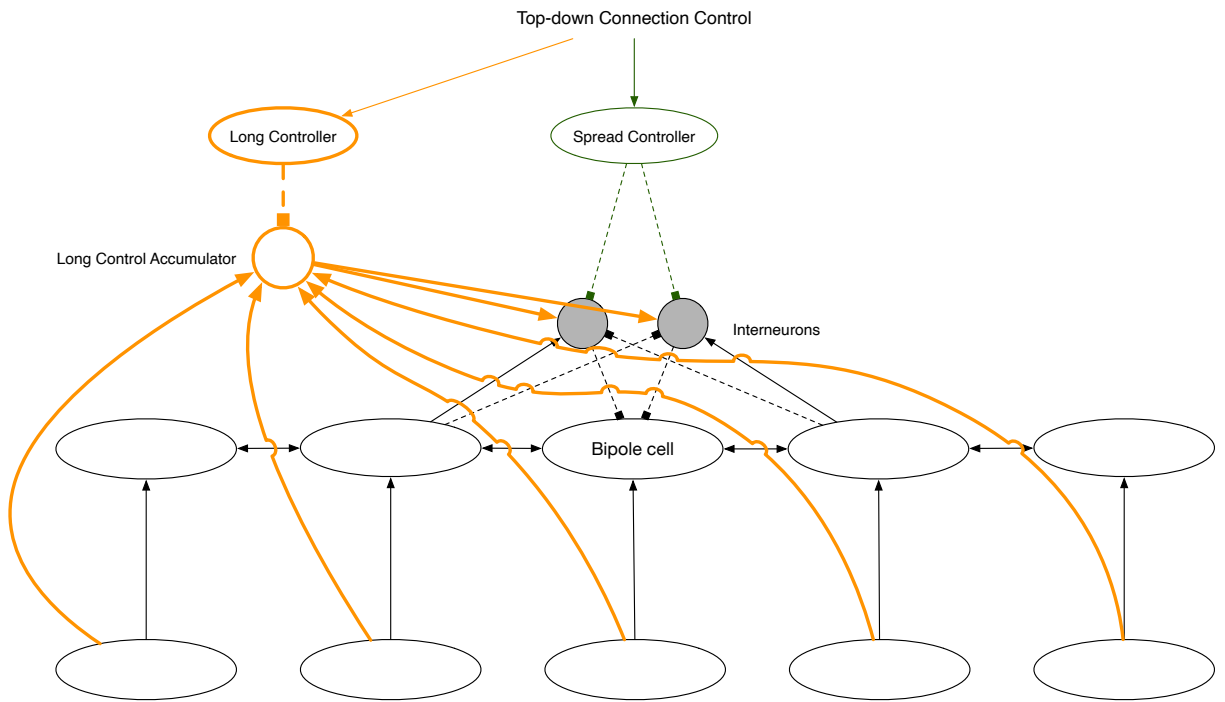


Figure 20. The components of the Long Controller Circuit for a horizontally-tuned bipole cell at a particular pixel location are shown in orange. Although (for the sake of reducing clutter in the image) only two cells (bottom row) to the left and right of this bipole cell are shown here, the horizontal Long Control Accumulator cell pools activity from 24 horizontally-tuned complex cells to its left and right.

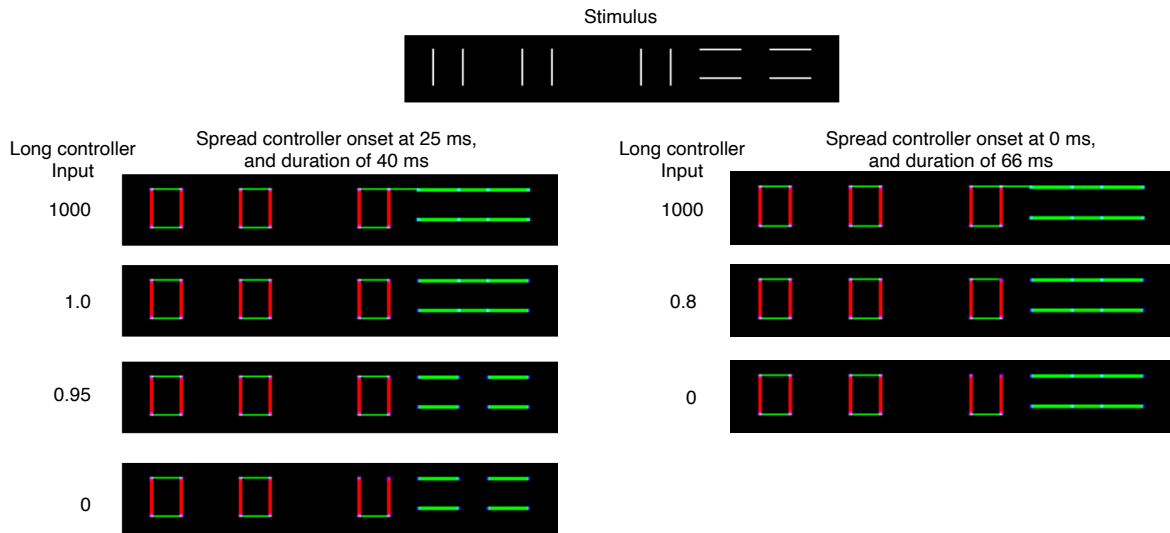


Figure 21. The connections resulting from simulations with different input to the Long Controller cell. The Long Controller Circuit is effectively off when the input to the Long Controller is very high, e.g., 1000, and has maximum restriction when the input is 0. For each simulation in the left column, Spread Controller onset was at 25 ms after stimulus onset and had a duration of 40 ms. For the simulations in the second column, Spread Controller onset was at 0 ms and the duration was 66 ms.

Here, the horizontal Spread Controller onset and duration were 25 ms and 40 ms, respectively. The vertical Spread Controller duration was set to 0, so no vertical connections form. For high Long Controller input (e.g., 1000), the Long Controller Circuit is effectively off: length of the input edges does not impact the formation of connections (the top-down control signal completely inhibits the Long Control Accumulator regardless of the number of signals feeding into its large receptive field). For the given Spread Controller duration, the very small horizontal edges at the top of each vertical line connect with their nearest neighbor. The top of the rightmost vertical line also connects with the longer horizontal lines on the far right.

With a weaker Long Controller input of 1.0, the spread from the long horizontal lines is reduced and can no longer reach to the top of the rightmost vertical line. This occurs because the long horizontal lines contribute a quite strong signal to the Long Control Accumulator neurons through their long range receptive fields. These strong signals can activate the Long Control Accumulator and thereby excite the interneurons, which then shut down the formation of connections. Thus, long edges do not form connections, but short edges (such as the tops of vertical lines) do form connections. For a slightly weaker input of 0.95, the long horizontal lines do not connect with each other.

With an even weaker value of 0, there is some “bleed” over to the top of the rightmost vertical line. The Long Control Accumulator cells for this line register the long horizontal edge to its right due to their large receptive fields. Figure 22, which shows V2, layer 4, activity over 50 ms, helps to visualize the location of these receptive fields for two bipole locations indicated by the center orange box of each ‘star’. For a given horizontally-tuned bipole cell, the pool consists of the 24 V2, layer 4 cells to the left and right of it that are horizontally-tuned, e.g., the horizontally-tuned cells that correspond to the positions of pixels in the horizontal white boxes in Figure 22. (On the other hand, if the cell at this location was vertically-tuned, it would pool from vertically-tuned V2, layer 4 cells above and below this location, e.g., from the vertically-tuned cells that correspond to the positions of pixels in the magenta boxes. Each diagonal Long Control Accumulator cell at that location pools V2, layer 4, activity, indicated by the relevant cyan box.)

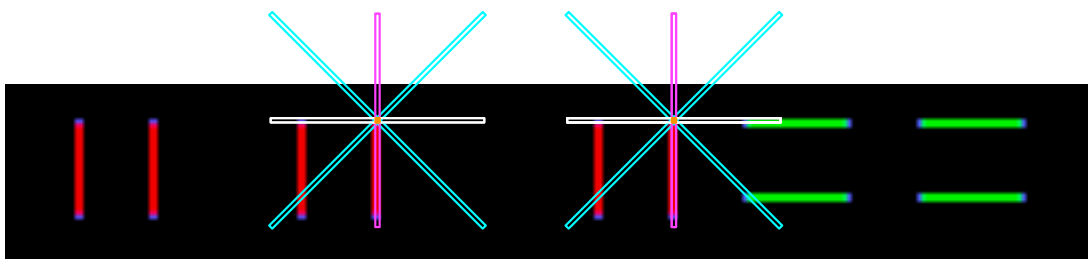


Figure 22. V2, layer 4, output from the grouping by orientation example to illustrate the difference between activity pooled by the Long Control Accumulator cell for a cell at the top of a vertical bar that is near the horizontal lines (right set of boxes) and that of a cell that is at the top of a vertical bar farther from the horizontal lines (left set of boxes). The white pair of boxes on the right overlaps with more active horizontally-tuned V2, layer 4 cells, and the Long Control Accumulator cell at the location indicated by the orange pixel at the center of these white boxes has stronger input since it pools from this set of cells. In contrast, the white pair of boxes on the left overlaps with far fewer and weaker horizontally-tuned cells. Since the Long Control Accumulator cell at the location indicated by the orange pixel between these white boxes is weakly excited, it does not excite the V2 interneurons as much. Thus, the Long Controller will more strongly excite the interneurons at the position on the right than the position on the left, and this prevents V2 connections forming between the far right vertical bars only given a Long Controller input of zero.

Regarding Figure 21 when input to the Long Controller is at 0, although the horizontal edges at the top of the vertical bars are quite short, the top edges of the far-right pair of vertical bars are near to and aligned with the top left horizontal bar. In effect, the Long Control Accumulator cells at these locations pool this signal along with that from the top left horizontal bar, an example of which is indicated in Figure 22 by the green pixels in the right white box. This

causes strong excitatory signals to be sent to the horizontally-tuned interneurons at the tops of the far-right pair of vertical lines that, in turn, inhibits their horizontal spread and prevents them from connecting. The ends of all other vertical lines, on the other hand, although also pooled by Long Control Accumulator cells, may only slightly excite the interneurons at those locations because the horizontal signal that these Long Control Accumulator cells end up pooling is quite weak (which is indicated by the green pixels in the left white box); yet this small excitation is not enough to prevent these short edges from connecting. In turn, the top edges of the rightmost pair of vertical do not generate connection spreading, while the other pairs of vertical lines maintain their connections even with a Long Controller input of 0.

In the simulations discussed thus far, the Spread Controller onset was fixed at 25 ms after stimulus onset. However, the Spread Controller onset interacts with the Long Controller to determine what connections are possible. For example, when Spread Controller onset is the same as stimulus onset, i.e., is at 0 ms, a somewhat different set of connections can be produced as shown in the right column of Figure 21. The connections in the top two rows are essentially the same (even when having slightly different values for the Long Controller).

With a Spread Controller onset of 25 ms, the simulations on the left side of Figure 21 show that it is possible (with Long Controller input between 0 and 0.95) to connect the vertical lines without connecting the long horizontal lines. Such a connection pattern is not possible when the Spread Controller has an onset of 0 ms and a duration of 66 ms (simulations on the right side of Figure 21). With the earlier appearance of Spread Controller inhibition, the interneurons are almost fully inhibited before signals from the Long Controller come in to play. With such strong inhibition, both long and short edges contribute to connection formation even when the Long Controller is at its maximum with an input of 0 (Figure 21, right column, last row). In contrast by delaying the Spread Controller onset by 25 ms and having 0 Long Controller input, long edges tend to generate excitation from the Long Control Accumulator to the interneurons just before the Spread Controller cell inhibits the interneurons. Thus, for a short time, the long edges cannot contribute to the formation of connections. Should the Spread Controller remain active long enough, eventually even the long edges would contribute to connection formation. But here, the Spread Controller turns off before the interneurons corresponding to long edges would be fully inhibited, so the long edges never get a chance to contribute to connection formation.

In sum, generally when the Long Controller Circuit is on, the connections that spread from long edges do not travel as far as those from shorter edges. This is due to the way in which the Spread Controller Circuit and Long Controller Circuit interact and the design of the Long Controller Circuit. In the Figure 21 example with Spread Controller onset at 0 ms, the Spread Controller inhibits the V2 interneurons for all horizontal edges for 40 ms and thereby encourages the spread of all horizontal edges, while the Long Controller constantly excites these interneurons in proportion to the length of the edge. In other words, the longer the horizontal edge (or, more precisely, the more horizontal signal that the Long Control Accumulator cell pools), the more the spreading of horizontal connections is inhibited. In turn, the Long Controller Circuit allows the Connection Circuit more flexibility: an observer can tune this parameter to connect pairs of vertical lines and horizontal lines without connections between vertical and horizontal lines and, at least for this stimulus, group elements by orientation.

While the specific connection patterns can vary a lot for different combinations of Spread Controller duration and Long Controller input, the parameter maps in Figure 23 demonstrate that the circuit behavior follows fairly simple rules that govern the connection patterns.

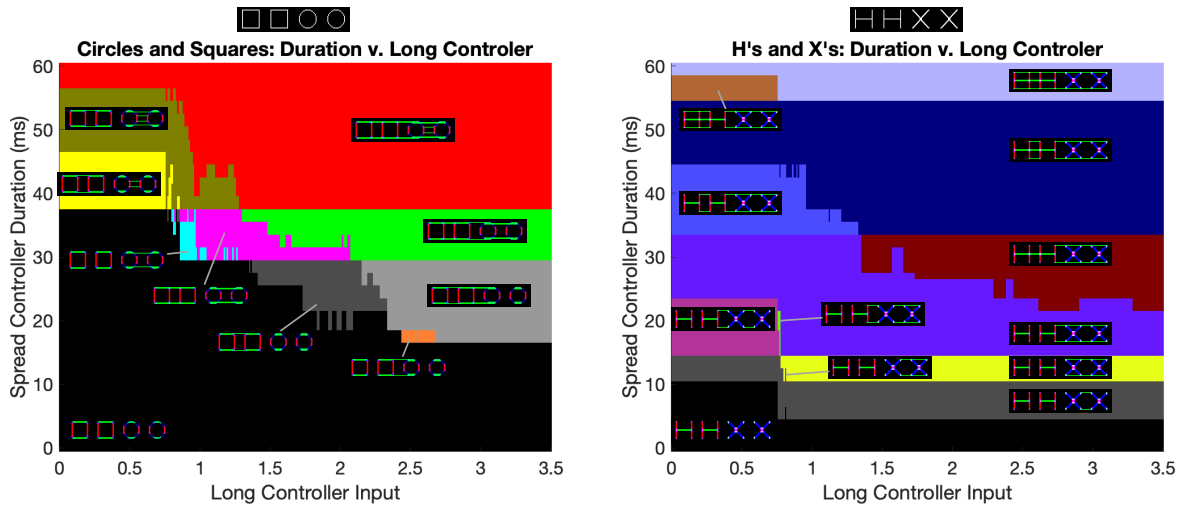


Figure 23. Parameter maps for the stimulus shown at the top of each plot. To provide a fine-grained picture of connections possible with different Long Controller inputs, Spread Controller duration ranged from 0 to 60 ms in increments of 1, and Long Controller input ranged from 0 to 3.50 in increments of 0.01. For all simulations here, the Spread Controller input was turned on 25 ms after stimulus onset.

Longer Spread Controller durations and larger Long Controller input values allow for more connections to form between elements (e.g., the red region in the upper right of the parameter map on the left side of Figure 23). With some fine tuning of the parameter pair (e.g., the yellow, olive green, and magenta regions), it is possible to form connections between squares and between circles, but not between a circle and a square. For very specific parameters values (e.g., the orange region around input 2.5 and duration 18 ms), it is possible to form connections between a square and a circle but not between same-shaped elements. I suspect that such a connection pattern would be difficult to maintain due to cortical noise.

The parameter map for the H's and X's image on the right side of Figure 23 shows similar relationships between Spread Controller duration and Long Controller input. Most connection patterns can be produced by a set of similar parameters, but a few connection patterns are very fragile.

Although it enables a variety of connection patterns, the Long Controller still has some limitations. The parameter maps indicate that the Long Controller may be adjusted so as to connect all and only same-shaped pairs in the stimulus with squares and circles (the yellow, olive green, and magenta areas in the left map of Figure 23), but there is no pair of parameters that form connections between the H's and connections between the X's, while not connecting the middle X and H. Thus, a connection circuit with a Long Controller cannot connect the elements so as to facilitate grouping by shape similarity for this stimulus.

Short Controller Circuit

The Short Controller Circuit reduces the spread of connections from short edges. Figure 24 shows how the Short Controller Circuit integrates into the bipole circuit described in Figure 20.

The new components are drawn with thick red lines. Similar to the Long Controller, the Short Controller Circuit operates by exciting the interneurons of the modified bipole circuit to prevent connection spreading. This excitation is provided by the Short Control Trigger cell, which exists at each pixel location. The trigger cell receives excitation from the complex cell at its pixel location (indirectly through an interneuron in order to equate the number of synapses for other inputs, which allows the inhibition from the Short Control Accumulator and the excitation from the gray interneuron with a red border to reach the Short Control Trigger at roughly the same time). The trigger cell also receives inhibition from the Short Control Accumulator. The accumulator

gathers excitatory input from nearby pixels that align with the bipole cell's orientation preference. The size of the accumulator receptive field does not matter much; I used 5 pixels on either side of the bipole cell in my simulations.

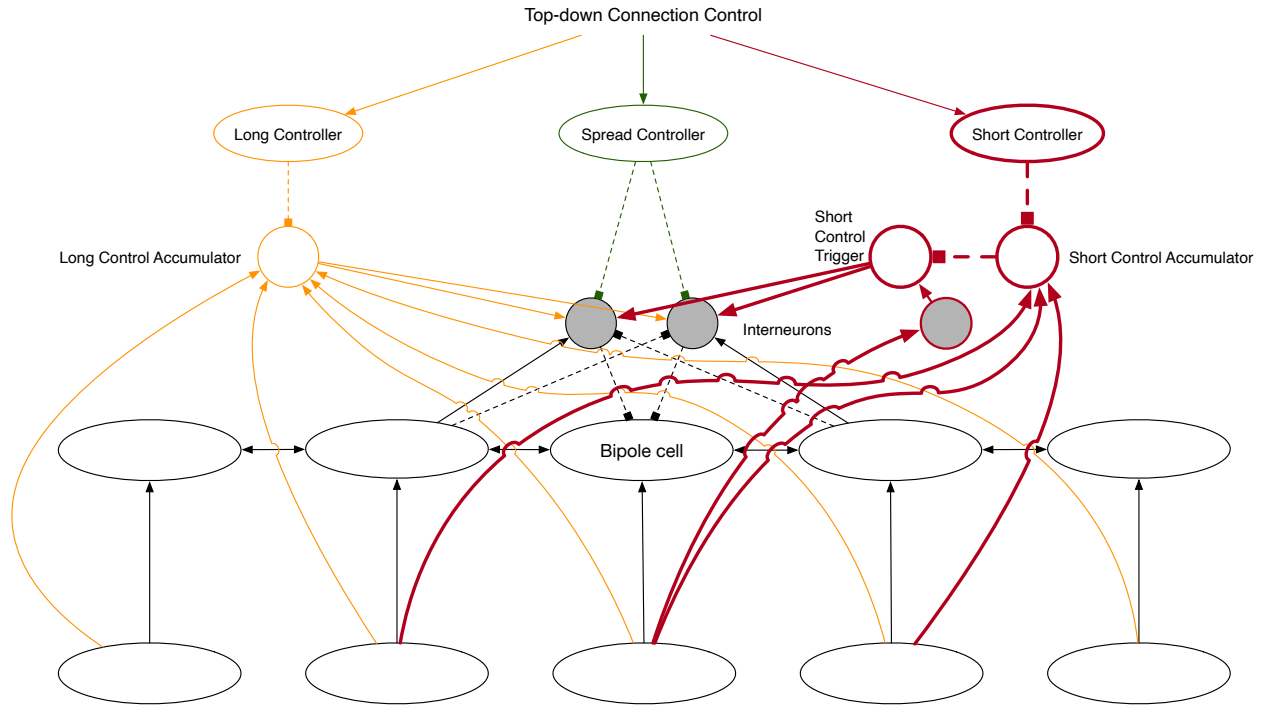


Figure 24. The components of the Short Controller Circuit are shown in red for a horizontally-tuned bipole cell. Although, for the sake of simplicity, this schematization indicates that the Short Control Accumulator only pools the activity from two complex cells corresponding to locations neighboring that of the bipole cell (and the complex cell that feeds into the bipole cell), in the simulations each horizontal Short Control Accumulator cell pools activity from the five horizontally-tuned complex cells (bottom row) that are to the left and right of the bipole's pixel location.

Let us first consider the circuit's behavior with no top-down control (input equals zero). Suppose the bipole cell of interest is in the middle of a long set of edges that covers the full receptive field of the Short Control Accumulator. The Short Control Accumulator will be strongly excited and will strongly inhibit the Short Control Trigger. Since the Short Control Trigger only receives input from a single complex cell (the cell at the same corresponding pixel location as the bipole cell, which is directly below the cell labelled "Bipole Cell" in Figure 24), it is inhibited by the Short Control Accumulator, which receives stronger net input. Thus, in this case, the Short Control Trigger does not excite the bipole circuit's interneurons, so connections spread normally (depending on whether the Long Controller or the Spread Controller Circuits are excited). Now,

suppose the horizontally-tuned bipole cell of interest is nearly isolated (say, it is the top of a thin vertical line). Then, the Short Control Accumulator receives little excitation that is nearly equivalent to what is received by the Short Control Trigger cell. The synapse strength between the complex cell and the Short Control Trigger is fairly weak, so the trigger is inhibited, in net. Thus, without any top-down control, i.e., without any input to the Short Controller cell, the Short Control Circuit allows spreading of both long and short edges.

The top-down short control input influences spreading by inhibiting the Short Control Accumulator. Figure 25 demonstrates the Short Controller's effects when this circuit is applied to the grouping by orientation example.

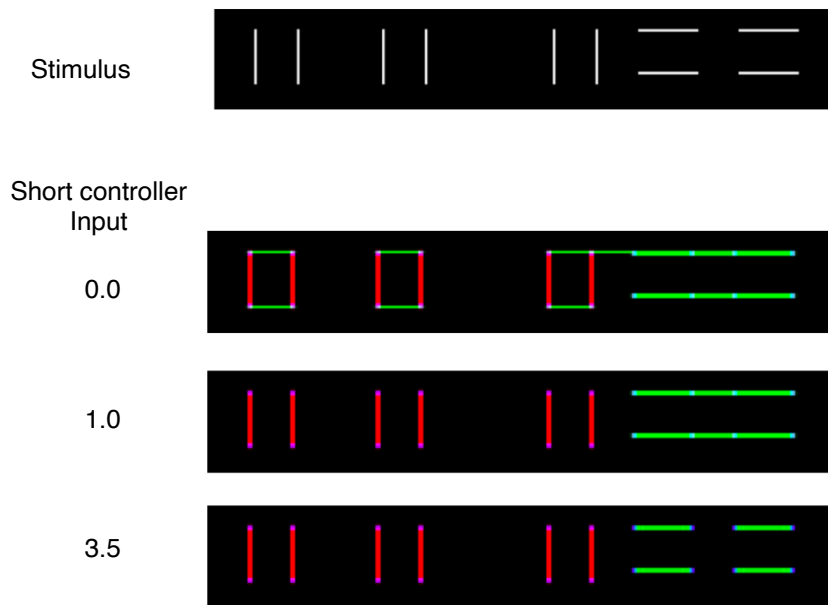


Figure 25. Connections formed for different top-down inputs to the Short Controller cell. Each image shows the activity of bipole cells for 50 ms of model time after 700 ms of a simulation. For each simulation, the other connection parameters were fixed: Spread Controller duration was set to 40 ms (beginning at 25 ms), and the Long Controller Circuit was effectively turned off with an input of 1000.

Horizontal Spread Controller onset and duration are fixed at 25 ms and 40 ms, respectively. The delayed onset allows the Short Controller Circuit to influence the interneurons as the Spread Controller starts to allow connections to form. As the simulations show, increasing the Short Controller input blocks short edges from contributing to connection formation. This control happens because the excitatory bottom-up input to the Short Control Accumulator needs to exceed

the inhibitory top-down input in order to prevent the Short Control Trigger from halting connection spread. As shown in Figure 25, when the top-down Short Controller input is 1.0, the short horizontal edges at the end of each vertical line do not generate connections. This occurs because such short edges do not provide sufficient excitatory input to the Short Control Accumulator to overcome the inhibition from the excited Short Controller cell, so the Short Control Trigger halts connection spreading. In essence the magnitude of the top-down Short Controller input determines the definition of “short” for a given stimulus. As shown in Figure 25, with a value of 3.5, the horizontal lines on the far right are deemed “short” and do not generate connections.

Together, the Long and Short Controllers can capture some sense of “similarity” between elements. Figure 26 shows simulation results for the row of shapes in Figure 18.

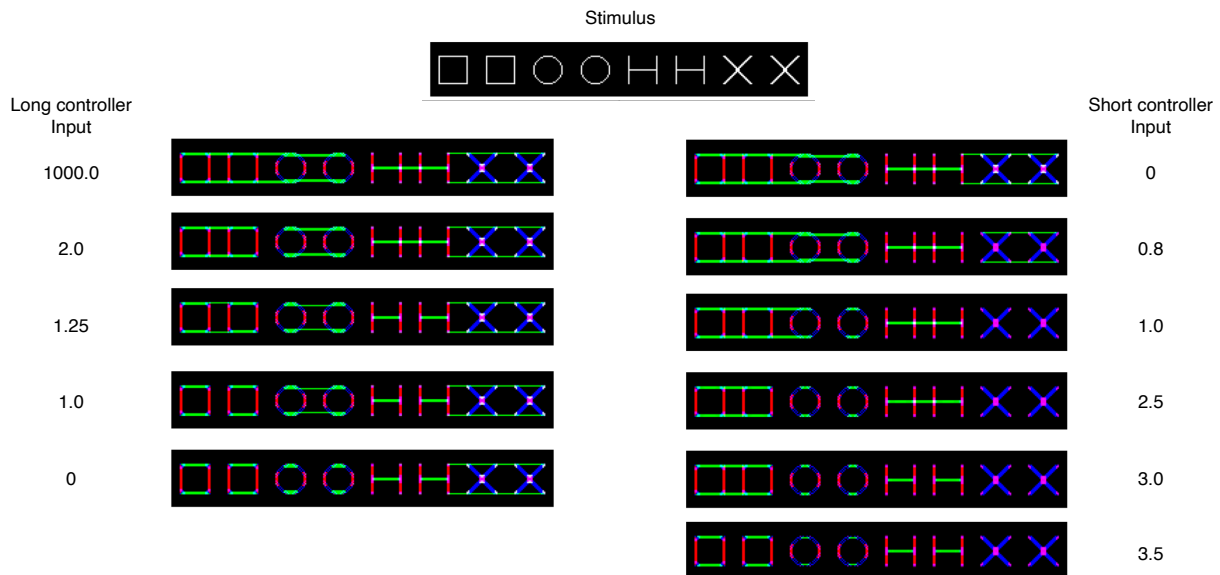


Figure 26. The impact of different inputs to the Long Controller (left column) and Short Controller (right column) for the stimulus at the top. For each simulation, Spread Controller duration was 30 ms and Spread Controller onset was at 25 ms, i.e., Spread Controller parameters at which at least all similarly shaped pairs were connected. For the left column, which demonstrates the effect on connections of different Long Controller input values, the Short Controller (Long Controller) was turned off for each simulation by having the top-down Short Controller input set at 0. To demonstrate the effect of different Short Controller input values, the Long Controller was turned off for each simulation in the right column by having the top-down Long Controller input set at 1000.

For all of these simulations, the Spread Controller onset is 25 ms with a duration of 30 ms. As we saw before, these Spread Controller parameters are insufficient, by themselves, to connect all and only each same-shaped pair. The simulation results on the left side of Figure 26 vary the

input to the Long Controller while keeping the Short Controller input at 0 (so that the Short Controller does not influence connection spreading). With large values of Long Controller input, all edges (both short and long) can easily spread connections, so elements connect both within and between pairs. For smaller Long Controller inputs, long edges are less able to spread connections. This impact is most notable for the value 1.0, where the long edges of the squares no longer connect but the small horizontal edges at the tops of the X's are able to connect (but they also connect with the nearby H).

The simulation results on the right side of Figure 26 vary the input to the Short Controller while keeping the Long Controller with an input of 1000 (so that the Long Controller does not influence connection spreading). For small top-down inputs to the Short Controller, both short and long edges can spread connections, so elements connect both within and between pairs. With large values of Short Controller input, only long sets of edges can spread connections. This impact is especially noticeable for the value 2.5, where the pair of squares and the pair of H's connect, but the pair of circles and the pair of X's cannot connect. For a large enough input value to the Short Controller (for these stimuli a value of 3.5 is sufficient), essentially all edge sets are deemed "short" and cannot spread connections.

The Short Controller interacts with the Spread Controller. Figure 27 plots two parameter maps, for different stimuli, that categorize different connection patterns for various combinations of Spread Controller duration and Short Controller input.

For these simulations, the Spread Controller onset was always 25 ms and the Long Controller was off. The resulting connections are robust in the sense that each possible pattern of connections can be produced by many pairs of duration and Short Controller input values, as indicated by the relatively large areas in the Figure 27 parameter maps.

Although the short and Long Controllers capture some aspects of grouping by shape similarity, they are insufficient, on their own, to impose some desired groupings. In particular, if an observer wanted only the same-shaped pairs in Figure 26 to connect, they would not be able to do so only by manipulating one parameter. The next section shows how combinations of parameters allow for complex connection patterns that facilitate a variety of Gestalt groupings.

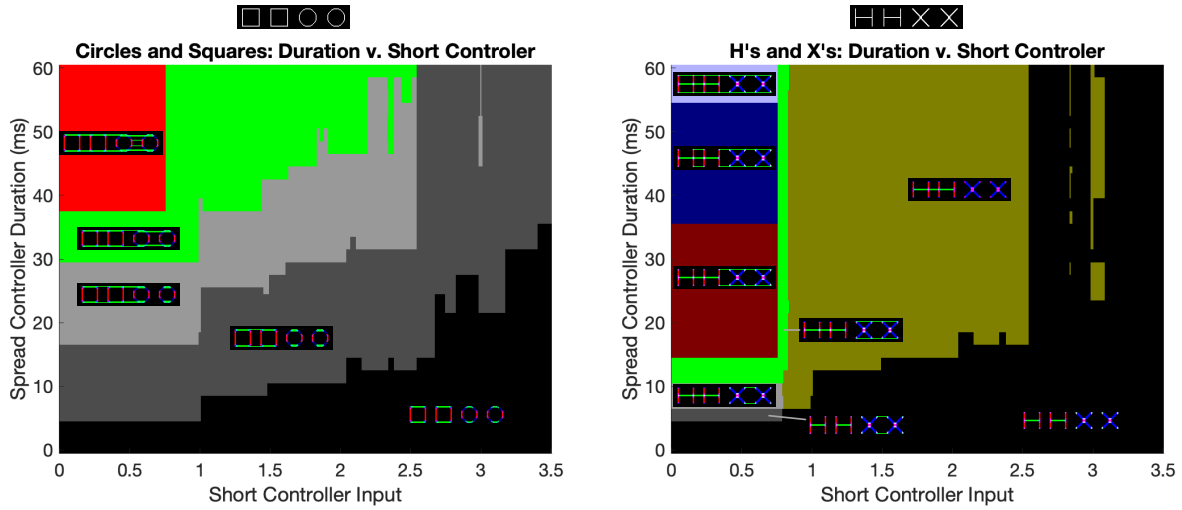


Figure 27. Parameter maps of Short Controller input and Spread Controller duration for the stimulus shown at the top of each plot. To provide a fine-grained picture of connections possible with different Long Controller inputs, Spread Controller duration ranged from 0 to 60 ms in increments of 1, and Long Controller input ranged from 0 to 3.50 in increments of 0.01. For all simulations, Spread Controller onset was 25 ms after stimulus onset and, thus, the second column of Figure 26 provides some examples of parameter values and resulting connections that are shown in these maps.

Combining Connection Controllers to Promote Gestalt Groupings

By adjusting the onset and duration of the Spread Controller and the top-down inputs for the Long and Short Controllers, an observer can influence what elements in a scene form connections, and thereby influence perceived groupings. For example, we already saw in Figure 21 that a combination of Spread Controller onset and duration and Long Controller input can result in connections that can be described as grouping by orientation for this stimulus. With Spread Controller onset at 0 ms and a duration of 66 ms, and with Long Controller input of 0.8, only the lines of the same orientation are connected together.

Figure 28 shows a situation where a combination of Spread Controller onset and duration and Short Controller input can connect stimulus elements in a way characterized by the Gestalt principle of size similarity, i.e., each adjacent same-sized pair is connected but there are no connections between the pairs.

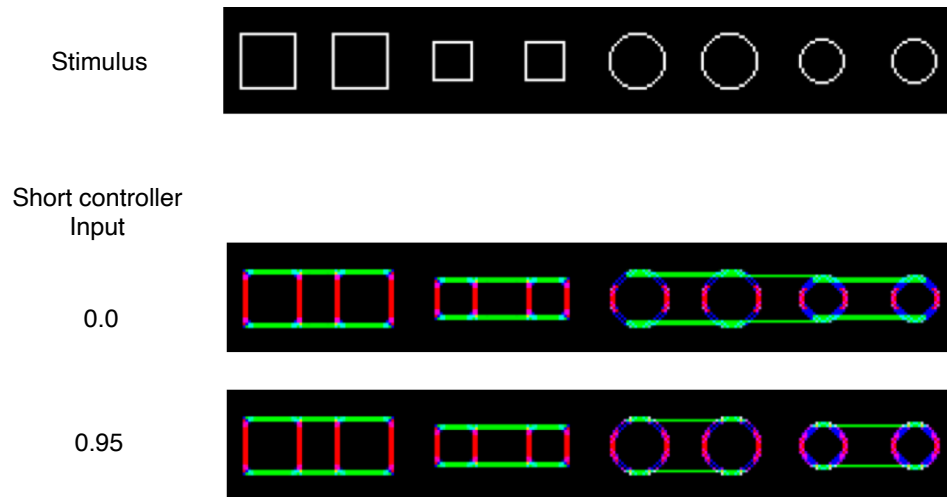


Figure 28. Model output for two simulations that have a Spread Controller onset of 0 ms and a duration of 55 ms for a stimulus that can be grouped by similarity of size. For the model to form connections between all and only same-sized elements, the Short Controller must be on with an input of around 0.95 in order to prevent the connections that result between the pairs of same-sized circles shown in the middle image.

Here, the stimulus consists of a pair of large squares, a pair of small squares, a pair of large circles, and a pair of small circles. For the squares, size grouping is easy to achieve even with no Short Controller input because the bipole connections spread only in the direction of the orientation preference (e.g., horizontal connections only spread horizontally). Due to their difference in size, the tops/bottoms of the small squares do not line up with the tops/bottoms of the large squares, so they cannot connect. However, due to the somewhat fuzzy response of an oriented filter to the top of a circle, there is some alignment in the horizontal responses to the large and small circles. When the Short Controller input equals 0, these weak horizontal signals connect across pairs. Such a grouping cannot be controlled simply by lowering the Spread Controller duration (here set to 55 ms with an onset of 0 ms) because doing so would prevent the formation of connections *within* the same-sized circle pairs.

Setting the Short Controller input to a value of 0.95 makes the weak/fuzzy horizontal signal generated at the top of a circle too weak to promote connections. The input value is chosen so that just the weakest parts of the fuzzy response are blocked; the stronger responses along the circle edge are “long” relative to the top-down controlled Short Controller input value, and so they form connections within each circle pair. As can be seen in Figure 28, with these top-down parameters, it is possible to form connections within each pair but not between pairs. It is the combination of top-down parameters that supports such groupings.

For the stimulus in Figure 18, all three connection circuits are needed to connect elements in accord with the Gestalt principle of grouping by shape similarity, i.e., all and only same-shaped adjacent pairs are connected. As can be seen in the simulation results in Figure 29, a combination of Spread Controller onset and duration and long and Short Controller inputs connects same-shaped pairs while preventing connections between pairs.

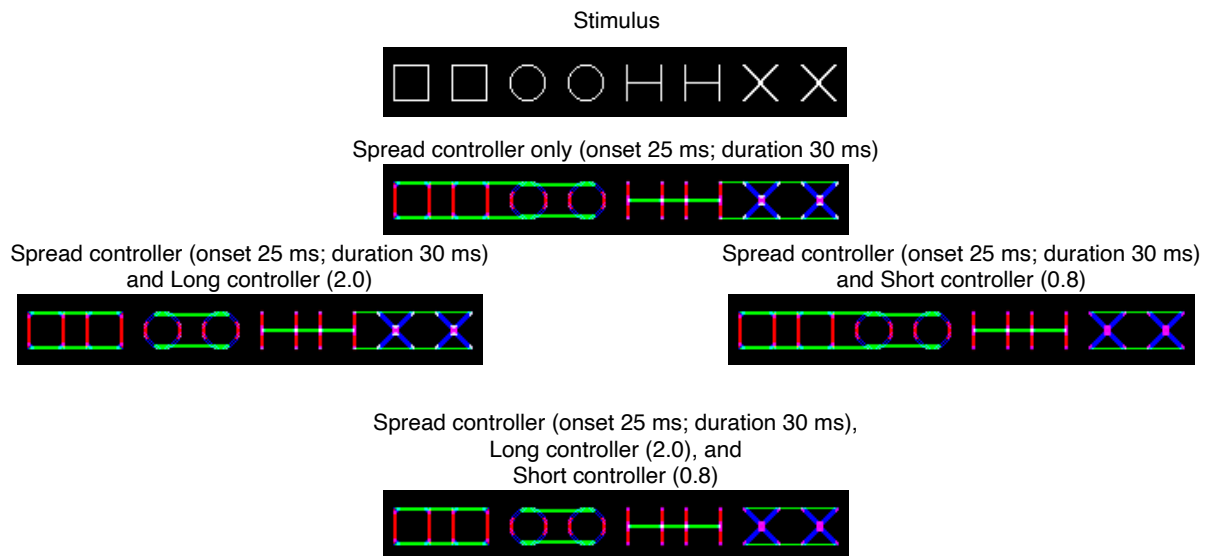


Figure 29. Model output for four simulations with different combinations of connection parameter values. For each simulation, Spread Controller duration is 30 ms with an onset of 25 ms. The Long Controller input of 2.0 eliminates connections between the square and circle, while the Short Controller input of 0.8 eliminates connections between the H and X. By combining these connection parameters, the model can produce connections that effectively group the stimulus elements by shape as shown in the bottom image.

Thus, with the appropriate top-down control parameters, it is possible to group the elements in this stimulus by shape similarity.

Figure 30 demonstrates that different combinations of the top-down parameters for the Spread Controller, Long Controller, and Short Controller Circuits can lead to connections that largely emulate the Gestalt principles of grouping for a variety of stimuli.

Grouping by	Row	Stimulus	Bipole cell activity	Top-down connection parameters			
				Spread controller onset	Spread controller duration	Long controller input	Short controller input
Proximity	1			0	70	Off	Off
	2			0	30	Off	Off
Similarity of orientation	3			0	66	0.80	Off
	4			0	5	Off	2.00
Similarity of size	5			0	55	Off	0.95
	6			0	75	Off	0.80
Similarity of shape	7			0	45	2.00	0.80
	8			0	45	2.00	Off
Closure	9			0	9	1.20	Off
	10			0	30	Off	Off
Symmetry	11			0	20	Off	Off

Figure 30. Grouping rules that can be accounted for by tuning the top-down control parameters of the connection circuits. The first proximity stimulus (row 1) is similar to examples in Köhler (1925). The stimuli with filled shapes (rows 2, 6, and 8) and the orientation stimuli (rows 3 and 4) are adapted from examples of “classical principles of grouping” listed by Palmer (1999; 2002). The first closure stimulus (row 9) appears in Köhler (1925), and the second (row 10) is from Pomerantz and Kubovy (1986). The symmetry stimulus (row 11) is adapted from Palmer (1999; 2002). Spread Controller onset was at 0 ms for all stimuli. For the Long Controller input, “off” means that a large value (1000) is provided. For the Short Controller input, “off” means that a value of zero was provided.

Rather than being “principles” in an absolute sense, I propose that such groupings reflect task-specific characteristics. For example, the stimulus in row 8 of Figure 30 may be regarded as exemplifying grouping by shape similarity (Palmer 1999; 2002). I hypothesize that such a statement motivates observers to search for top-down parameters that form connections between shape pairs and not form connections between different shapes. The ease with which such parameters are identified may be an indication of how well the given stimulus exemplifies the principle.

The parameter maps in Figures 19, 23, and 27, suggest that it should be fairly easy for observers to identify parameters that produce desired connection patterns. If current parameter values are insufficient to produce a desired connection pattern, the observer can change the values with a pretty good understanding of the impact. Indeed, this is precisely what I have done when producing the simulations presented in this section. To generate the parameter values in Figure 30, I fixed the onset of the Spread Controller to match stimulus onset and then found the shortest

Spread Controller duration where all elements of a stimulus that I wanted to connect were connected. Then, depending on whether long and/or short edges had undesired connections, I tuned the input parameter to the Long and/or Short Controller Circuit(s) to eliminate those connections. I speculate that human observers are quite skilled at tuning such parameters and do so more or less automatically.

Although sometimes used as a prescription for how elements in a scene group together, I believe the Gestalt principles are better understood as a high-level description of the output of simpler underlying mechanisms. (In the final chapter, I elaborate more on the relation between Gestalt grouping principles and the model.) With this view in mind, it is worthwhile to explain how the connection circuits are able to emulate the principles. In general, there are often several ways to form a desired connection pattern.

Proximity

The role of proximity on grouping reflects the way connections spread outward from an edge and must connect with another edge in order to maintain its representation. For the two proximity examples in Figure 30, it is easy to control the extent of spreading with the Spread Controller Circuit so that connections form between nearby neighbors and not between farther away elements.

Similarity of Orientation

Elements of a similar orientation tend to group together. This property reflects that the cells in a bipole circuit represent a preferred orientation. The stimulus in row 3 of Figure 30 was previously discussed at length. For the stimulus in row 4, Spread Controller duration for each orientation was set to 5 ms. The Short Controller is on to prevent short aligned horizontal edges from connecting between the top of the upper vertical line segment and top of the diagonal line segment.

Similarity of Size

For the stimuli in rows 5–6 of Figure 30, connections form only between elements of a similar size because only such size matches have edges that are appropriately aligned. The bipole

connections cannot form between misaligned edges. With other orientations, it would be straightforward to generate other kinds of groupings for the circles in row 6. For example, a small circle could connect with a large circle at some orientation tangent, respectively, to each circle such that they would be joined by diagonal connections. Nevertheless, it is easy to find top-down connection parameters for horizontally-tuned cells that group stimuli in a way that matches the similarity of size principle.

Similarity of Shape

As discussed above, forming connections to reflect similarity by shape requires careful setting of the top-down parameters, at least for some stimuli. Even so, for the stimuli in rows 7–8, it is not difficult to find parameters that connect pairs of similar shapes but not different pairs.

Closure

In the model, elements can be grouped together if such grouping results in a “closed” set of contours. For some stimuli, such as that in row 10, this property is just a variation of the principle of proximity. Here, careful tuning of Spread Controller duration allows the ends of the parentheses to connect with each other in the traditional way (open on the left, closed on the right) rather than with their (slightly farther away) neighbor (e.g., closed on the left and open on the right).

Tuning of the Spread Controller does not fully account for the groupings in row 9, where each bracket has the same edge-to-edge separation with its neighboring brackets. However, the presence of a vertical edge on the outside of the bracket slightly alters the response of orientation filters at the top and bottom corners of each bracket. In particular, the response of a horizontally-tuned cell at the corner of a bracket is slightly weaker than the response of a horizontally-tuned cell at the end of a horizontal line. This difference in strength means that a connection is more readily formed between the inside ends of brackets compared to the outside corners of brackets. This difference can be exaggerated with the Long Controller Circuit, and this influence enables an observer to tune the Spread Controller duration and Long Controller input to find parameters that allow only the relatively stronger signals to connect.

Symmetry

Some experimental work suggests that particular symmetric elements are more likely to group together than asymmetric elements (Locher & Wagemans, 1993; Machilsen et al., 2009). Row 11 of Figure 30 shows an example stimulus from Palmer (1999; 2002). Here people often report that mirror symmetric lines group together into pairs, with the element on the far right not being grouped with anything. Symmetry is an inherent part of the bipole circuits (as indicated in the AND-gate exposition above). If the left and right sides of input to a horizontally oriented bipole cell are not of similar magnitude, the interneurons will inhibit the bipole cell and thereby not form a connection. On the other hand, the model operates on local information and so is only indirectly influenced by large-scale symmetry.

For this stimulus, the bipole cell activity in Figure 30 can connect the mirror symmetric lines into distinct pairs. These connections are formed by tuning the Spread Controller. The mirror symmetric lines have some parts that are closer than any non-symmetric lines. By setting the duration of the Spread Controller Circuit, it is possible to form short horizontal connections between symmetric neighbors but not between asymmetric neighbors.

Summary

In sum, the Connection Circuit provides a model of how an observer can use top-down control to group elements of a single stimulus in a multitude of ways via V2 connections. By tuning four connection parameters, two of which are timing parameters that promote the spread of V2 connections while the other two reduce this spread from edges that are relatively long or short, an observer can group particular stimulus elements. Additionally, as illustrated above, the connection parameters can be tuned such that elements of a given stimulus can form connections and, thereby, group in accord with a number of different Gestalt grouping principles, which are summarized in Figure 30.

Finally, I should highlight the fact that this demonstration is consistent with my rejection of the assumption that there is only one way to group elements of a given stimulus. Although it is important that the model is flexible enough to group the shapes in the table in accord with classic Gestalt grouping principles, it is not restricted to producing only these particular groupings via V2 connections.

TOP-DOWN CIRCUITS II: THE SELECTION CIRCUIT

Before presenting the Selection Circuit, it is useful to first develop the notion of a grouping strategy, which I am introducing in this project, and highlight its dependence on a particular task and stimulus set.

Generalizing from Francis et al. (2017), I propose that the overarching goal of an observer when performing some visual tasks, e.g., looking for or tracking a target, is to isolate the target in a segmentation layer, which allows for easy identification of the target. Exactly how this isolation is accomplished depends on the stimulus details and the task. Thus, an observer must develop a strategy that utilizes connection and selection processes in a way to quickly find and identify the target stimulus.

In terms of strategies, the main aim of a grouping strategy is to promote performance on a given task and stimulus set. Recall that I propose that a grouping strategy consists of a connection strategy and a selection strategy. As will become more evident in the simulations of experiments that study grouping, the connection strategy one chooses, i.e., how an observer tunes the four top-down input parameters to the Spread Controller, Long Controller and Short Controller Circuits to join the elements of a given stimulus, can depend on the particular task and stimulus set. For example, suppose the task is to find the location of a pair of adjacent I's in a row of I's and H's, e.g., Figure 31, first image from the left.

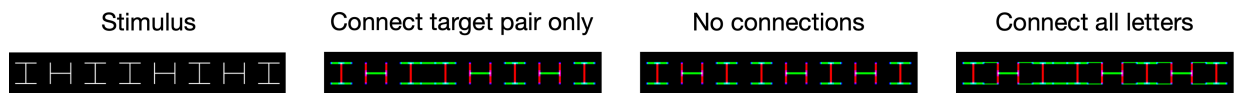


Figure 31. A toy example stimulus, i.e., a row of letters (left image), and task, i.e., identify the location of the pair of repeated I's, to illustrate connection and selection strategies in this chapter. The simulation images show three possible ways to connect the elements of this stimulus.

Suppose that on a given trial, the target pair could appear in any position, e.g., the target pair happens to be in the third and fourth positions from the left in the stimulus image in Figure 31. Given this stimulus and task, there seems to be only one connection strategy that would promote performance, i.e., tune the connection parameters such that only the nearby I's connect (Figure 31, second image). With a Spread Controller onset of 0 ms, this can be achieved by having

a positive Spread Controller duration (30 ms), turning off the Long Controller Circuit, and having a Short Controller input of 2.0 (which prevents spreading from the tops of H's). By connecting the I's, the observer may more easily select and segment out the target pair of letters.

In contrast, consider the alternative of connecting no letters (Figure 31, third image). If no letters are connected, then the selection process would have to do all the work in grouping, selecting and segmenting out the target pair, e.g., the observer would have to rely upon a selection signal that could overlap with both target letters. However, for this alternative strategy to work, the selection signal must be fairly large and/or placed quite precisely so that it overlaps with both target letters and no distractors. Additionally, unless it was known in advance where in the row the target pair will be located, it would be difficult for an observer to consistently place the selection signal so that it would both group and select the target pair only, on each trial. So, although this alternative is possible, it is not optimal given this task and stimulus set in which the target pair could appear at any location in the row. At the other extreme, an observer may connect all letters (Figure 31, fourth image). However, if all letters are connected, it is very difficult for a selection signal to segment out only the target pair since a selection signal segments out any boundary signal it falls on and any boundary signal that is connected to the boundary signal it falls on. Thus, it seems that the connection strategy in which only the target pair is connected will best promote performance on this task.

So, suppose the observer chooses to use the connection strategy in which only the target pair of letters group via V2 connections. How might an observer use this connection strategy to more easily find the target pair? I propose that the observer uses the connection strategy in tandem with a selection strategy to quickly find target elements.

In broad terms, the Selection Circuit models the selection and segmentation of stimulus elements into different layers with the aim of segmenting target items from distractors efficiently. The main components of this circuit, i.e., segmentation layers, selection signals, a reset, and a feature filter, will be presented in detail in the next section. For the present purpose of providing an overview of the task-dependent strategies investigated by the model, a selection strategy involves a strategy for determining the size, location, number, and timing of selection signals. The model's Selection Circuit allows for top-down control of these features and, thus, allows us to simulate possible selection strategies that an observer may be using given a particular task and stimulus set. For example, in the above toy example task of finding a pair of adjacent I's that are

among H's in a row, suppose that the observer uses the connection strategy that connects only the target pair of I's. Given this connection strategy, the target pair will tend to have a substantial horizontal signal since the V2 connections between them are horizontal. In turn, a possible selection strategy that would allow the observer to quickly select and segment the target pair is to place the selection signal on areas where more horizontal signal is detected and to choose a size for the selection signal that, say, will not cover more than two letters in the stimulus. Thus, an observer may use a connection strategy in tandem with a selection strategy to promote performance on a given task and stimulus. So, the next step in model development is to create a Selection Circuit subject to top-down control that can generate such selection strategies.

The Selection Circuit

The Selection Circuit is presented in somewhat general terms in the sense that, although this project focuses on visual search tasks, the circuit's main components may be used to simulate other visual tasks, e.g., multiple object tracking tasks. The toy example stimulus introduced in Figure 31 and a simple visual search task, i.e., find the location of the adjacent repeated I's, are used in this section to demonstrate some of this circuit's features. So, although this circuit has the potential to simulate a variety of visual tasks, the example used below is a visual search task due to the focus of this project.

The segmentation process in Francis et al. (2017) could perform two sorts of segmentation: boundary segmentation and surface segmentation. The present model only has boundary segmentation because surface segmentation is not required in the simulations conducted for this project.

The Selection Circuit as developed in this project (Figure 32) consists of five main components: segmentation layers, selection signals, a Boundary Segmentation Circuit, a reset, and a feature filter. Although the first three of these components are largely the same as those used in Francis et al. (2017), it is worth explaining how they function since they play a central role in the Selection Circuit.

The structure of the Boundary Segmentation Circuit, which is given in Figure 32, has been simplified compared to that of Francis et al. (2017), but its function is generally the same: to shift selected boundaries from the base segmentation layer to a different segmentation layer, which was described in the second chapter.

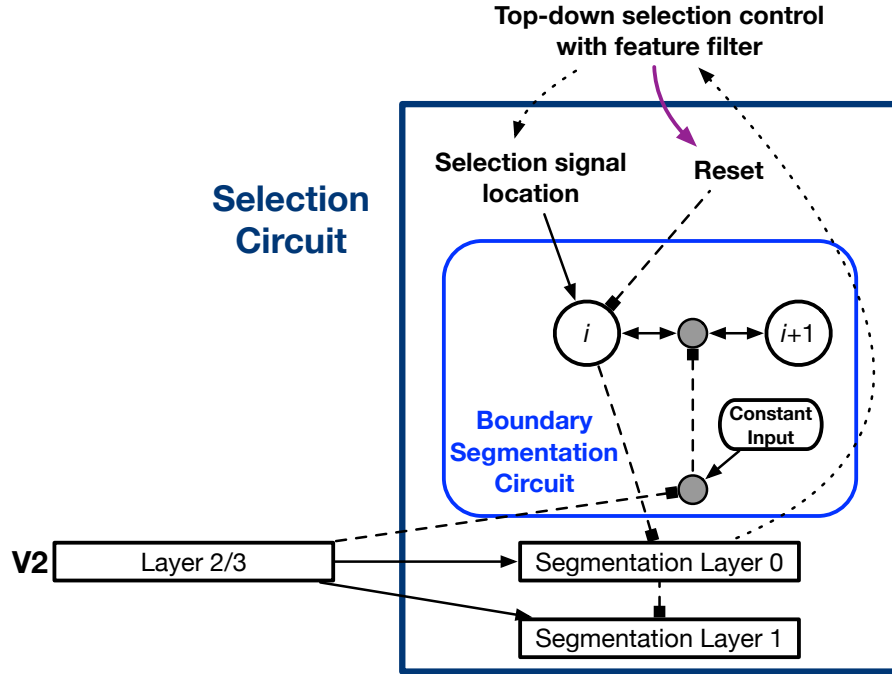


Figure 32. A schematization of the Selection Circuit. Again, excitatory synapses are represented by arrows with solid shafts, while an inhibitory connection is represented by a dashed line capped with a square. Arrows with dotted shafts represent the flow of information, and the purple arrow to Reset represents a timed, positive current input to a reset cell. The bottom gray interneuron receives constant excitatory input, which can be inhibited when there is activity at the corresponding position in V2, layer 2/3. In turn, the bottom interneuron prevents runaway spread of the selection signal. The Boundary Segmentation Circuit box shows the circuit created for the V2 bipole cell that corresponds to position i . This circuit involves the boundary segmentation cell (labelled ' $i+1$ ') for its neighboring cell that corresponds to position $i+1$.

In the models of both Francis et al. (2017) and this project, the bipole cell activity in V2, layer 2/3, feeds directly into one or more segmentation layers. Selection signals are applied to one layer, i.e., segmentation layer 0, which will also be referred to as ‘the base segmentation layer’, via the Boundary Segmentation Circuit. There is one such circuit for each non-base segmentation layer. This circuit, shown in Figure 32, operates as follows. If there is no selection signal present, strong inhibition to each pixel in segmentation layer 1 from the corresponding pixel in the base segmentation layer prevents any activity in segmentation layer 1.

Next, consider a segmentation signal placed in the base segmentation layer. This is achieved by creating a map that has the same dimensions as V2, layer 2/3, with the location of the selection signal having a constant non-zero value. This map will be referred to as a ‘selection signal map’. In Figure 32, this selection signal map is labelled ‘Selection signal location’, and Figure 33 provides an example of this map, consisting of a white circle. (It is shown overlapped

with the stimulus image to give a sense of the signal's location relative to the stimulus elements.) If location i in this map has a non-zero value, then the boundary segmentation cell labelled i in Figure 32 receives constant excitatory input. In this case, cell i will eventually inhibit the activity of the cell that corresponds to position i in segmentation layer 0, which, in turn, reduces the inhibition from this segmentation layer 0 cell to the segmentation layer 1 cell that corresponds to position i . This allows the signal from V2, layer 2/3, at this position to effectively shift from segmentation layer 0 to segmentation layer 1. In effect, a selection signal transfers activity from the base layer to layer 1 by sending excitatory signals to all cells in the locations it 'covers', which is achieved by sending excitatory input to each boundary segmentation cell at those positions. Figure 32 shows the components of the Boundary Segmentation Circuit for one such cell labelled by its position i .

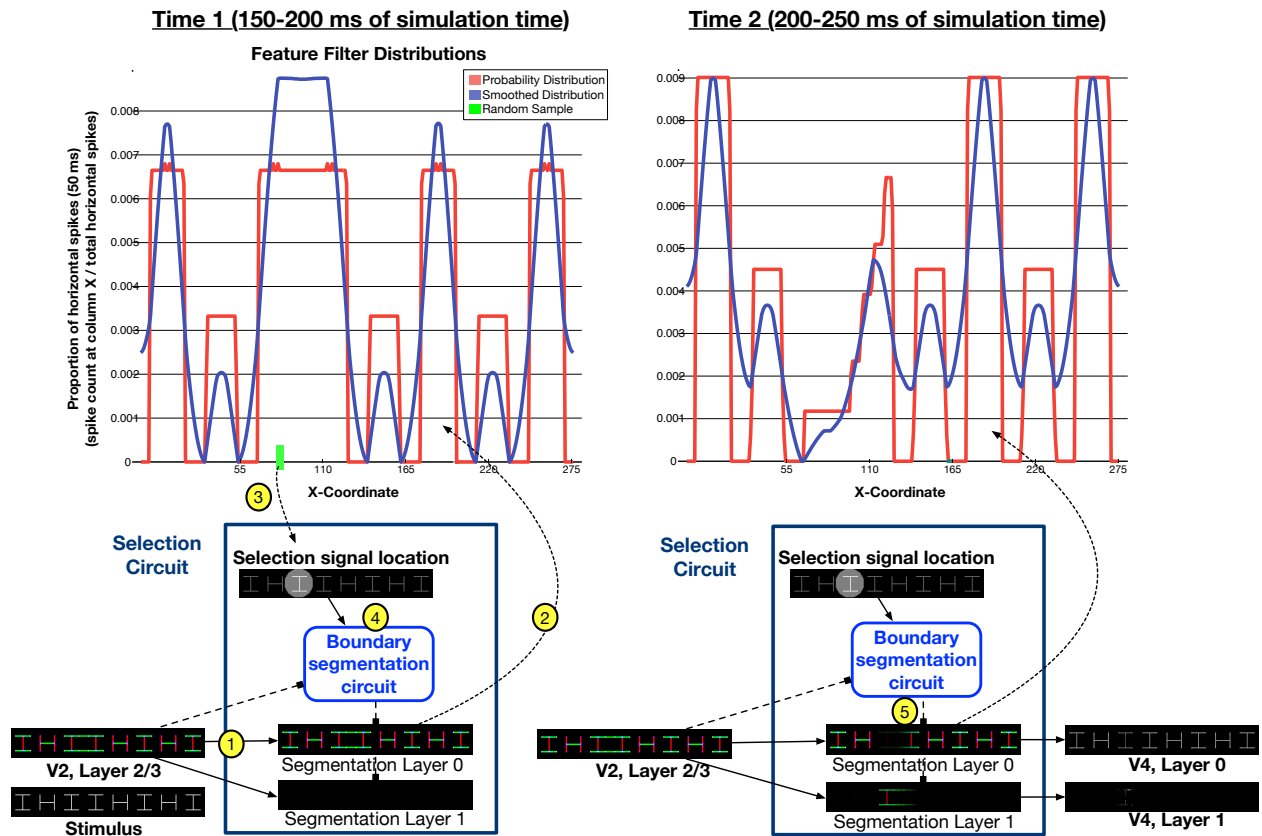


Figure 33. The activity, feature filters and selection signal map of two successive 50 ms timesteps in a simulated trial.

Additionally, edges that are not covered by a selection signal but connected to an edge that the selection signal covers are also eventually pushed into segmentation layer 1. To see how the selection signal can spread along boundaries that are not ‘covered by a selection signal’, e.g., to locations with boundary signal that are adjacent to a location in the selection signal map with a non-zero value, suppose there is a horizontal edge at locations i , and $i+1$. So, the V2, layer 2/3, horizontally-tuned bipole cells are active at the corresponding retinotopic locations. Further suppose that the selection signal covers location i , but not $i+1$. The selection signal does not cover $i+1$ and, thus, the boundary segmentation cell labelled $i+1$ does not receive the same constant excitatory input directly from the selection signal map that cell i receives (i.e., there is no excitatory input from ‘Selection signal location’ to the boundary segmentation cell labeled $i+1$ in Figure 32). However, the selection signal input to boundary segmentation cell i will spread to boundary segmentation cell $i+1$ via the bidirectional arrows between these neighboring boundary segmentation cells to the gray interneuron: over time, activity from a neighboring location, e.g., cell i , that is covered by a selection signal can spread to, e.g., cell $i+1$ via the interneuron.

To see how this spread of selection signal occurs, refer to Figure 32. In the example above, recall that an edge is detected at positions i and $i+1$. So, the bipole cells in V2, layer 2/3, that correspond to these locations are active. In turn, the bipole cell corresponding to position i , will inhibit the bottom interneuron, which also receives constant excitatory input by default. In turn, the inhibition of the top interneuron is reduced and, thus, selection signal can spread from boundary segmentation cell i to its neighbor, i.e., to boundary segmentation cell $i+1$. If this activation is strong enough, the boundary segmentation cell $i+1$ will behave as if there is a selection signal at that location. More specifically, boundary segmentation cell $i+1$ can: (a) functionally move signal at the corresponding location in segmentation layer 0 to segmentation layer 1 by inhibiting activity in segmentation layer 0 at a particular location, and (b) cause selection signal to spread to its neighboring boundary segmentation cell corresponding to location $i+2$ via the bidirectional synapses and top interneuron. Thereby, the Boundary Segmentation Circuit allows the activation from the selection signal at a particular location to spread from active cells at that location to neighboring active cells. Additionally, due to the architecture of the Selection Circuit, this spread of selection signal can result in the selection and segmentation of, e.g., connected lines (Figure 3) and entire lines (Figure 4), rather than only the edge(s) that is covered by the selection signal.

Thus, the Boundary Segmentation Circuit is designed to promote spread of the activation from a localized selection signal along boundaries that are connected to boundaries that the signal covers.

While only a single selection signal and associated non-base layer are presented in Figure 32, it is possible to have multiple selection signals and corresponding segmentation layers. Indeed, the simulations in both Francis et al. (2017) and reported below use multiple selection signals. Following Francis et al. (2017), the structure shown in Figure 32 is extended as follows to implement multiple selection signals. For each non-base segmentation layer, a separate Boundary Segmentation Circuit is created. And, each additional segmentation layer is inhibited by all previous layers, e.g., segmentation layer 2 would be inhibited by both layer 0 and layer 1. The selection signal associated with segmentation layer 2 would pull boundary signals it covers and any activity connected to it from layer 0 to layer 2. (Examples of resulting segmentation layer output when the model implements multiple selection signals are provided below in applications of the model to various experiments as well as in Francis et al., 2017.)

Several modifications were required to allow the model to simulate different visual tasks, e.g., visual search tasks, in which an observer typically scans a stimulus to find or track a target (or targets). Notably, the circuit has the ability, through top-down control, to make the selection signals more dynamic. In other words, this circuit can model a shift in the observer's attention from one area of a stimulus to another. It does so by allowing for different selection signal maps over time. For example, in a visual search task where the task is to find a target as quickly and as accurately as possible, the selection signal map changes after every 150 ms of model time, which shifts the selection signal to a different location every 150 ms. This time was chosen because it seems the shortest period of time that can produce a clear enough signal in model V4 for a target identification algorithm to determine whether the segmented boundaries are from a target. If the selection signal is at the position for a shorter time, the V4 signal is often too weak to determine whether a target has been detected. And, since these visual search tasks typically emphasize speed and accuracy and measure reaction time, it seems inefficient to choose a longer time before shifting a selection signal if the target was not selected. I refer to an interval in which a selection signal is at a particular location as a 'selection cycle'. For example, the 150 ms interval used in the preceding example is referred to as a '150 ms selection cycle' since the selection signal map is the same, i.e., the selection signal remains in the same place, for 150 ms.

This shifting selection signal gives rise to two issues. First, a dynamic selection signal has to be moved to other areas of the stimulus. How is it determined where the selection signal will move if it did not isolate the target in a non-base layer? Second, if a selection signal changes location, the boundaries that it selected, which often comprise shapes with closed edges, can cause positive feedback loops to occur in the non-base segmentation layer. In other words, a selected object will stay selected and remain in the non-base segmentation layer even if the selection signal has moved and no longer covers any part of the object (e.g., the bottom row of Figure 36). In a task where, say, one is to count target items in a stimulus image where there are no distractors, this buildup of selected elements may be advantageous: eventually the shifting selection signal could segment out all target elements. However, in a visual search task, for example, where the task is to find a target item, such a buildup of stimulus elements could be detrimental to performance. If, say, the initial selection signal overlapped with a distractor but subsequently shifted to the target, then the selected distractor would be segmented along with the target and, thus, would prevent the observer from easily segmenting the target from the distractor.

To address these two issues and, in turn, allow the model to simulate a wider variety of visual tasks, I have incorporated feature filters and a reset signal into the Selection Circuit. A third factor, namely selection signal size, is also relevant to selection strategy; as will be discussed below, larger selection signals allow for faster segmentation yet are more likely to overlap with distractors, while smaller selection signals offer more precision yet tend to take more time. In the model, these three components (i.e., choice of feature filter, timing of reset, and choice of selection signal size) are subject to top-down control, which allows for more options in choosing a selection strategy that is suitable for the visual task at hand. To again emphasize the flexibility of this circuit (although the examples below are visual search tasks since that is the focus of this project), the addition of these features allows the Selection Circuit the potential to simulate a number of different selection strategies that are likely used in a variety of visual tasks.

Feature Filters

To guide the placement of a selection signal during each trial, I implemented an algorithm that functions as a feature filter. Something like a feature filter is an integral part of many theories of visual perception (e.g., Itti & Koch, 2001; Wolfe, 1994), where it is used to guide attention and can be used to classify scenes or elements.

Figure 33 shows how feature filters are integrated within the Boundary Selection Circuit and segmentation layers introduced above.

To better organize my presentation of this circuit, numbered yellow circles are provided in Figure 33. These indicate (somewhat) sequential steps that occur in this process. This figure has two columns labelled ‘Time 1’ and ‘Time 2’, which show model output during two consecutive time steps of a simulation using the toy stimulus and implementing the connection strategy where only the pair of adjacent I’s connect. In this section, I go through each step, explaining relevant components of the model, and subsequently show how the feature filters can be used to strategically place selection signals so as to promote performance on our example task.

At Time 1 (Figure 33, left column), during Step 1 the activity of V2 bipole cells feeds into the Selection Circuit, as described above. At Step 2 information from the base segmentation layer is used to create a probability distribution represented by the red curve in the plot above the circuit diagram. To see how this distribution is generated, consider Figure 34a.

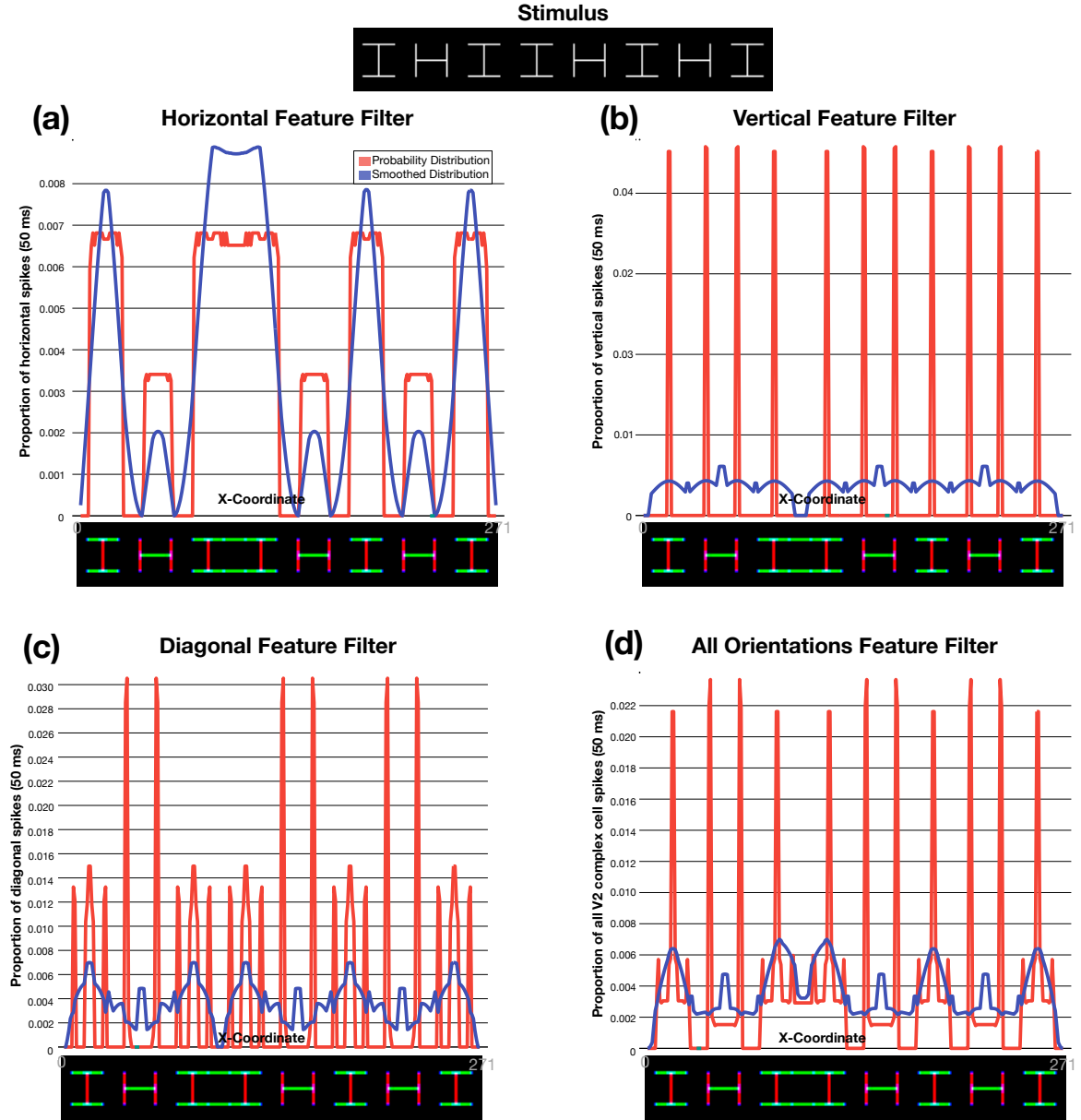


Figure 34. Examples of different feature filters given the stimulus image at the top and the connection strategy where only adjacent letter I's connect.

The image in Figure 34a below the plot shows activity in segmentation layer 0 over 50 ms. Here it is quite similar to V2 bipole activity. Along the x-axis are the x-coordinates of the stimulus. In this case the stimulus is 271 pixels wide, and so there are 271 ticks along the x-axis. The plot has two distributions. The red distribution is created as follows. The height of each point represents the ratio

$$\frac{V_c}{\sum V_c},$$

where V_c is sum of the activity of horizontally-tuned bipole cells in column c during 50 ms, and $\sum V_c$ is the sum of this activity across all columns, i.e., the sum of horizontal activity across the entire image during 50 ms. For each column in the image, the value of this ratio is plotted above the coordinate on the x-axis that corresponds to the column. Thus, the red curve in Figure 34a depicts the ratio of horizontal activity in each column to the horizontal activity in the entire image during 50 ms. As is obvious in this plot, the highest peaks are above columns with a lot of horizontal signal, i.e., at columns where the letter I's are located. These peaks are roughly twice as high as those over the horizontal H bars because letter I's have twice much horizontal activity at each column, i.e., they have two horizontal bars rather than one. Since it only uses activity of horizontally-tuned V2 cells, I call it a 'horizontal feature filter'. Other feature filters are possible. For example, Figure 34b shows a vertical feature where the red curve was generated using the same calculation described above except activity from vertically-tuned cells, rather than horizontal, was used. Hence, it is a 'vertical feature filter'. Here each vertical bar in the output results in a very high peak in the red curve. Figure 34c shows the diagonal feature filter for this image, and Figure 34d shows the distribution generated if activity from all cells is used. Other feature filters are possible, e.g., a horizontal-vertical feature filter that uses activity from horizontally-tuned and vertically-tuned cells.

In each plot shown in Figure 34, the blue curve was generated by smoothing the red curve using a 1D Gaussian filter. The blue curve plays a key role in Step 3 shown in Figure 33. An x-coordinate is produced by generating a random sample from the distribution represented by the blue curve (the Python library numpy has a function, `random.choice()`, that performs this kind of sampling). This x-coordinate is used as the x-coordinate for the center of the selection signal for the next 50 millisecond time step. (In this case where the stimulus elements are along a single row, the y-coordinate was fixed to half of the height of the stimulus for simplicity.) In the example depicted in Figure 33, this x-coordinate is indicated by the green rectangle at the bottom of the plot. Here the randomly sampled x-coordinate happened to be at the left target I. An appropriate selection signal map is created, shown under Selection Signal Location in Figure 33 and overlaid on the stimulus image to make clearer over what region the selection signal falls.

The selection signal map is then input at Step 4 to the Boundary Segmentation Circuit, which was described earlier. The effects of this circuit can be seen at Time 2 (Figure 33, right column) at Step 5 where the pair of I's is starting to become segmented into layer 1. Further, the edges of the target pair have just begun to move to V4 layer 1. Additionally, to highlight the dynamic nature of the model, the probability distributions at Time 2 are shown even though, as explained above, the selection signal will stay at the same location until the next selection cycle begins at Time 5 (i.e., 150 ms after the first selection cycle, which began at Time 2), the location of which will be sampled from the probability distribution at Time 4. It's also important to highlight the change in the feature filter distribution at Time 2. Since the distributions are created using activity in segmentation layer 0 only, the distribution changes when a selection signal is applied to this layer that pulls activity down into a non-base segmentation layer.

How do these feature filters allow the model to strategically guide the placement of selection signals so as to promote performance on a given task? Consider the example task of identifying the location of the adjacent pair of I's given the toy stimulus. As discussed above, a good connection strategy is to connect only the target pair of I's. That way, if one of the I's is selected, then both will be segmented, isolated in a layer and, thus, easier to identify. If a horizontal feature filter is used in conjunction with this connection strategy, then the model will tend to guide selection signals to the location of the target pair. As shown in Figure 34a, the horizontal feature filter results in a distribution with a relatively high and wide peak at the location of the target pair. Since the connected target pair has a lot of horizontal signal, it is likely that the horizontal feature filter will guide the selection signal to this location, i.e., it is more likely that the sampled x-coordinate will be at its location. In turn, the use of a horizontal feature filter seems to be a good choice for performing this task efficiently; it is more likely that the resulting selection signal will fall on this location, which will tend to result in the observer finding the target pair faster. Thus, this horizontal signal filter models how an observer may guide attention to areas with more horizontal signal so as to quickly find the target and perform well on this task.

Of course, since the feature filter uses a random sample to determine selection signal location, the observer may not always immediately find the target pair. For example, consider the simulation shown in Figure 35.

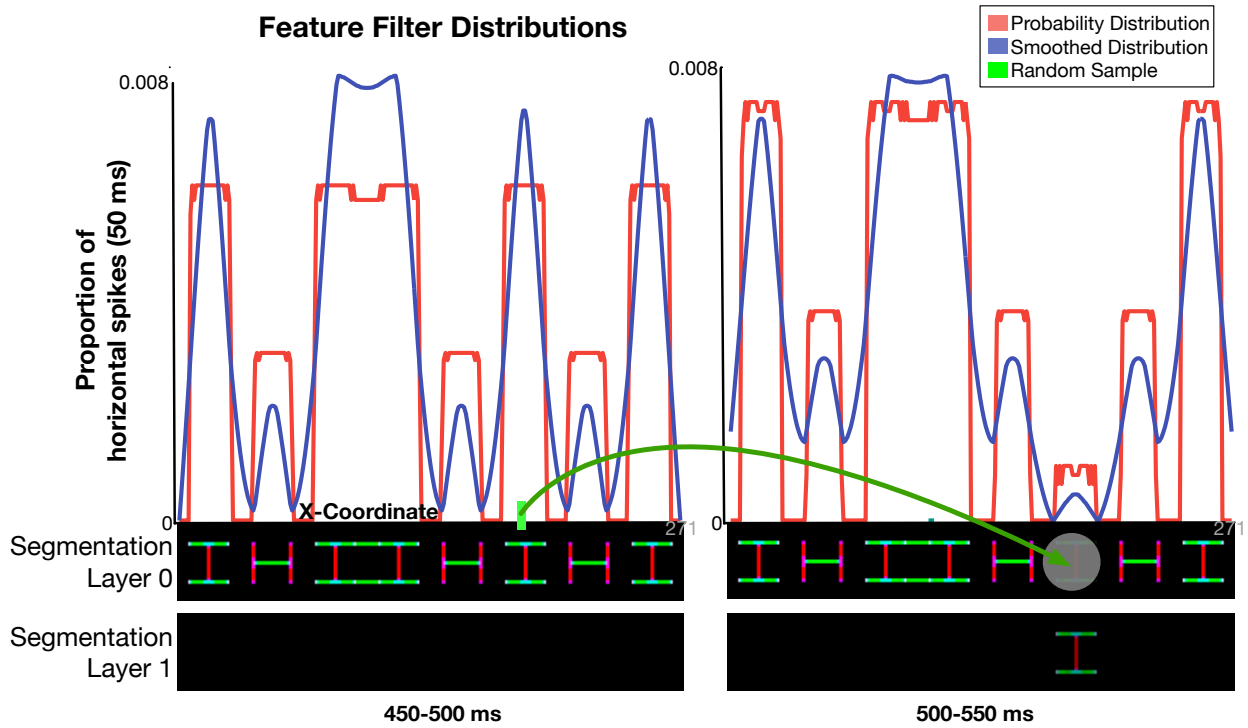


Figure 35. An example of a horizontal feature filter given the toy example discussed at the beginning of this chapter where the connection parameters are tuned such that only the target pair of nearby letter I's connected. Left column shows the layer 0 activity and resulting feature filter distribution from which the selection signal's x-coordinate at the next time step was sampled. Right column shows the resulting location of the selection signal at the subsequent time step, which is superimposed on the layer 0 activity to show its position relative to layer 0 activity. The green arrow indicates that the sampled x-coordinate is used as the x-coordinate of the selection signal center on the subsequent time step.

At this point in the simulation, a coordinate above a distractor letter I was randomly selected from the horizontal feature filter distribution. This is to be expected since this distribution also has relatively high peaks above distractor I's given this image and connection strategy. Here this resulted in the selection of a distractor, rather than the target pair. Since the model (like human observers) does not always immediately select and find the target in a visual search task, it was necessary to add a reset signal that allows search to proceed.

The Reset Signal

The reset signal, labelled 'Reset' in Figure 32, is inhibitory input to the Boundary Selection Circuit. Its function is to provide the observer with the option of periodically, i.e., at the start of each selection cycle, resetting each non-base segmentation layer (e.g., Figure 36, top row).

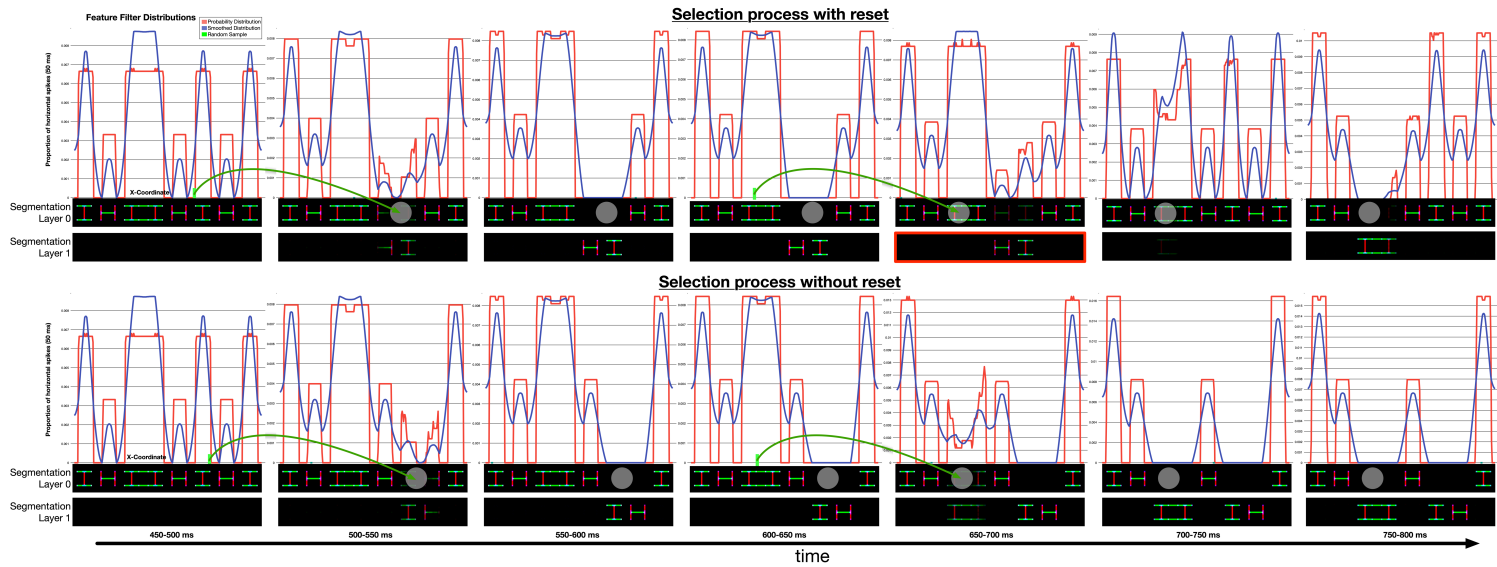


Figure 36. Examples of how the selection process operates over time. In both examples the feature filter uses horizontal signal. Top row: An example of the selection process with the reset of segmentation layer 1, which occurs at the beginning of the 50 millisecond time step indicated by a red box. Bottom row: An example of the selection process without a reset. Even though the selection signal moves to a different location at 650 milliseconds, the pair of letters selected earlier remain in layer 1.

Such a reset is needed to ‘clear’ the non-base segmentation layer; since the reset signal eliminates the activity of all boundary selection cells, they can no longer inhibit segmentation layer 0, which, in turn, returns to strongly inhibiting the non-base segmentation layer. Without such a reset, items that have been selected and segmented by selection signals in other locations remain in non-base segmentation layers (e.g., Figure 36, bottom row).

For an example of a task in which this reset would be part of a selection strategy that promotes performance, consider a visual search task where both speed and accuracy are emphasized. For such a task, it would be beneficial for performance to use the quickest possible reset that completely clears the non-base segmentation layers so that no items previously selected yet not identified as targets would remain segmented from the base segmentation layer. For all of the visual search tasks simulated in this project, it turned out that a reset of 5 ms at the beginning of each 150 ms selection cycle was the shortest amount of time in which the contents of a non-base segmentation layer could be cleared and a shifted selection signal could segment an element to V2 and still produce enough V4 signal for a target identification algorithm to determine whether a target was selected. With this setup only elements covered by (and connected to elements covered by) the selection signal during a selection cycle are shifted to a non-base segmentation layer; elements that the selection signal segmented during previous selection cycles are cleared from the non-base segmentation layer by this periodic reset such that accuracy and speed are both maintained and, thus, likely subject to some sort of top-down control.

Selection Signal Size

I propose that selection signal size is another top-down controlled component of an observer’s task-dependent and stimulus-set-dependent selection strategy. To see the costs and benefits of different selection signal sizes, consider the running toy example of a row of letters in which the task is to identify the location of the pair of adjacent letter I’s. Three simulations were run, each of which had a selection signal of a different size centered at the same location, as shown in Figure 37.

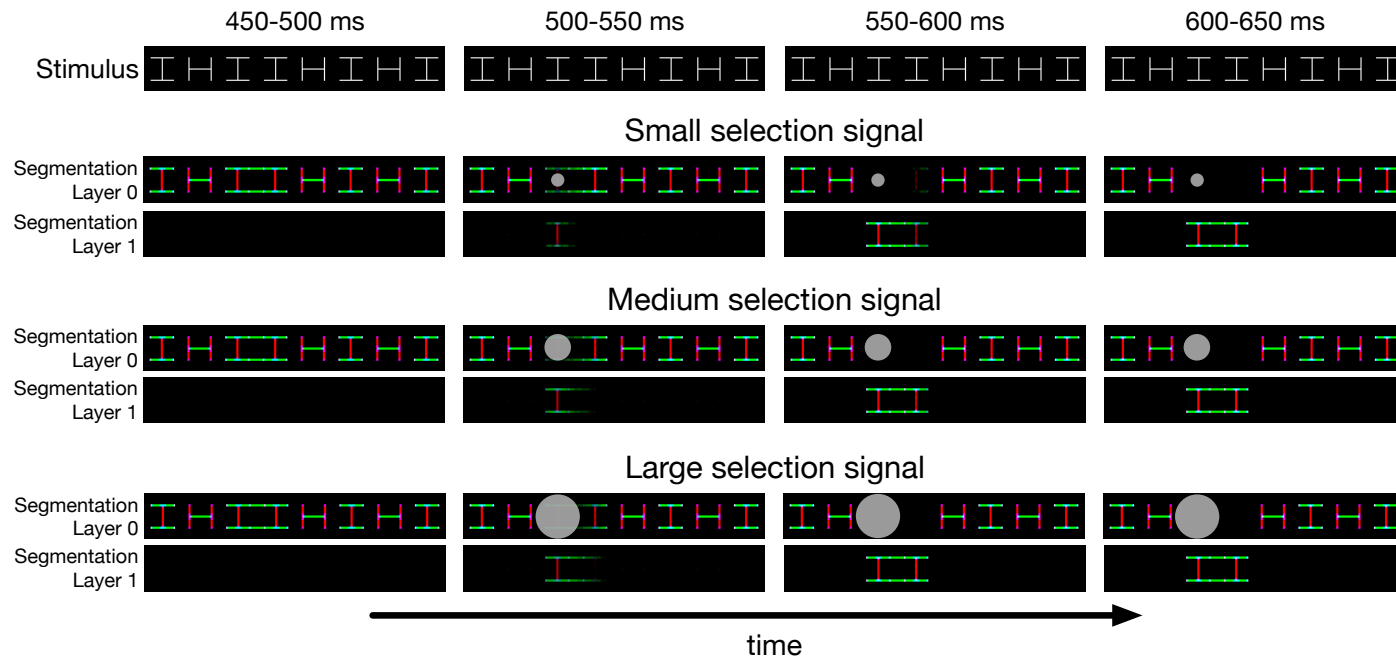


Figure 37. An illustration of how selection signal size impacts the speed of segmentation through a comparison of the segmentation layers of three simulations, each of which only differed in the size of selection signal used, which are indicated by the white circles. Each selection signal was centered at the same location (on the left target I). The smallest selection signal has a diameter that is half the width of a letter. The medium selection signal diameter is the width of a letter. And, the large selection signal diameter is twice the width of a letter. A comparison of segmentation layer 1 at 500-550 ms for each simulation shows that the segmentation process is faster, i.e., there is more boundary signal in layer 1 for the simulation with the large selection signal, than for the other simulations.

An advantage of using a smaller selection signal is better precision. Consider the example of finding a pair of nearby letter I's with the connection strategy of only connecting the target pair and a selection strategy that involves a horizontal feature filter. The smallest selection signal is unlikely to overlap with a neighboring distractor. In turn, it is highly likely that a small selection signal that lands on a target letter will segment out the target pair only, which will make it easier for the observer to identify the location of target pair. The largest selection signal, on the other hand, has a greater chance of landing on a target letter yet also overlapping with a distractor. For example, the large selection signal in the simulation above is quite close to the distractor even though it is centered on a target. So, if this large selection signal landed slightly off center of this letter I and to the left, this selection signal would likely overlap with a distractor. In such cases, a target pair and a distractor would be selected and segmented out, making it difficult to identify the location target pair. So, smaller selection signals are more precise in that it is less likely for them to overlap with distractors.

However, there is also a cost with using small selection signals. As shown in Figure 37, it takes an additional time step, i.e., 50 milliseconds, for the smallest selection signal to segment out the target pair from the distractors compared to the medium and large selection signals; at 550-600 ms, the boundary signal corresponding to the right letter I is not completely in layer 1 for only the simulation with the small selection signal. This is because the small selection signal only covers a small portion of the connected target pair and, in turn, it takes more time for the selection signal to spread across edges that are connected to edges covered by the small selection signal. Thus, smaller selection signals can be more precisely placed yet can take more time to segment out a target, while larger selection signals take less time to segment out a target yet are more likely to overlap with a distractor. In turn, a choice of selection signal size involves something like a speed-accuracy tradeoff but at the segmentation level: larger selection signals can result in faster segmentation yet are more likely to segment a non-target item, while smaller selection signals can result in slower segmentation yet are more precise and, thus, more likely to segment only targets.

Application to Other Stimulus Images

Thus far we have only considered stimuli in which elements are in a row, but the selection process can be generalized to elements in a grid (which is similar to stimuli used in the Trick & Enns, 1997, experiment simulated below). For elements in a row, the feature filter distribution

presented above provides a way to determine the x-coordinate of the selection signal's center. For stimuli that involve elements in a grid, the simplifying assumption that the y-coordinate of the selection signal is fixed to, e.g., half the height of the stimulus image, may not provide a reasonable selection strategy. For these cases, the x- and y-coordinates can be randomly sampled from a 2D distribution that is smoothed using a gaussian kernel, e.g., Figure 38.

Similar to the example of the selection process for a row of shapes, for a stimulus in a grid a particular pixel location is randomly sampled from these smooth distributions (the sampling process was done with the `random.choice()` function, which was also used for the 1D distribution, after I converted the probability assigned to each coordinate into a 1D array). The x- and y-coordinates of this pixel location determine the location of the center of the selection signal for the next time step.

Summary

In sum, according to the model, a selection strategy involves determining placement and timing of the selection signal. Selection signal placement is broken down in the model in terms of: choice of a feature filter, which determines where a selection is more likely to be placed, and choice of selection signal size, which involves a segmentation speed-precision tradeoff. Timing of the selection signal involves when to start selection signals, the duration of selection cycles, and when to use the segmentation reset signal. Just as for a connection strategy, what selection strategy an observer uses depends on the given task and stimulus set. Further, a selection strategy is chosen in conjunction with a connection strategy, e.g., in the above two examples, a horizontal feature filter was chosen partly because horizontal connections united target pairs due to the chosen connection strategy. Within the framework of the model, a connection strategy and selection strategy, which are chosen in tandem so as to promote performance on a given task and stimulus set, comprises a grouping strategy.

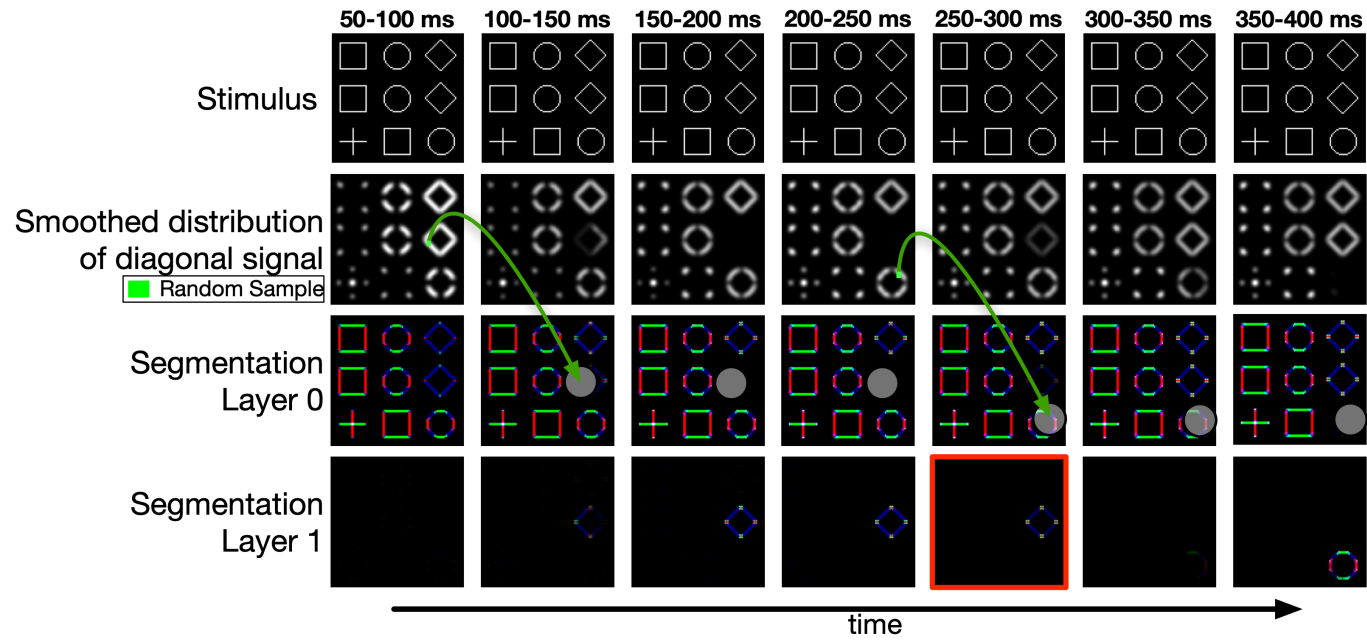


Figure 38. An example of the selection process for a matrix of shapes where a smoothed probability distribution is constructed from V2 signal. Here the feature filter used diagonal signal. Thereby, areas with diagonal signal are more likely to be the center coordinate of the selection signal. As in the previous figure, the red box indicates a reset of layer 1.

This modified Selection Circuit, with top-down controlled feature filters and reset, allows the model to implement a wide variety of selection strategies and, thus, gives it the ability to simulate many kinds of visual tasks compared to the model of Francis et al. (2017). For example, it is now possible to simulate visual search tasks in which observers likely use a dynamic and strategic segmentation process in which each selection signal tends to go to locations where a stimulus has particular feature(s). In the simulations below and, foreseeably, simulations of other visual search tasks where one must respond as quickly and accurately as possible, if the target pair is not segmented out after a 150 millisecond selection cycle (which was chosen to emphasize speed and accuracy), the selection signal shifts to another location and the segmentation process repeats.

APPLICATION OF THE MODEL TO THREE EXPERIMENTS THAT USE VISUAL SEARCH TASKS TO INVESTIGATE GROUPING

In the previous two chapters, I presented model modifications in detail. Modifications to the Connection Circuit give the model the flexibility to group stimulus elements by V2 connections in a variety of ways. The addition of a dynamic selection circuit that can use features to guide selection signals to areas with contours of particular orientations enable the model to simulate visual search tasks. Further, I developed the concept of a grouping strategy in terms of the mechanisms of the model, i.e., a grouping strategy is a combination of a connection strategy and a selection strategy that promotes performance on a given task. In the next three chapters, I apply the model to three visual search experiments designed to investigate grouping. For each experiment, I conducted a replication experiment and several sets of simulations in which the model used particular grouping strategies. In some cases, I conducted additional experiments to test model predictions.

General Experiment Methodology

Three experiments designed to investigate grouping with a visual search task were replicated, i.e., experiment 1 of Palmer and Beck (2007), experiment 2a of Vickery (2008), and experiment 2 of Trick and Enns (1997). I conducted two additional experiments to test model predictions. Methodological procedures common to all experiments are as follows, and additional details are indicated under the Methods section of the relevant experiment.

Participants

To identify an appropriate sample size for each experiment that I compared with simulation results (i.e., Experiments 1, 5, and 6), I first determined that I generally wanted to measure each mean response time with a precision that would have a standard error of 10 milliseconds. Response time standard deviations across observers for identification tend to be around 100 milliseconds, so I planned for a sample size of around 100 observers because this would give a standard error of $100/\sqrt{100} = 10$. For other experiments, justification for sample size is given below. All participants were naïve undergraduates from Purdue University who took part in exchange for

course credit. All participants provided informed consent in accordance with Purdue's Institutional Review Board.

Apparatus

All experiments were conducted online. I programmed each experiment in custom JavaScript and HTML scripts and uploaded them to a local server. Participants used a computer to navigate to the webpage and take an experiment, and those who attempted to use a tablet or phone were prompted to switch to a laptop or desktop computer. Participants used keys on their keyboard to register responses. Specific keys used are indicated under the Methods section for each experiment below.

Stimuli

All stimuli were generated by the JavaScript program. Since the study was conducted online, a participant's distance from the monitor and the monitor's size are unknown, and, thus, the visual angle subtended by the stimuli is also unknown. To provide some sense of the size of stimulus elements, below I provide the degrees of visual angle that stimulus elements would subtend for a hypothetical participant who used a laptop on a desk with the monitor 18 inches away. The hypothetical participant's laptop has a 13.3-inch (diagonal) monitor with a width of around 11.25 inches and resolution of 2560x1600 (making pixel density 227 pixels per inch).

Procedure

In all experiments, participants navigated to a webpage. After providing consent, participants began by reading instructions at the top of a webpage and subsequently scrolled down to a black rectangle where the experiment took place. The instructions encouraged participants to enlarge their browser, if needed, so they can view the entire black rectangle. The rest of the procedure varied by the task and is detailed below for each experiment.

Simulation Methodology

All simulations were programmed using Python2 scripts and used the package NEST 2.14.0 (Peyser et al., 2017) for creating the cells and synapses and managing network dynamics.

A single cell type (`iaf_psc_alpha`, which is a leaky integrate-and-fire neuron model with alpha-function shaped synaptic currents) and synapse type (`static_synapse`) were used, but synapse weights were manually set. For the simulations conducted, each trial takes approximately 12 – 40 minutes, depending on the size of the stimulus and on the time at which a trial was terminated due to a target or set of targets being found.

Stimuli for each simulation were made using custom Python3 scripts with standard packages (`numpy`, `random`). For the Palmer and Beck and Vickery simulations, the stimuli were written to BMP files using the package `ImageIO` (Klein et al., 2018). For the Trick and Enns simulations, on each trial a stimulus image was generated in the main LAMINART code to better match the stimuli used in the experiment, where the location (and added jitter) of each stimulus element was randomly generated on each trial.

To reduce overall runtime, the simulations were run on several systems in parallel: a 2019 MacBook Pro (8 cores, 32 GB RAM), a 2018 Linux running Ubuntu (6 cores, 16 GB RAM), two virtual machines with each running Ubuntu 20.04.2 (44 cores, 120 GB RAM) on XSEDE Jetstream (Stewart et al., 2015; Towns et al., 2014), and a local computer cluster that runs Linux with Debian (24 nodes with 20 cores and 64 GB RAM per node). Batches of trials for particular simulations were run on the cluster (Trick & Enns) and virtual machines (Vickery), which was merely due to having access to these systems at the time. Checks of the different computing systems indicate that they give the same results.

Each stimulus was presented until the trial was terminated. Every 50 ms following onset of the selection signals, the program summed the neural action potentials in V4 in the non-base segmentation layer(s). This activity was used to determine whether a target was detected using a target identification algorithm, which is described below for each simulated experiment. If a target was detected in the Palmer and Beck and Vickery simulations, the trial terminated and the simulated time until target detection, which corresponds to reaction time, was recorded. For each target detected in the Trick and Enns simulation, a counter was increased by one. If this counter remained the same after 250 ms, then the trial terminated and the value of the counter, which corresponds to the number of targets enumerated, and simulation time, which corresponds to reaction time, were recorded. Otherwise, the simulation continued for another 50 ms increment. For the Palmer and Beck and Vickery simulations, a trial terminated when either a target was found or a maximum time was reached.

The results of model simulations will be qualitatively compared with the empirical results for human subjects who performed the same task. Because the model simulations occur over time, the model time for each condition can be averaged across simulated trials and can be directly compared with the averaged reaction time of human subjects measured in each of the target experiments. Similarly, accuracy can be measured in terms of whether elements segmented out from the base segmentation layer are identified as targets or not, or, for the Trick and Enns simulation, whether the final value of the counter on each trial matches the number of targets in the stimulus.

PALMER AND BECK (2007)

Palmer and Beck (2007) developed a visual search task meant to provide an “objective” measure of grouping. Each stimulus in their experiment 1 consisted of a row of nine shapes. Along the row the circles and squares alternated except for one repeated pair, which was the target and was either a pair of adjacent circles or a pair of adjacent squares. The task was to press a key corresponding to the target pair shape as quickly as possible. An element of the target pair could appear in any position in the row except for the outer two positions. There were nine spacing conditions, which are shown in Figure 39 above the dashed line for a square target pair in positions 5 and 6 (counting from the left).

The spacing conditions were specifically designed to investigate effects of proximity by varying target pair separation (the distance between target pair elements) and nontarget pair separation (the distance between a target element and the neighboring nontarget shape). In each condition, spacing between the shapes was varied to form three grouping conditions. In the *neutral conditions*, each shape was equidistant from its neighbors (conditions numbered 1, 5 and 9, in Figure 39). In *within-group conditions*, the target pair was relatively close together (conditions 2, 3 and 6). In *between-group conditions*, the target pair was relatively far apart (conditions 4, 7, and 8).

For this experiment, Palmer and Beck predicted that proximity should bias the target pair in within-group conditions to be perceived as a group. For between-group conditions, they predicted that proximity should bias the target elements to be perceived as part of two different groups. For the neutral conditions, they claimed that proximity should not bias the target pair in either direction. Generally, they expected the target pair to be detected more quickly when it is perceived as part of the same group. In turn, they predicted that response times will tend to be highest for between-group trials, lowest for within-group trials, and somewhere in between for neutral trials.

Palmer and Beck’s results (Figure 40, top left plot, $n = 11$, and Figure 41a, left set of bars) indicated that performance tends to be best (fastest) for the within-group conditions (white bars, i.e., conditions 2, 3, and 6), middling for the neutral conditions (gray bars, i.e., conditions 1, 5, and 9), and worse for between-group conditions (black bars, i.e., conditions 4, 7, and 8).




















		Nontarget Pair Separation								
		0.5			1.0			1.5		
Experiment 1 Stimuli (with square target pair)	Target Pair Separation	0.5	1		2		3			
		1.0	4		5		6			
		1.5	7		8		9			
Additional Stimuli for Experiment 2 (with square target pair)		2.0	10		11		12			
Experiment 3 Stimuli (with circle target pair)	W1		N1		B1					
	W2				B2					
	W3				B3					

Figure 39. Screenshots of stimuli from Experiments 1-3. All spacing conditions are shown with the target pair in positions 5 and 6 (counting from the left) for ease of comparison.

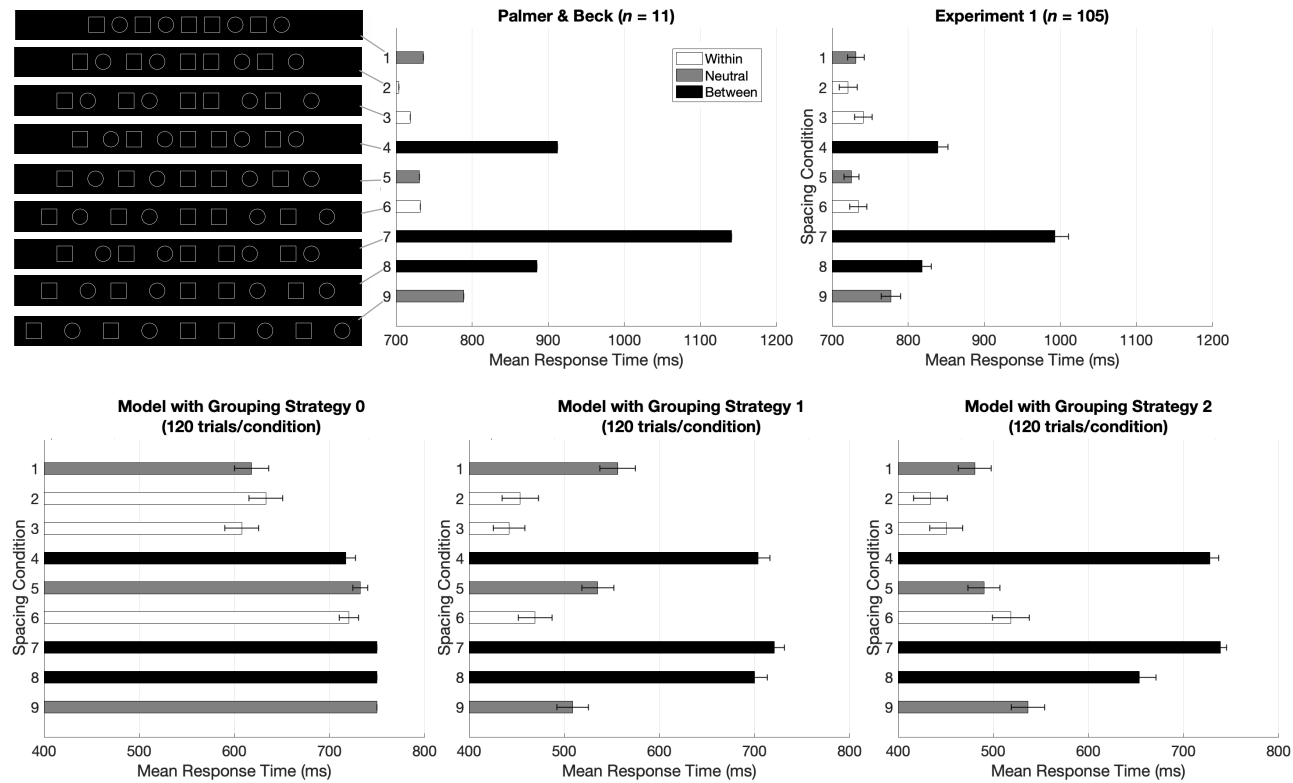


Figure 40. Results by spacing condition. Top row: results from the Palmer and Beck (2007) experiment and Experiment 1 for each spacing condition. No error bars are shown for the original data since Palmer and Beck did not provide any measure of variability. Bottom row: results from simulations with three different grouping strategies. Error bars represent one standard error of the mean. Each spacing condition is numbered from 1-9 (see Figure 39), and each spacing condition falls under one of three grouping conditions, i.e., neutral, within-group, and between-group, indicated by gray, white, and black bars, respectively.

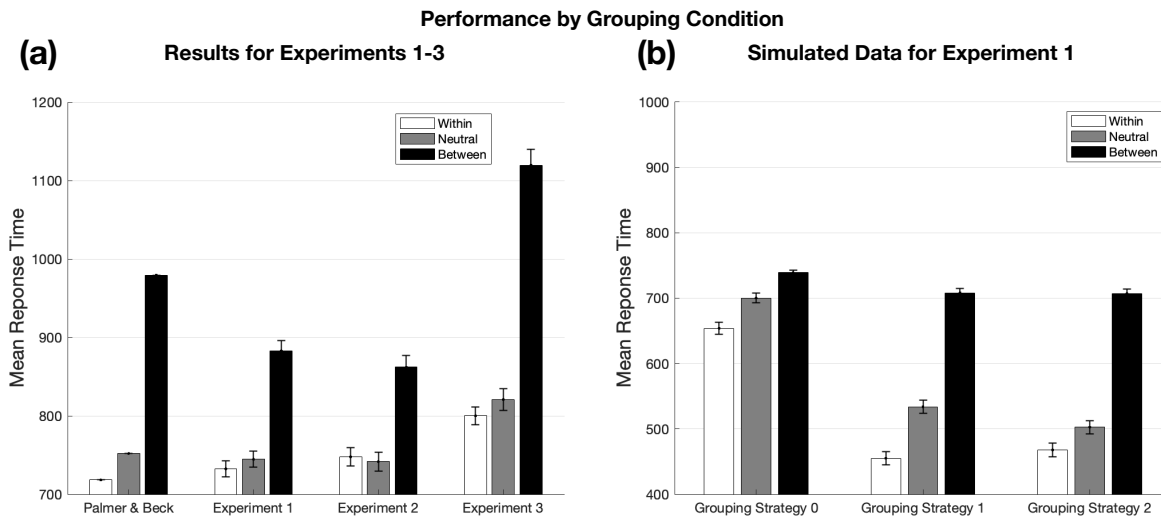


Figure 41. Results by grouping condition. Plot (a) shows mean response times for grouping conditions for the original data (which are an estimate given mean response times for spacing conditions reported by Palmer & Beck, 2007; means for grouping conditions were not provided) and Experiments 1-3. Plot (b) shows grouping condition mean response times for each simulated grouping strategy. Error bars represent one standard error of the mean.

Their prediction that grouped pairs of repeated elements would be discriminated faster than ungrouped pairs was supported by these results: the targets in within-group conditions could form groups by proximity. Their prediction about the relative performance on within-group and neutral conditions was also supported: performance for neutral conditions was worse than the within-group conditions. However, even though their results indicated that the difference in response times between the neutral and within-group conditions was statistically significant, they noted that the mean response times for these grouping conditions were relatively similar and speculated that this similarity was due to the target pair in neutral conditions being perceived as a group due to shape similarity.

Experiment 1: Replication of Palmer and Beck (2007)

Due to the original experiment's small sample size, I conducted a replication of experiment 1 in Palmer and Beck (2007). Deviations between the original experiment and the replication experiment are as follows. First, the replication, unlike the original lab-based experiment, was conducted online. This format led to the following modifications to the experiment: (a) the replication used participants' keyboards to register responses rather than a button box, (b) the replication was self-paced where the participant was prompted to press a key to start each trial

instead of having the next trial start automatically after completion of the previous trial (This was done to reduce the risk of participants who become distracted by something in their environment from giving incorrect and/or very slow responses for multiple consecutive trials; with self-paced trials, participants should still be able to get in to a similar rhythm as for automated paced trials without the risk of environmental distractions impeding performance on multiple trials.), and (c) the total number of experimental trials was reduced from 576 (in 8 blocks) to 144 (the equivalent of 2 blocks) to keep the entire experimental session within a 30 minute time slot with the aim of increasing the likelihood that participants complete the experiment and maintain focus throughout.

Second, rather than providing auditory feedback (a beep) for incorrect responses, I used visual feedback (the word 'Incorrect'). Third, although the original experiment counterbalanced between left and right keys on a button box, I used the T-key and Y-key to register responses and did not counterbalance them between participants. Finally, the original experiment used black line figures on a white background, but for the model simulations it was better to use white line figures on a black background. So, I used the same coloring for the experiment.

Method

Participants. Recall that since my aim is to measure each response time with a precision that would have a standard error of 10 milliseconds, I planned for a sample size of 100 observers. (This is an estimate based on similar studies; Palmer & Beck, 2007, do not provide the standard deviations for their data.) Due to excess sign-ups for the experiment, a total of 130 observers participated in the experiment. Data from 25 participants who had an error rate greater than 10% was removed from analysis. In the original and similar experiments (Beck & Palmer, 2002; Palmer & Beck, 2007) subjects with an error rate above 5% were excluded. In the present replication experiment, the data from this subset of subjects ($n = 80$) largely followed the same pattern of results for each spacing condition as those from the set of subjects ($n = 105$) with an error rate of 10% or less.

Apparatus. Participants used the T-key and Y-key on their keyboard to register responses for circle and square target pairs, respectively. The keys were chosen since they are close together and relatively centered on most keyboards.

Stimuli. Examples of stimuli used in the experiment are provided in Figure 39 above the dashed line. As in the original experiment, there were two possible target shapes (square or circle), four possible positions of the target pair (3-4, 4-5, 5-6, 6-7) and 9 spacing conditions (conditions numbered 1-9 in Figure 39). For the hypothetical participant described in the “General Experiment Methodology” section above, the diameter of a circle subtended approximately 1.0° of the visual angle, a separation of 0.5 entailed that there was a gap between two shapes that was half the width of a square (i.e., the gap subtended approximately 0.5°), and the width of the entire row of figures, i.e., the distance from the leftmost shape to the rightmost shape, subtended between approximately 13.0° to 20.8° depending on the spacing condition. The row was in a black rectangle with a constant width and height that subtended approximately 22.2° and 14.8° , respectively.

Procedure. After reading instructions that explained the task, encouraged them to make responses as quickly and as accurately as possible, participants scrolled to the bottom of the webpage, where the experiment took place. Figure 42 schematizes what occurred on a trial. A participant initiated the first trial by pressing the B-key on their keyboard. 500 milliseconds after this key was pressed, the stimulus appeared and was shown until the participant responded. The participant was to respond by pressing the T-key if the target pair were circles or the Y-key if the target pair were squares. Since the original experiment used a button box and asked participants to place an index finger on each button, this was emulated by instructing participants to rest their left finger on the T-key and right finger on the Y-key for the duration of the experiment. After completing 36 practice trials in which all spacing conditions were experienced, the participant completed a total of 144 experimental trials. All conditions were randomly interleaved. If a response was incorrect, too fast (if the response time was less than 100 ms), or too slow (if the response time was greater than 3000 ms), participants were given feedback at the end of a trial and subsequently were prompted to press the B-key to initiate the next trial. Otherwise, i.e., if a correct response was given and the response time was not too fast or slow, the participant was prompted to press the B-key to start the next trial. Incorrect trials and trials with reaction times lower than 100 ms or greater than 3000 ms were omitted from analysis.

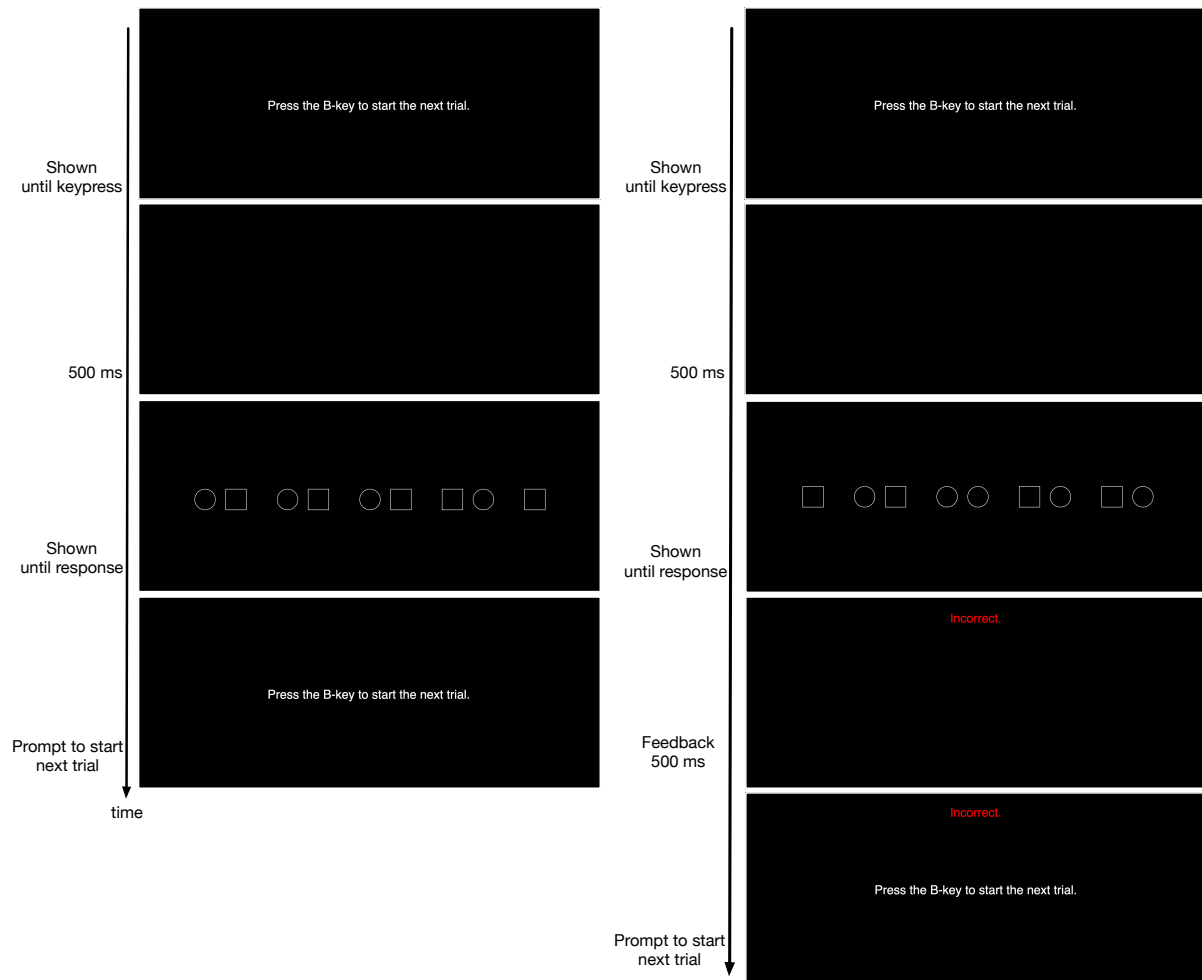


Figure 42. The left column shows the time course of a correct trial with a between-group stimulus, and right column shows the time course of an incorrect trial with a within-group stimulus. This trial structure was used for Experiments 1-3. The black rectangle has been cropped to take up less room here.

Results and Discussion

Mean response times for correct trials for each spacing condition are shown in the top right plot of Figure 40 and in Table 1.

Table 1. Experiment 1 Descriptive Statistics and Correlations for Spacing Conditions

Condition	Response Time (ms)									
	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1	731.45	113.12	—							
2	721.17	122.03	.77	—						
3	741.32	116.90	.75	.75	—					
4	839.37	132.97	.66	.72	.67	—				
5	725.58	99.04	.78	.75	.75	.71	—			
6	734.60	115.02	.80	.78	.72	.69	.75	—		
7	993.09	178.32	.62	.67	.62	.73	.65	.60	—	
8	818.35	127.05	.77	.73	.71	.73	.76	.69	.68	—
9	777.50	127.60	.78	.71	.68	.75	.76	.76	.68	.74

Note. *M* indicates mean, *SD* indicates standard deviation, and columns 1-8 indicate correlations across spacing conditions.

The spacing conditions were pooled by grouping condition. Mean response times for correct trials for each grouping condition are shown as the second set of bars in Figure 41a and in Table 2.

Since I am interested in whether there is a significant difference in mean response times for grouping conditions, I ran an ANOVA model with three contrasts. This and all ANOVA models below were run in R (version 4.0.2; R Core Team, 2020) using the ez package (Lawrence, 2016).

Table 2. Experiment 1 Descriptive Statistics and Correlations for Grouping Conditions

Condition	Response Time (ms)			
	<i>M</i>	<i>SD</i>	Within	Neutral
Within	732.47	107.96	—	
Neutral	744.67	104.45	.88	—
Between	882.89	130.82	.82	.84

Note. *M* indicates mean, *SD* indicates standard deviation, and columns labelled “Within” and “Neutral” indicate correlations across grouping conditions.

A repeated measures ANOVA showed grouping condition had a significant effect on response time, $F(1.73, 179.45) = 328.71, p < .001$ (since Mauchly’s test indicated a violation of sphericity, $\epsilon = .86$, Huyn-Feldt corrected results are reported). Planned contrasts indicated that response times were significantly higher for the between-group condition compared to the within-group condition, $t(208) = -23.08, p < .001$, and for the between-group condition compared to the neutral condition, $t(208) = -21.21, p < .001$. There was no significant difference in response times for the neutral condition compared to the within-group condition, $t(208) = 1.87, p = .063$.

The pattern of results for the spacing conditions of this experiment largely replicates that in Palmer and Beck (2007). A key similarity is that the mean response times for the between-group condition are significantly higher than the neutral condition. A difference is that the response times for neural and within-group conditions did not significantly differ. The low sample size of the original experiment and the results of their experiment 3 (in which they used color, rather than proximity, to investigate differences in performance by grouping condition), which indicated no significant difference between performance on neutral and within-group conditions, suggest that this discrepancy could be the result of sampling variations. Additionally, response times are compressed for my experiment, which probably reflects differences in equipment and context.

Model Simulations of Experiment 1

Simulated Grouping Strategies

To explore the impact of V2 connections and selection signal size on performance, simulations implementing three grouping strategies were conducted. Recall that a grouping strategy consists of a connection strategy (what elements are connected) and a selection strategy (size, location and timing of selection signals). The simulations differed in their connection strategy and selection signal size strategy as indicated in Figure 43b. These components of the grouping strategies and as well as the selection signal timing strategy common all of these grouping strategies will be presented in the next two subsections.

Connection strategies. A connection strategy that would facilitate selection, segmentation, and identification of the target pairs would be to connect only the target pairs if possible. Since target pairs are always the same shape and next to each other, this strategy can be implemented by, in terms of Gestalt grouping principles, grouping by shape. In terms of the model, an observer can tune the four connection parameters so as to promote connections only between nearby same-shaped objects provided that these objects are appropriately aligned and spaced, e.g., equidistant from each other and objects of other shapes (which was demonstrated in the grouping by shape example in the “Combining Connection Controllers to Promote Gestalt Groupings” section above). With such connection parameter values, i.e., with at least a positive Spread Controller duration value, which promotes connection spread, and the Long Controller Circuit on, which can be used to connect nearby same-shaped objects only, the model is able to connect only target pairs in the neutral (spacing conditions 1, 5, 9) and within-group (spacing conditions 2, 3, 6) conditions. Since there is a different minimum inter-shape distance in each of the within-group conditions, there were a total of three sets of parameters. In other words, the further apart the target pair is, the longer the Spread Controller needs to be on to allow the connections to spread far enough to bridge the space between them. The set of parameters used for each spacing condition is indicated by the Roman numerals next to each image in Figure 43c (see the caption for the connection parameter values used in these simulations).

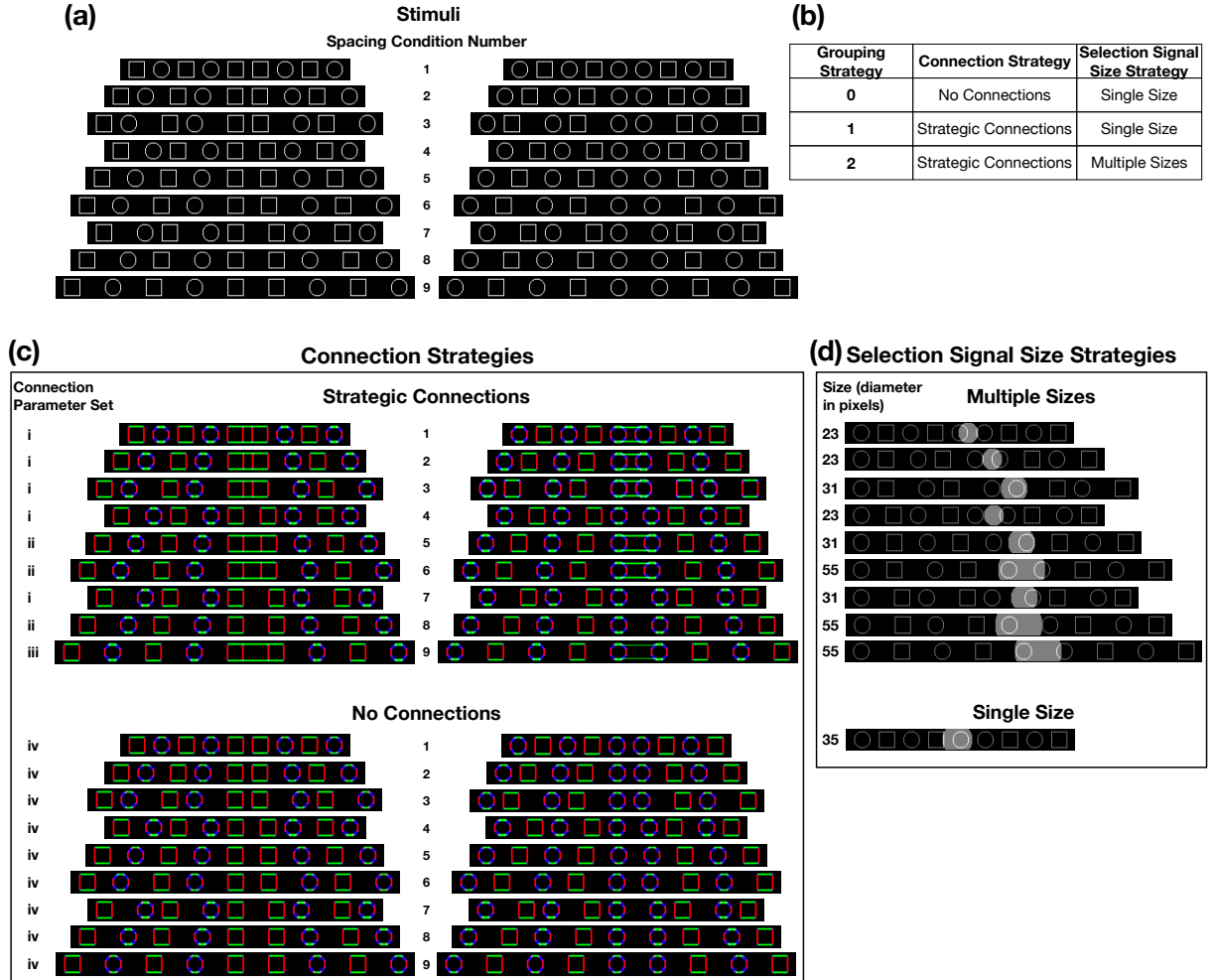


Figure 43. (a) All of the stimuli for the Experiment 1 simulations. (b) A table summarizes the differences in grouping strategies, which references (c) and (d). Box (c) shows two connection strategies, and box (d) indicates two selection signal size strategies. The connection parameter sets numbered by i-iv in box (c) correspond to the following values for the horizontal connection circuits: (i) Spread Controller duration 30 ms, Long Controller input 2.0, Short Controller input 0.8; (ii) Spread Controller duration 60 ms, Long Controller input 1.7, Short Controller input 0.8; (iii) Spread Controller duration 95 ms, Long Controller input 1.0, Short Controller input 0.8; (iv) Spread Controller duration 0 ms, Long Controller input 1000, Short Controller input 0.8. Spread Controller onset was fixed at 0 ms for all simulations.

For the between-group conditions (spacing conditions 4, 7, 8), targets and distractors are too close to prevent them from connecting while also connecting target pairs; there is no set of connection parameters that connects target pairs only for the between-group conditions. So, the most advantageous connection strategy for the between-group conditions is to have no connections.

This connection strategy gives rise to two related questions. First, do we need an additional set of connection parameters for the between-group conditions, i.e., a set in which the Connection Circuit is turned off so there are no connections? Second, how might an observer know when to apply a particular set of connection parameters? There are at least three sets of connection parameters, but the stimuli are randomized. So, an observer would not know whether, e.g., a stimulus from the neutral condition with a target pair separated by half a square width (i.e., spacing condition 1) or by 1.5 square widths (i.e., spacing condition 9) will be presented on the next trial, or whether a neutral or between-group stimulus will be presented on the next trial. So, how does the observer know which set of connection parameters to use for any given trial?

To answer the second question for the neutral (spacing conditions 1, 5, 9) and within-group (spacing conditions 2, 3, 6) conditions, I propose that an observer applies a particular set of connection parameters based on the smallest inter-shape distance between the shapes. For example, if the smallest inter-shape distance in a given stimulus image is half a square width, then the observer implements connection parameter set (i) in Figure 43a. If we apply connection parameters (i) to the between-group conditions, e.g., apply connection parameters set (i) if the smallest inter-shape distance for any stimulus image is half a square width, it turns out that this set of parameters not only connects target pairs in the within-group condition, e.g., condition 2, and in the neutral conditions, e.g., condition 1, but also deters connections in the between-group conditions, e.g., condition 4 (see Figure 43c under ‘Strategic Connections’ for the connections that form when this strategy is used). So, to answer the first question, only three sets of connection parameters are needed to implement this connection strategy. In turn, one connection strategy that an observer might learn to promote performance on this task is to associate the smallest distance between elements in a stimulus with the connection parameters that promote V2 connections among target pairs if it is possible to connect them. If it happens to be a between-group stimulus, there is no disadvantage in using this strategy: the target pair in these cases will never connect together in a way that allows only the target pair to be selected and segmented out.

To see how these strategic connections impact performance, simulations were conducted with (Grouping Strategies 1 and 2 – see Figure 43b) and without connections (Grouping Strategy 0). The V2 bipole activity is shown in Figure 43c under ‘No connections’, which used connection parameter set (iv) that effectively turned off the connection circuit.

Selection strategies. To explore the impact of different selection signal sizes on performance, two selection signal size strategies were used. Grouping Strategies 0 and 1 used the Single Size Strategy where a selection signal of a single size was used for all conditions, while Grouping Strategy 2 used the Multiple Sizes Strategy. For each size strategy, Figure 43d shows the selection signal maps overlaid on to the stimulus image to provide a sense of size of the selection signal relative to stimulus elements. The size for the Single Size Strategy was chosen to be large enough to possibly select and segment out the unconnected target pair in between-group conditions and, thereby, promote grouping via selection signals. Additionally, the selection signal was not made so large that it might overlap with a distractor in the spacing conditions where the shapes are relatively close and, thus, reduces such interference. The sizes for the Multiple Sizes Strategy were chosen in view of the results of pilot simulations. As for the Strategic Connections Strategy, which involves three sets of connection parameters, a similar question arises for the Multiple Sizes Strategy: How does an observer know which selection signal size to use on a given trial? I propose that observers who use the Multiple Sizes Strategy select a particular size by using gist information about row width, i.e., the smallest size is used for stimuli with a relatively short distance from the extreme elements in the row (spacing conditions 1, 2, 4), the largest size for rows with elements that span a relatively long distance (spacing conditions 6, 8, 9), and a medium size of rows of intermediate width (spacing conditions 3, 5, 7).

Apart from this difference in selection signal size, the selection strategy implemented in each grouping strategy was the same. The model used a feature-based strategy for determining the position of selection signals. A horizontal feature filter was used because targets, which tend to be united by horizontal connections via the Strategic Connections strategy, often have more horizontal signal. This can be seen by comparing the feature filters produced using bipole cell activity with Strategic Connections compared with those from No Connections shown in Figure 44.

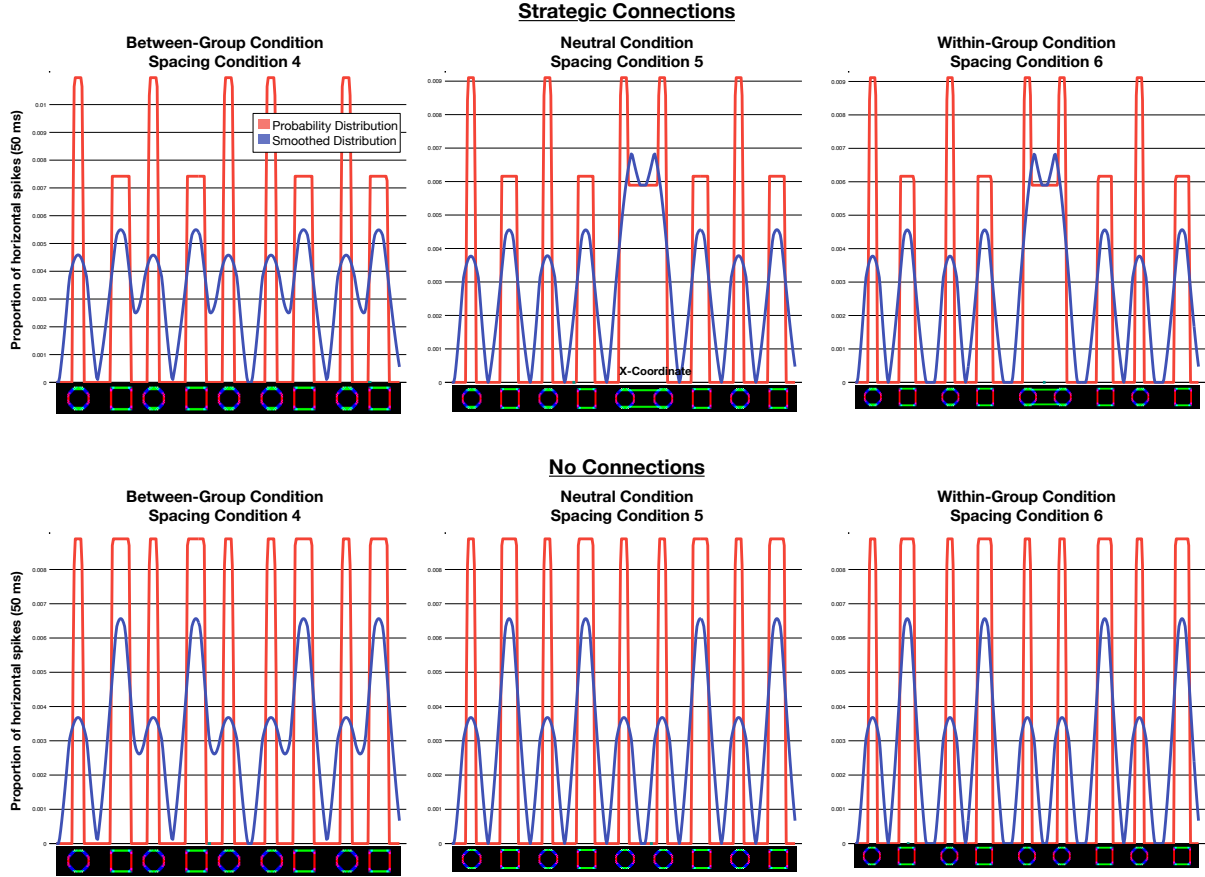


Figure 44. Horizontal feature filters for spacing conditions 4, 5, and 6. Top row provides those for Strategic Connections, and bottom row for No Connections.

With Strategic Connections, there are relatively wide and high peaks in the distribution above the target pair due to the horizontal connections between them in the neutral and within-group conditions. So, given this connection strategy, this feature filter was chosen since it will tend to guide selection signals to the area where a connected target pair is located for the majority of spacing conditions.

There were two selection signals, which is the same number used by Francis et al. (2017) in simulations of vernier discrimination tasks. Unlike the simulations conducted by Francis et al. in which location of selection signals were at a fixed location for the duration of a trial, here each selection signal shifted every 150 ms to a location sampled from the probability distribution created by smoothing the sum of horizontal activity in each column of the V2 output from the previous 50 ms. 150 ms was chosen because this is the shortest time possible for V2 to elicit V4 signals that are strong enough to identify whether the selected element was a target. The selection

signals were temporally staggered, e.g., the first started at 100 ms of simulation time and the second started at 150 ms (Figure 45).

Target identification algorithm. Once an element was selected and segmented into a segmentation layer, the associated V4 activity was checked for symmetry: if the segmented element was (nearly) symmetrical and of appropriate width, e.g., wider than a single shape, then it was identified as a target pair and the trial ended. If the segmented element did not meet these conditions, then the non-base segmentation layer was reset and the trial continued (for examples, see Figure 46).

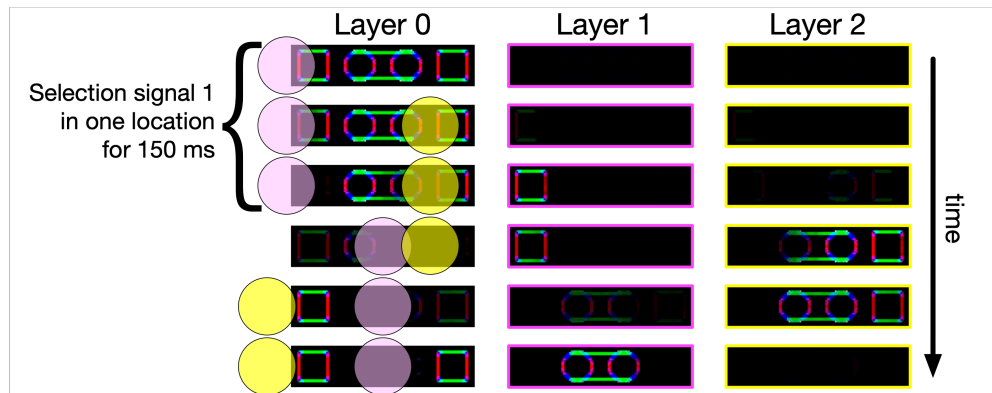


Figure 45. Illustration of the timing of selection signals used for the simulations of Experiment 1. Two selection signals are used but are temporally staggered: the selection signal represented by a yellow circle begins 50 ms after the pink selection signal. After 150 ms at one location, i.e., after a 150 ms selection cycle (which was defined in Section 5.1), each selection signal shifts to a different location and its segmentation layer is reset.

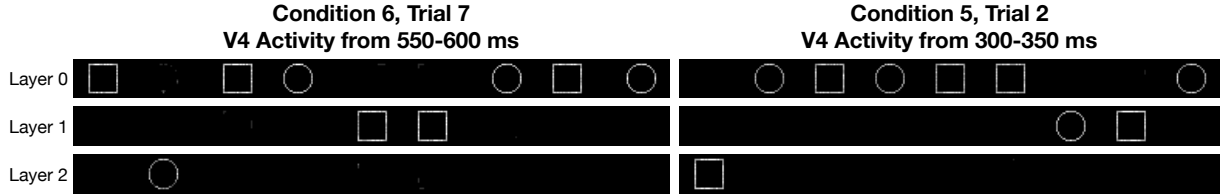


Figure 46. Examples of V4 activity that are input to the target identification algorithm. Left column: The target identification algorithm identified a target in layer 1 because the V4 activity is generally symmetrical and of suitable width. In turn, the trial ended at 600 ms. Right column: Layer 1 has activity that is of suitable width yet is not symmetrical. Layer 2 has a single object. Although this activity is symmetrical, it is not of suitable width for a target pair. So, the target identification algorithm did not detect a target in either V4 layer, and the trial continued.

Simulation Stimuli, Method, and Procedure

The stimuli for the simulations were the set of stimuli shown in Figure 43a. These are BMP files I created using a custom Python script and were used as input for the model. I attempted to make the proportions of stimulus elements and spacing as close as possible to those used in Experiment 1 while also reducing the size of the stimuli to facilitate faster runs of simulations. The diameter of a square was 19 pixels, and the distance from the edges of the farthest shapes in the row varied from 251 to 403 pixels, depending on the spacing condition. Target pairs were always in the same positions, i.e., in positions 5-6. Palmer and Beck (2007) reported slower reaction times for targets in positions 3-4 and 6-7; however, since, among other things, the model lacks a retina that might explain better central acuity, the positions of the targets are irrelevant for the simulations.

On each simulated trial, a stimulus was shown to the model until either the target identification algorithm indicated that the target pair was detected or a 750 ms maximum time was reached. For each spacing condition, 120 trials were conducted (60 trials with a circle target pair, and 60 with a square target pair). Simulations were run in batches of 10 trials due to a memory leak that occurs when using NEST 2.14.0, but such batch runs do not impact model performance. Since the order of presenting the conditions is irrelevant for model performance, a single stimulus image with a particular spacing condition and specific target pair (e.g., spacing condition 1 with a square target pair) was input to the model on each batch.

Three sets of simulations were conducted, and each set implements one of three grouping strategies, which are summarized in Figure 43b. Grouping Strategy 0 implemented a connection strategy where no connections formed and used a fixed selection signal size for all conditions.

Grouping Strategy 1 also used a fixed selection signal size but implemented the Strategic Connection Strategy shown in the top set of images in Figure 43c. Like Grouping Strategy 1, Grouping Strategy 2 used strategic connections, but it used the Multiple Sizes selection signal strategy indicated in the top set of images in Figure 43d.

Model Results and Discussion

For each condition, the time it took for the model to find a target or time out on each trial was averaged across simulated trials to provide a measure of mean response time that can be compared with that of human observers. Model results for each spacing condition are shown in Figure 40, bottom row. Model results from Grouping Strategies 1 and 2 are well correlated with the results from my Experiment 1 ($r = .81$ and $r = .88$, respectively) relative to Grouping Strategy 0 and Experiment 1 ($r = .53$).

As shown in Figure 40, bottom row, Grouping Strategy 0 has much higher response times overall compared with the other grouping strategies, and performance tends to decrease as distance between the target pair increases. The latter trend is to be expected since, due to the fixed selection signal size that can just barely cover two stimulus elements at 1.5 target-pair separation, it is less likely that the observer simultaneously selects both unconnected target elements as their separation increases.

Unlike Grouping Strategy 0, Grouping Strategies 1 and 2 result in relatively good performance for the within-group and neutral conditions (bottom row of Figure 40, white and gray bars, respectively) compared to the between-group conditions (black bars). The patterns of results more closely resemble those of human observers in Experiment 1 (Figure 40, top right plot). In turn, it seems the connection strategy used in these grouping strategies, i.e., connect target pairs only (if possible), promotes performance for within-group and neutral conditions.

Interestingly, the pattern of results for Grouping Strategy 1 reflects Palmer and Beck's prediction that response times should be lowest for within-group conditions (2, 3, 6), middling for neutral conditions (1, 6, 9), and highest for between-group conditions (4, 7, 8). This can be readily seen in Figure 41b, middle set of bars, which shows results for each grouping condition. However, the pattern produced for Grouping Strategy 2 more closely resembles the results from Experiment 1 in that response times for neutral and within-group conditions are similar (compare Figure 41a, second set of bars, with Figure 41b, third set of bars). In turn, out of the grouping strategies

simulated, Grouping Strategy 2, which had a strategic connection strategy coupled with a selection signal size strategy with multiple sizes, leads to results that best match those of human observers. This suggests that for this apparently simple task, participants are using a somewhat complicated grouping strategy that combines a connection strategy that involves three sets of connection parameters with a selection signal size strategy that involves multiple selection signal sizes.

Experiment 2: Same Task, Different Stimulus Set

One of my main claims is that the grouping strategy an observer adopts depends on the task and stimulus set. I investigated this claim by conducting Experiment 2, which is a variation of Experiment 1 that has the same task, number of spacing conditions, and number of possible row widths, i.e., distance from the leftmost to the rightmost stimulus elements. However, rather than having the majority of stimulus images being spaced such that an observer could group them by shape, i.e., 2/3 of the stimulus images were either within-group or neutral in Experiment 1, Experiment 2 has 2/3 of its stimulus images from the between-group condition. According to the model, an observer should group stimulus elements such as to promote performance. In Experiment 1 it seems beneficial to connect target pairs (if possible) since this will promote performance for 2/3 of the stimuli. For the other 1/3 of the stimulus images, i.e., for the between-group condition, this connection strategy was not detrimental to performance since, although target pairs were too far to connect, target elements did not connect to distractors. This connection strategy coupled with a horizontal feature filter, led to lower performance in the between-group conditions, consistent with the empirical data.

Since 2/3 of the stimulus images in Experiment 2 are of the between-group condition, the model predicts that an observer will likely not use the same grouping strategy; since this strategy will only promote performance for 1/3 of the stimuli, it seems likely that an observer would use an alternative grouping strategy. If, as for Experiment 2, the connection strategy causes the elements in the between-group conditions to not connect, then a possible selection strategy is to attempt to select a single element, perhaps using stimulus element features to guide attention to the location of an element, and then use a location-based placement strategy to guide attention to the element at the opposite side of the relatively wide gap. The two selected elements can thereby be isolated and compared to see if they are of the same shape. This should cause faster response times for the between-group conditions where the target elements are separated by a relatively

wide gap. (The mechanics of this selection strategy would have to be worked out, but it does not *prima facie* seem to be an unreasonable strategy. Small selection signals would probably work best here to avoid also selecting the closest nearby element)

Assuming No Connections are used, the model predicts an increase in response time for the within-group and neutral conditions. If the aim is to initially select a single item, then having no connections would facilitate such selection; however, given the hypothetical selection strategy that involves directing attention to the neighboring object across the wider gap, the relatively nearby target in within-group conditions would be less likely to be selected and identified. Thus, if No Connections are used, we should see an increase in response times for the within-group and neutral conditions.

Would Strategic Connections promote performance for this stimulus set? This would connect target elements only in the within-group and neutral conditions, and it would result in no connections in the between-group conditions. It's possible for a step to be incorporated into the selection strategy described above: after the first item is selected, the segmented signal is checked to see whether a pair of same-shaped items was selected. If so, then the trial stops since the target pair was found. If not, then a second selection signal is directed to the neighboring element that is farther away from the selected element. So, if Strategic Connections are used, the model predicts similar performance on within-group and neutral conditions for this experiment as for Experiment 1. A cost of this connection strategy is that it may be more difficult to select a single element at the beginning of a trial with a between-group condition due to the spread of boundary signal before the system settles down to equilibrium; a selection signal may overlap with spread from a neighboring circle and, thereby, select a nontarget pair, especially when a target element is very close to a distractor.

In sum, the model predicts that between-group response times should be lower in Experiment 2 than Experiment 1. If a connection strategy with no connections is adopted, then response times for within-group and neutral conditions should be higher in Experiment 2 than Experiment 1. If a connection strategy in which target pairs only are connected (if possible) is adopted, then within-group and neutral response times should be similar across both experiments.

I generated additional between-group spacing conditions by expanding the set of spacing conditions (Figure 39, fourth row of stimuli) used in Experiment 1. Experiment 2 uses the set of spacing conditions numbered 4-12 in Figure 39, which consists of one within-group (spacing

condition 6), two neutral (conditions 5 and 9), and six between-group (conditions 4, 7-8, and 10-12) spacing conditions.

Method

Participants. For ease of comparing data from this experiment with that of Experiment 1, I aimed to gather data from the same number of participants as in Experiment 1 who had error rates no greater than 10% ($n = 105$). I ended up with a total of 121 participants, 105 of which had an acceptable error rate and were used in the analysis.

Apparatus, stimuli, and procedure. All aspects were identical to those of Experiment 1 except the spacing conditions are those numbered 4-12 in Figure 39. For this set of stimuli, the width of the entire row of figures, i.e., the distance from the leftmost shape to the rightmost shape, subtended between approximately 15.0° to 22.8° , depending on the spacing condition, for the hypothetical participant described in the “General Experiment Methodology” section above. Additionally, to accommodate the widest spacing condition, the width of the black rectangle (where the experiment took place on the webpage) was increased so that it subtended approximately 27.0° .

Results and Discussion

Mean response times for correct trials for each spacing condition are shown in Table 3.

Spacing conditions were also pooled by grouping condition. Mean response times for correct trials for each grouping condition are shown as the third set of bars in Figure 41a and in Table 4.

As in Experiment 1, I ran an ANOVA model with three planned contrasts to compare performance by grouping condition.

Table 3. Experiment 2 Descriptive Statistics and Correlations for Spacing Conditions

Condition	Response Time (ms)									
	<i>M</i>	<i>SD</i>	4	5	6	7	8	9	10	11
4	803.81	150.40	—							
5	739.29	137.11	.82	—						
6	747.83	122.69	.83	.82	—					
7	922.85	184.58	.86	.82	.80	—				
8	789.43	145.86	.87	.84	.84	.85	—			
9	744.36	119.31	.87	.86	.86	.86	.91	—		
10	991.67	194.15	.78	.73	.71	.83	.78	.78	—	
11	862.57	167.14	.85	.81	.83	.87	.83	.82	.78	—
12	807.29	153.58	.80	.83	.82	.77	.85	.84	.78	.77

Note. *M* indicates mean, *SD* indicates standard deviation, and columns 4-11 indicate correlations across spacing conditions.

Table 4. Experiment 2 Descriptive Statistics and Correlations for Grouping Conditions

Condition	Response Time (ms)			
	<i>M</i>	<i>SD</i>	Within	Neutral
Within	747.83	122.69	—	
Neutral	741.79	123.34	.87	—
Between	862.05	151.94	.87	.93

Note. *M* indicates mean, *SD* indicates standard deviation, and columns labelled “Within” and “Neutral” indicate correlations across grouping conditions.

A repeated measures ANOVA showed grouping condition had a significant effect on response time, $F(1.86, 193.11) = 223.90, p < .001$ (since Mauchly's test indicated a violation of sphericity, $\epsilon = .86$, Huyn-Feldt corrected results are reported). Planned contrasts indicated that response times were significantly higher for the between-group condition compared to the within-group condition, $t(208) = -17.84, p < .001$, and for the between-group condition compared to the neutral condition, $t(208) = -18.78, p < .001$. There was no significant difference in response times for the neutral condition compared to the within-group, $t(208) = -0.94, p = .347$.

Regarding the grouping conditions, the results of Experiment 1 are replicated in this experiment. However, because the main purpose of Experiment 2 is to test whether performance is dependent on the stimulus set, I compared mean response times for spacing conditions that appeared in both Experiment 1 and 2. These are shown in rows 2 and 3 of Figure 39, and Figure 47 plots the response times across experiments for these conditions (the number beside each point in Figure 47 corresponds to the spacing condition number in Figure 39).

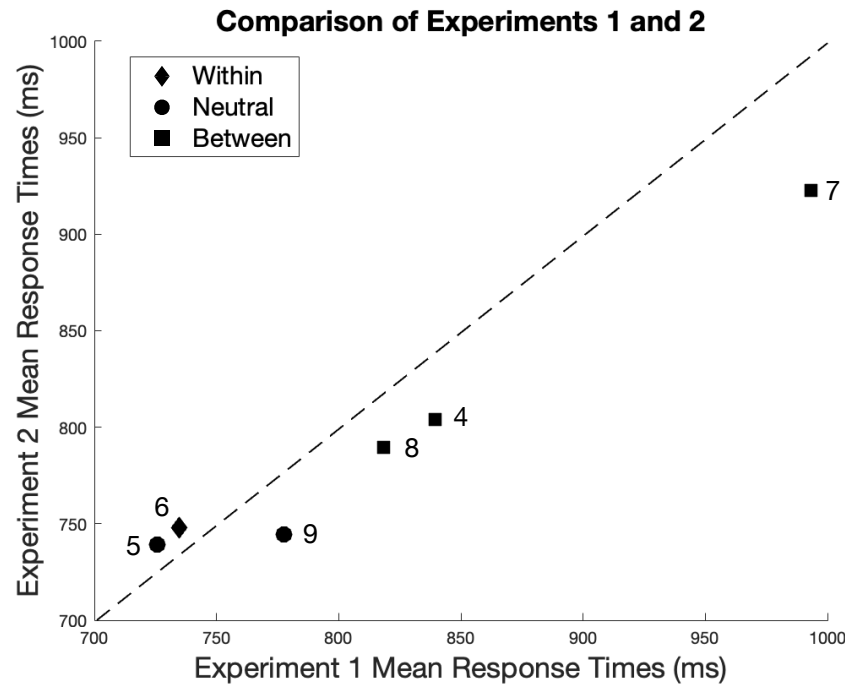


Figure 47. A plot comparing response times for Experiments 1 and 2 for shared spacing conditions. The number labels correspond to the conditions shown in rows 2 and 3 of Figure 39.

If a point falls on the dashed diagonal line, then the mean response time was the same for both experiments. If the point is above the dashed line, then the mean response time was lower in Experiment 1 than in Experiment 2 for that spacing condition. If the point is below the line, then the mean response time was higher in Experiment 1. Qualitatively, mean response times for all between-group conditions (squares in Figure 47) were faster in Experiment 2, and mean response times for all within-group and neutral conditions (diamond and circle points in Figure 47, respectively) except condition 9 were slower in Experiment 2.

To compare the performance of participants in Experiment 1 and Experiment 2, which had different stimulus sets, on conditions shared between these experiments (spacing conditions 4-9), a mixed ANOVA was run. There was a significant effect of spacing condition, $F(3.70, 770.17) = 313.69, p < .001$, and a significant interaction between stimulus set and spacing condition, $F(3.70, 770.17) = 11.64, p = .011$ (since Mauchly's test indicated a violation of sphericity, $\epsilon = .74$, Greenhouse-Geisser corrected results are reported). Stimulus set did not have a significant effect on response time, $F(1, 280) = 1.89, p = .171$. Planned contrasts indicated that response times were significantly lower for between-group spacing conditions (4, 8, 7) for Experiment 2 than in Experiment 1, $t(208) = -2.28, p = .023$. There was no significant difference in response times for the within-group and neutral conditions in Experiment 1 compared with response times for those conditions in Experiment 2, $t(208) = -0.13, p = .0894$.

The stimulus set of Experiment 2 had a majority of between-group conditions, while that of Experiment 1 had a majority of within-group and neutral conditions. The results of Experiment 2 are in line with the expected results: performance for between-group conditions is improved for this stimulus set compared to that of Experiment 1, which indicates that participants are using an alternative selection strategy. The similar performance for within-group and neutral conditions across the experiments indicates that the same connection strategy was used. Overall, this supports my claim that the grouping strategy used is dependent on the stimulus set and chosen to promote overall performance.

Experiment 3: Controlled Row Width to Induce a Simpler Selection Strategy

In this experiment, I investigate whether controlling for total row width of each stimulus image impacts reaction time. Recall that in view of how well the results of Grouping Strategy 2 matched those of Experiment 1, it seemed that observers might be using a complicated selection

strategy in which the chosen selection signal size depends on the total width, i.e., the distance from the leftmost to rightmost shape in a row, of each stimulus image. Sets of spacing conditions with the same row width are easily generated by extending the matrix of spacing conditions used in Experiment 1 shown in Figure 39. Each spacing condition that falls along a diagonal that starts at the bottom left and ends to the top right of the matrix, e.g., conditions 3, 5, and 7, have the same row width. The set of spacing conditions for Experiment 3 are shown in Figure 39 below the solid line. If the matrix above the dashed line is extended to have 7 rows and 7 columns (with a maximum pair separation of 3.5), then this set of stimuli constitutes the matrix's counterdiagonal.

If participants use a selection strategy involving a fixed selection signal size, which was implemented in Grouping Strategy 1, then reaction times should be more indicative of the connection strategy and feature filter that was used and, thus, would support my claim that observers are using a more complicated selection strategy in the original Palmer and Beck experiment since total width was not controlled for. In other words, the results should mirror the model results when Grouping Strategy 1 was implemented, which had a fixed selection signal size, i.e., times should tend to be lowest for within-group conditions, worst for between-group conditions, and somewhere in between for neutral conditions.

Method

Participants. I planned to gather data so that I measured mean response time with a standard error of around 10 milliseconds. Experiment 1 above found standard deviations across participants around 114 milliseconds, so I again planned for approximately 100 participants. I ended up with a total of 143 participants. As in Experiment 1, I used data only from subjects with an error rate of 10% or less ($n = 102$).

Apparatus, stimuli, and procedure. All aspects were identical to those of Experiment 1 except the stimulus set is that shown below the solid line in Figure 39. It consists of a total of seven spacing conditions: three within-group (labelled W1, W2, W3, in Figure 39), one neutral (N1) and three between-group (B1, B2, B3). For this set of stimuli, the width of the entire row of figures was a constant, i.e., the distance from the leftmost shape to the rightmost shape, subtended approximately 24.7° .

Results and Discussion

Mean response times for correct trials for each spacing condition are shown in the second column of Figure 48 and in Table 5.

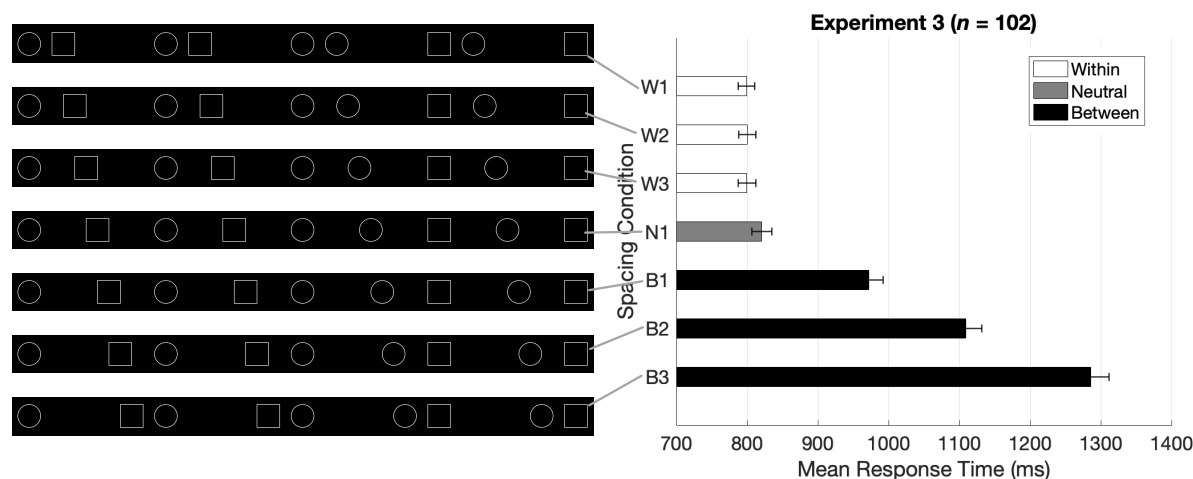


Figure 48. Mean response times for each spacing condition in Experiment 3. Error bars represent one standard error of the mean, and bar color indicates which grouping condition the spacing condition falls under.

Spacing conditions were pooled by grouping condition. Mean response times for correct trials for each grouping condition are shown as the fourth set of bars in Figure 41a and in Table 6.

I again ran an ANOVA model with three planned contrasts to compare performance by grouping condition.

A repeated measures ANOVA showed grouping condition had a significant effect on response time, $F(1.59, 160.34) = 382.91, p < .001$ (since Mauchly's test indicated a violation of sphericity, $\epsilon = .86$, Huyn-Feldt corrected results are reported). Planned contrasts indicated that response times were significantly higher for the between-group condition compared to the within-group condition, $t(202) = -24.73, p < .001$, and for the between-group condition compared to the neutral condition, $t(202) = -23.12, p < .001$. There was no significant difference in response times for the neutral condition compared to the within-group condition, $t(202) = 1.61, p = .109$.

Table 5. Experiment 3 Descriptive Statistics and Correlations for Spacing Conditions

Condition	Response Time (ms)							
	<i>M</i>	<i>SD</i>	W1	W2	W3	N1	B1	B2
W1	799.41	119.56	—					
W2	800.55	123.44	.74	—				
W3	799.89	125.72	.76	.80	—			
N1	820.90	139.84	.70	.68	.27	—		
B1	972.24	203.19	.60	.56	.65	.78	—	
B2	1109.67	225.89	.54	.43	.56	.58	.71	—
B3	1285.55	263.49	.55	.40	.51	.51	.53	.68

Note. *M* indicates mean, *SD* indicates standard deviation, and the other columns indicate correlations across spacing conditions.

Table 6. Experiment 3 Descriptive Statistics and Correlations for Grouping Conditions

Condition	Response Time (ms)			
	<i>M</i>	<i>SD</i>	Within	Neutral
Within	800.07	112.87	—	
Neutral	820.90	139.84	.76	—
Between	1119.67	199.78	.66	.70

Note. *M* indicates mean, *SD* indicates standard deviation, and columns labelled “Within” and “Neutral” indicate correlations across grouping conditions.

Although there is a small difference in response time between within-group and neutral conditions of around 20 ms, this difference was not significant. This is counter to the prediction that controlling for row width would control for selection signal size and, thereby, result in the lowest mean response time for the within-group condition and a middling mean response time for the neutral condition.

However, this prediction was ill-formed for two reasons. First, given that I claim that grouping strategy is dependent on stimulus set, I should have considered model behavior for this particular stimulus set, rather than rely solely on inferences from the Experiment 1 simulations. In future work, I plan to run simulations with the stimulus set used in Experiment 3 to determine what grouping strategy(s) optimizes performance for this stimulus set (which may also offer an explanation for the pattern of results for the between-group spacing conditions). Second, when making this prediction and designing the experiment, I did not consider that it was the combination of selection signal size and crowding that was a main reason for the pattern of results produced with Grouping Strategy 1 in the simulation of Experiment 1. Consider the simulated results in Figure 40, bottom row. Grouping Strategy 1 led to much higher response times for spacing condition 1, which was a neutral condition, compared with spacing condition 2, which was a within-group condition. With the target pair being connected in both conditions, this difference in response times is primarily due to the within-group target pair being farther from distractors than the neutral condition target pair. In turn, a selection signal that happened to land on the target pair in the neutral condition was more likely to overlap with surrounding distractors and, thus, less apt to segment and identify the target pair than the within-group condition. When Grouping Strategy 2 is used, which has the same small selection signal size, mean response time for spacing condition 1 is greatly reduced and more similar to that of spacing condition 2 (compare the bars for conditions 1 and 2 of Figure 40, bottom middle plot with those of the bottom right plot). Thus, with a medium selection signal size, grouping by selection of targets with neighboring distractors was more likely for this neutral condition than for the within-group condition, which led to the difference in performance.

Compared to the Experiment 1 stimuli, the neutral and within-group stimuli of Experiment 3 had much wider gaps surrounding the target pair for within-group and neutral conditions. So, unless a selection signal with a large size that tended to overlap with neighboring distractors was used, response times for the neutral and within-group stimuli should be similar, which is what was

observed. In turn, a selection signal size strategy with a fixed, small size is compatible with the Experiment 3 results but not directly supported by them.

VICKERY (2008)

Vickery (2008) used the visual search task developed by Palmer and Beck (2007) to investigate induced perceptual grouping, which refers to the influence of perceived grouping of some elements on other elements in the scene. The stimuli for Vickery's experiment 2a, which were reproduced for Experiment 4 and shown at the top of Figure 49, demonstrate this phenomenon with proximity.

Each stimulus consists of a top row of 15 large and small circles that are alternated across the row except for the target pair, which is either an adjacent pair of large circles or small circles. The task was to press a key as soon as the target pair was found. In all conditions the center of each circle is the same distance from that of its neighbors. However, Vickery attempted to measure the effects of induced grouping by adding a bottom row of crosses that is irrelevant to the task. Vickery's experiment had three grouping conditions: uniform, within-group, and between-group. Examples of each spacing condition are shown in the top three images of Figure 49.

		Grouping Condition
Experiment 4 Stimuli (with small target pair)	Within-group	
	Uniform	
	Between-group	
Experiment 5 Stimuli (with large target pair)	Within-group	
	Uniform	
	Between-group	

Figure 49. Screenshots showing stimuli that exemplify the grouping conditions used in Experiments 4 and 5. For ease of comparison, here the target pair is always shown in positions 4 and 5, counting from the left.

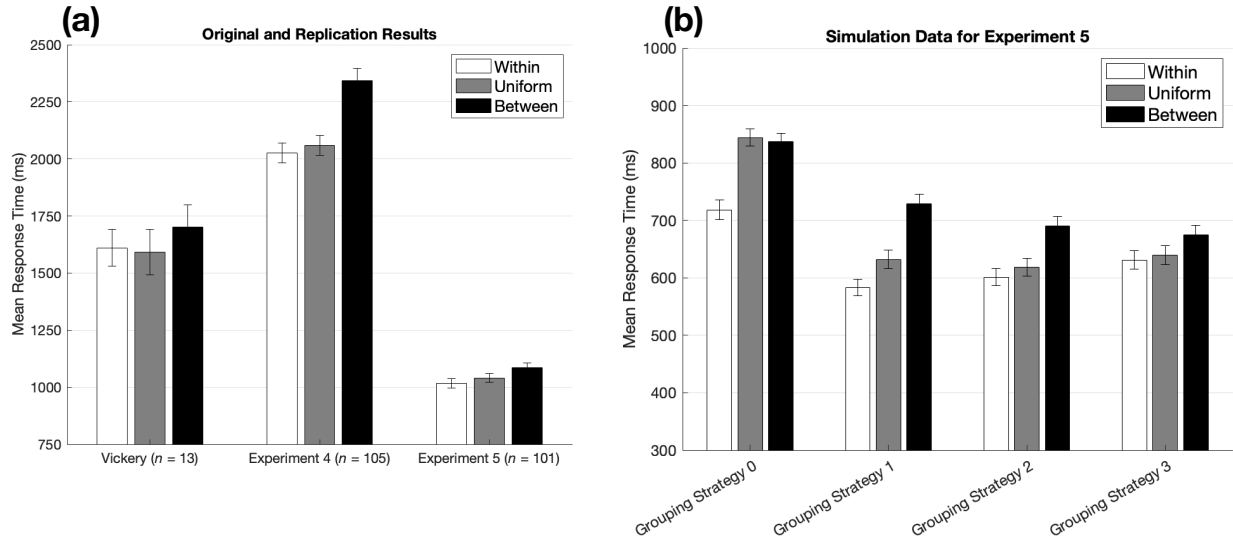


Figure 50. Plot (a) provides mean response times for grouping conditions for the original experiment and Experiments 4 and 5. Plot (b) shows the results of simulations running a version of Experiment 5 with the model implementing different grouping strategies. Error bars represent one standard error of the mean, and bar color indicates the grouping condition.

In the uniform condition, each cross is directly below the center of a circle and, thus, equidistant from neighboring crosses. In the other conditions, crosses were spaced so as to form perceived groups by proximity. Grouped crosses fall under the target pair in the within-group condition, while grouped crosses fall below a target circle and a distractor circle in the between-group condition. Since Vickery hypothesized that the grouped pairs of crosses will induce perceived grouping of the uniformly spaced circles, i.e., the grouping of the crosses ‘spreads’ to the row of equidistant circles, he predicted that the crosses in the within-group condition will tend to improve performance compared to that of the between-group condition.

In line with his prediction that response times should tend to be faster for the within-group condition and slowest for the between-group condition due to the grouping of irrelevant crosses, Vickery reported that mean reaction time for the between-group condition was significantly longer than that of the within-group condition, while the difference between the uniform and within-group conditions was not significant (Figure 50a, left set of bars, $n = 13$). Thus, he concluded that grouping of the irrelevant crosses due to proximity influenced the perception of target items.

Curiously, Vickery (2008) presents this experiment as a version of the Palmer and Beck (2007) experiment; however, he gives no reason for why he increased the number of non-cross elements from 9 to 15. Additionally, Vickery’s experiment also modified how participants

responded. Participants responded in two stages: first they pressed a button as soon as the target pair was found, then a black letter X replaced each circle and the subject clicked on the X's at the locations of the target pair.

To see whether Vickery's results replicate and to conduct the experiment so that it is closer to Palmer and Beck's version of the task, I conducted a replication experiment (Experiment 4) with 15 shapes per row as well as a version with 9 shapes per row (Experiment 5). Additionally, since simulations were only conducted with 9 shape stimuli (due to the computational resources required to simulate the large stimuli with 15 shapes), Experiment 5 provides results for the stimulus set that I was able to simulate.

Experiment 4: Direct Replication of Vickery (2007)

Experiment 4 is a replication of experiment 2a in Vickery (2008). However, Experiment 4 deviates from the original in several ways. First, Experiment 4 is an online, rather than lab-based, experiment. In turn, unlike the original experiment, this experiment is self-paced with participants being prompted to press a key to start each trial. Second, the task used in Experiment 4 was kept closer to Experiment 1, i.e., rather than the two-stage response described above, participants responded by pressing one key if the target was small circles and another key if the target was large circles. Third, the original experiment discarded trials with response times less than 150 ms or greater than 5000 ms. However, in view of pilot results, a 5000 ms cutoff was too short because too many trials were excluded and, in turn, a cut off of 8000 ms was used. Finally, the original experiment used black filled figures on a white background, but for the model simulations it was better to use white filled figures on a black background. So, I used the same color scheme for the experiment.

Method

Participants. Recall that my aim is to measure each response time with a precision that would have a standard error of around 10 milliseconds. The findings in Vickery (2007) indicate that the standard deviations are around 330 milliseconds. Provided that Vickery's standard deviations are representative, running 300 participants seemed rather excessive to achieve a

measure with a standard error of around 10 ms (particularly since I did not plan to simulate this version of the experiment), and a standard error of around 30 ms achievable by running 100 participants seemed reasonable. So, I planned for a sample size of 100 observers. Due to excess sign-ups for the experiment, I ended up with a total of 162 observers. Because this experiment used the same task as Experiment 1, I used the same exclusion criterion and analyzed data only from subjects who had an error rate no greater than 10% ($n = 105$).

Apparatus. Participants used the T-key and Y-key on their keyboard to register responses for large and small target pairs, respectively. The keys were not counterbalanced and were chosen since they are close together and relatively centered on most keyboards.

Stimuli. Examples of stimuli used in the experiment are shown in the first three rows of Figure 49. There were two target sizes (small or large), 3 grouping conditions (within-group, uniform, between-group), and 10 possible positions of the target pair (a target element could appear in any location except the two outer positions [1, 2, 14, or 15] in the row). For the hypothetical participant described in the “General Experiment Methodology” section above, the diameter of a small circle subtended approximately 0.7° of the visual angle, the diameter of a large circle subtended 0.9° , and the width of a cross was the same as the diameter of a small circle. The width of a line composing the cross subtended approximately 0.1° . The center-to-center distance for each adjacent pair of circles and the center-to-center distance of pairs of crosses in the uniform condition subtended approximately 1.8° . In non-uniform conditions, the center-to-center distance of relatively close crosses subtended approximately 1.1° , and that of relatively far crosses subtended approximately 2.6° . The vertical distance from the center of a circle to the center of a cross in the uniform condition subtended approximately 1.9° . The row was in a black rectangle with a constant width and height that subtended approximately 28.3° and 14.8° , respectively.

Procedure. The procedure was identical to that of Experiment 1 with the following exceptions. As shown in Figure 42, the screen was blank for 1000 ms, rather than 500 ms, and, of course, the subsequent stimulus image was, e.g., one shown in the first three rows of Figure 49. The participant was instructed to respond as quickly and as accurately as possible by pressing the T-key if the target pair were large circles or the Y-key if the target pair were small circles. After completing 24 practice trials in which all spacing conditions were experienced, the participant

completed a total of 120 experimental trials. So, there were 40 experimental trials per grouping condition, which is the same as the original experiment. If a response was incorrect, too fast (if the response time was less than 100 ms), or too slow (if the response time was greater than 8000 ms), the participant was given feedback at the end of the trial and were prompted to press the B-key to initiate the next trial. Otherwise, if a correct response was given and the response time was not too fast or slow, the participant was prompted to start the next trial by pressing the B-key. Trials with reaction times lower than 150 ms or greater than 8000 ms were omitted from analysis. Incorrect trials were also omitted from analysis.

Results and Discussion

Mean response times for correct trials for Experiment 4 are shown as the second set of bars in Figure 50a and in Table 7.

I ran an ANOVA with three planned contrasts to compare performance for grouping conditions.

Table 7. Experiment 4 Descriptive Statistics and Correlations for Grouping Conditions

Condition	Response Time (ms)			
	<i>M</i>	<i>SD</i>	Within	Uniform
Within	2025.25	441.80	—	
Uniform	2058.57	440.88	.88	—
Between	2344.08	526.77	.82	.78

Note. *M* indicates mean, *SD* indicates standard deviation, and columns labelled “Within” and “Uniform” indicate correlations across grouping conditions.

A repeated measures ANOVA showed grouping condition had a significant effect on response time, $F(1.71, 177.59) = 78.20, p < .001$ (since Mauchly’s test indicated a violation of sphericity, $\epsilon = .85$, Huyn-Feldt corrected results are reported). Planned contrasts indicated that response times were significantly higher for the between-group condition compared to the within-

group condition, $t(208) = 11.38, p < .001$, and for the between-group condition compared to the uniform condition, $t(208) = 10.19, p < .001$. There was no significant difference in response times for the uniform condition compared to the within-group condition, $t(208) = 1.19, p = .236$.

The results of this experiment largely replicate the pattern of results in Vickery (2008). Compared with the results reported by Vickery, response times were longer in Experiment 4, and there was a much greater difference between the mean response time for the between-group condition and those of the other conditions. The former difference is likely due to the difference in how participants responded: unlike the original experiment where participants pressed a button as soon as they found the target pair, this experiment required participants to press a key that corresponded to the size of the target pair. The process of determining which key to press in this experiment should take more time and, thus, should lead to higher response times.

Experiment 5: Nine Elements Per Row

There are two reasons for conducting Experiment 5. First, I wanted to see whether a version of Experiment 5 with nine shapes per row, which is the same as the number of elements in Palmer and Beck's experiment, also provides evidence of induced grouping. Second, this provides experimental results that I can directly compare with model results. Given the computational resources I had at the time of simulating this experiment, I could only run simulations using the Vickery stimuli with nine shapes.

Method

Participants. I again planned to gather data so that I measured mean response time with a standard error of around 10 milliseconds. Experiment 4 above found standard deviations across participants to be rather large (approximately 470 ms). Reasoning that this was due to the row being much longer than a version with nine shapes per row, I stipulated that standard deviation for Experiment 5 should be closer to Experiment 1 (approximately 114 ms). So, I again planned for approximately 100 participants. I ended up with a total of 143 participants. As in Experiment 5, I used data only from subjects with an error rate of 10% or less (final $n = 101$).

Apparatus, stimuli, and procedure. All aspects were identical to those of Experiment 4 with the following exceptions. Rather than having 15 shapes per row, each stimulus had 9 shapes per row; examples of stimuli used in the experiment are shown in the bottom three rows of Figure 49. There were two target sizes (small or large), 3 grouping conditions (within-group, uniform, between-group), and 4 possible positions of the target pair (a target element could appear in any location except the two outer positions [1, 2, 8, or 9] in the row). The size of the black window where the experiment took place was the same size as in Experiment 1. And, as in Experiment 1, a response time was considered to be too fast if it was greater than 3000 ms. Response times less than 150 ms or greater than 3000 ms were excluded from analysis.

Results and Discussion

Mean response times for correct trials for Experiment 5 are shown in third set of bars in Figure 50a and in Table 8.

Table 8. Experiment 5 Descriptive Statistics and Correlations for Grouping Conditions

Condition	Response Time (ms)			
	<i>M</i>	<i>SD</i>	Within	Uniform
Within	1016.99	197.45	—	
Uniform	1040.96	189.41	.93	—
Between	1084.84	206.41	.91	.89

Note. *M* indicates mean, *SD* indicates standard deviation, and columns labelled “Within” and “Uniform” indicate correlations across grouping conditions.

As for Experiment 4, I ran an ANOVA with three planned contrasts to compare performance for grouping conditions.

A repeated measures ANOVA showed grouping condition had a significant effect on response time, $F(1.92, 192.25) = 33.63$, $p < .001$ (since Mauchly’s test indicated a violation of sphericity, $\epsilon = .96$, Huyn-Feldt corrected results are reported). All planned contrasts were

significant: response times were significantly higher for the between-group condition compared to the within-group condition, $t(200) = 8.09, p < .001$, and compared to the uniform condition, $t(200) = 5.23, p < .001$. Response times for the uniform condition were also significantly higher than for the within-group condition, $t(200) = 2.86, p = .005$.

Again, the spacing of irrelevant crosses negatively impacted performance in the between-group condition. However, unlike Experiment 4, the crosses in this experiment offered some performative advantage for the within-group condition compared to the uniform condition.

Model Simulations of Experiment 5

Simulated Grouping Strategies

Four grouping strategies were implemented by the model, which are summarized in Figure 51b. Since a grouping strategy is a combination of a connection strategy and selection strategy, I go through these components in the next two subsections.

Connection strategies. A connection strategy that would allow for easier identification of the target pair is to connect only the target pair of circles together. The target pairs were always horizontally aligned, adjacent and of the same size. An implementation of this connection strategy (one that also ignores connections that happen to form among irrelevant crosses) is to have a relatively long Spread Controller duration (57 ms) so that even stimuli with a target pair consisting of small circles can connect. However, as in the grouping by size example discussed in the “Combining Connection Controllers to Promote Gestalt Groupings” section above, this causes connections between different sized circles; this high Spread Controller duration, which is needed to connect the distant but adjacent small circles also causes small and large circles to connect. To eliminate these connections, the Small Controller Circuit was on and the Long Controller Circuit was off. So, just as in the grouping by size example discussed in the “Combining Connection Controllers to Promote Gestalt Groupings” section, the Small Controller Circuit must be on to prevent the top and bottom horizontal edges of each small circle from connecting with the large circles. One set of connection parameters was used to connect the same sized nearby target pairs

for all conditions. The resulting connections for each stimulus are shown in Figure 51c under ‘Strategic Connections’. (The set of parameter values used to produce these connections is given in the caption of Figure 51).

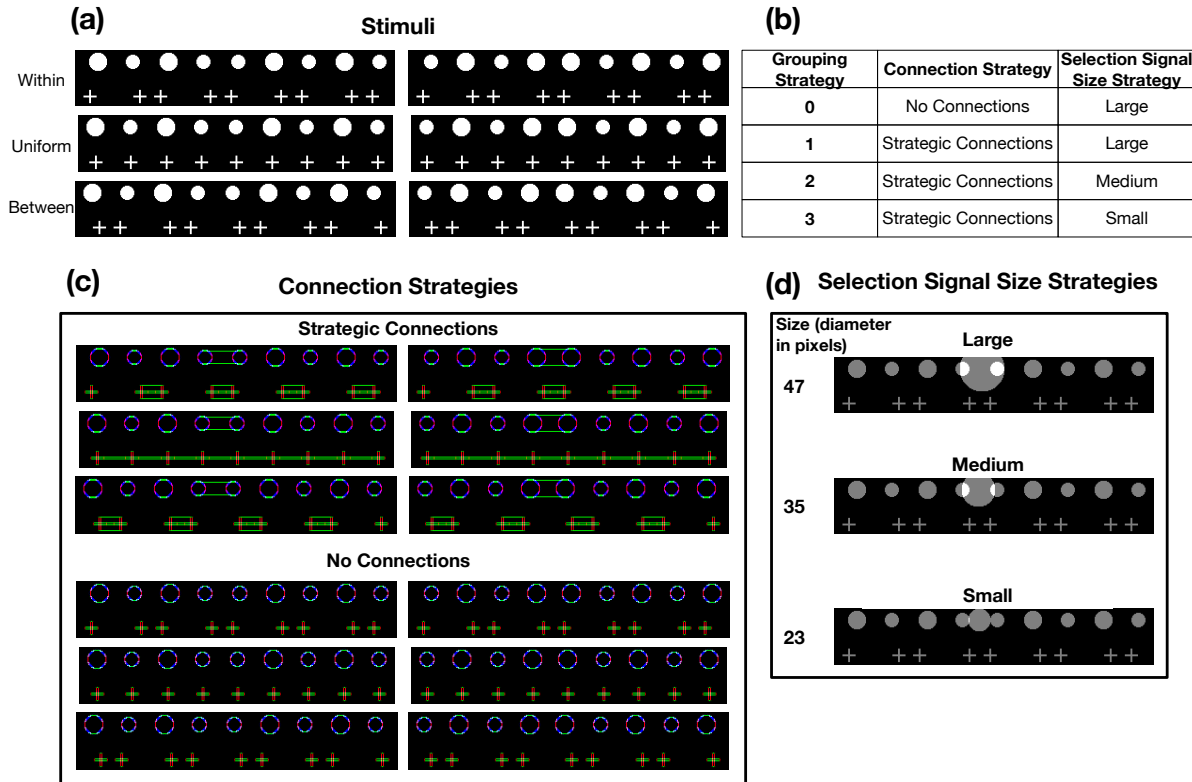


Figure 51. (a) All stimuli input to the model. (b) Grouping strategies implemented by the model, which reference boxes (c) and (d). Box (c) shows connections formed by two connection strategies. For Strategic Connections, one set of connection parameters was used for the horizontal connection circuit: Spread Controller duration was 57 ms, Long Controller input was 1000, and Short Controller input was 0.8. For No Connections, the set of connection parameters was: 0 ms duration, 1000 Long Controller input, and 0 Short Controller input. Spread Controller onset was fixed at 0 ms for all simulations. Box (d) shows the selection signal map for three selection signal size strategies. The map is overlaid with a stimulus to give a sense of a signal’s size relative to stimulus elements.

To evaluate the role of connections in performance on this task, simulations were conducted with Grouping Strategy 1, which had Strategic Connections, and Grouping Strategy 0, which had no connections.

Selection strategies and target identification algorithm. For ease of comparing the role of connections, Grouping Strategies 0 and 1 had the same selection strategy, which is shown in Figure 51d under ‘Large’. This selection signal size was chosen because it was as big as possible without risking overlap with an irrelevant cross. If Grouping Strategy 0 is to succeed, the observer implementing this grouping strategy must group target pairs by selection. So, the signal needs to be big enough to possibly cover a target pair consisting of two small circles, but also not so large that it selects an irrelevant distractor.

Two additional grouping strategies (Grouping Strategies 2 and 3) were run in order to see whether an observer using a smaller selection signal and connections would perform differently than an observer using Grouping Strategy 1 with the large selection signal. Grouping Strategy 2 and Grouping Strategy 3 were the same as Grouping Strategy 1 except for the selection signal size strategy. As shown in Figure 51d, Grouping Strategy 2 used the Medium size strategy, and Grouping Strategy 3 used the Small size strategy.

Apart from selection signal size, all other components of the selection strategy were identical across all grouping strategies implemented. The timing and placement strategies and target identification algorithm are the same as those for the Palmer and Beck simulation with the following exceptions. Possible locations of each selection signal were restricted to the row of circles; a selection signal could not overlap with a cross (Figure 52).

The feature filter probability distribution was constructed from horizontal activity for the entire image, including horizontal signals from crosses (Figure 53).

This distribution was used to guide selection signals, specifically the x-coordinate of the center of a selection signal was randomly sampled from this distribution. The y-coordinate was fixed for all trials so that each selection signal was vertically centered on the row of circles. So, although horizontal signals from crosses influenced selection signal placement across the image, crosses could not be selected.

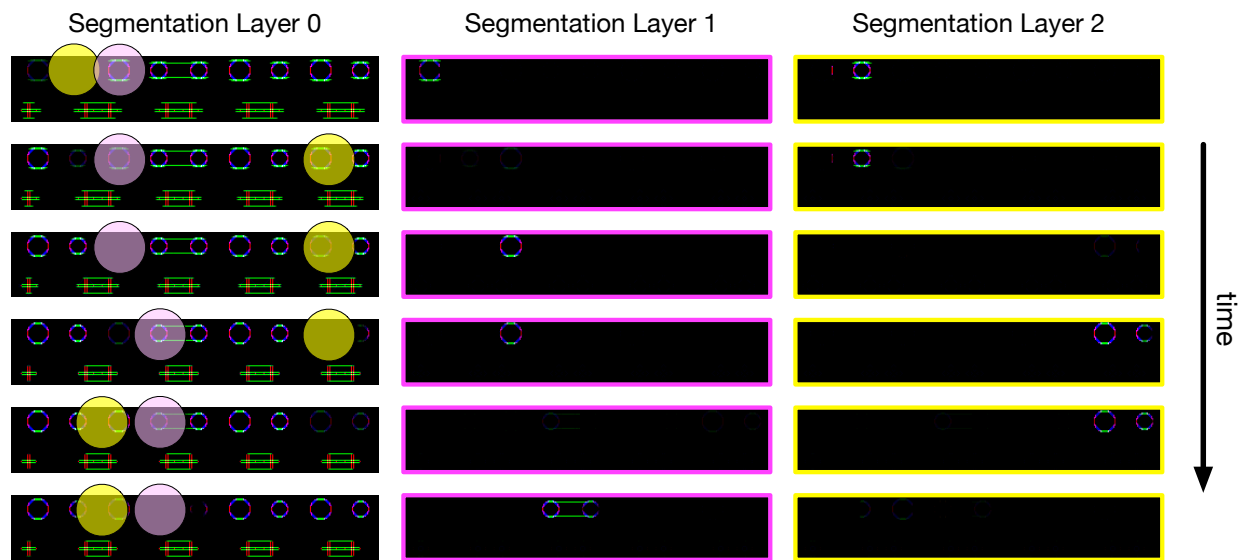


Figure 52. Example of the selection strategy used for the Vickery simulations. The output of each segmentation layer from the final 300 milliseconds (simulation time) from a particular trial, with each row showing segmentation layer activity during 50 ms. The y-coordinate of each selection signal was fixed such that it was centered on the row of circles yet did not overlap with the crosses. Shown here is the Large selection signal size strategy.

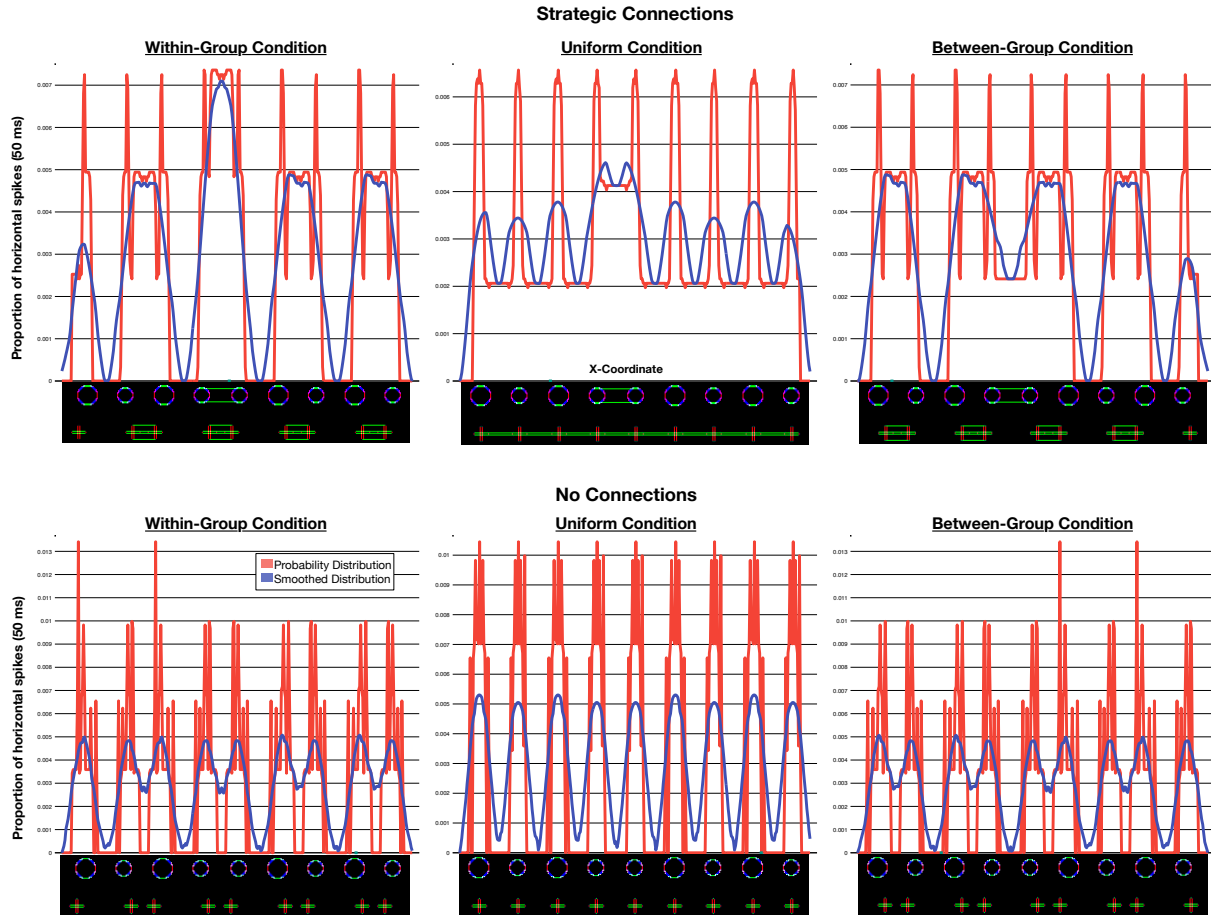


Figure 53. Examples of the distributions used in each condition of the Vickery simulation for both connection strategies implemented by the model.

Simulation Stimuli, Method, and Procedure

Stimuli input to the model are shown in Figure 51a and were created using a custom Python script and converted to BMP files. All aspects of the simulation method and procedure are identical to those of the simulations for Experiment 1 except, rather than a trial timing out at 750 ms, the maximum time of a trial was 1000 ms. In view of the pilot simulation results, the Vickery simulations often took longer to find a target than the Palmer and Beck simulations, which seldom found a target after 750 ms. The reason for this difference is discussed below.

Model Results and Discussion

Experiment 6 indicated that mean response times for all grouping conditions was significantly different with the longest response times for the between-group condition and shortest

for the within-group condition (Figure 50a, right set of bars). This pattern was mirrored qualitatively in the results of the simulations with Grouping Strategy 1 (Figure 50b, second set of bars). Grouping Strategies 2 and 3 also produced a similar but compressed pattern of results (Figure 50b, third and fourth set of bars). Grouping Strategies 1, 2, and 3, were well correlated with the Experiment 5 data ($r \approx 1.00$, $r = .99$, and $r = .98$, respectively) compared to Grouping Strategy 0 ($r = .74$). Of the grouping strategies simulated, Grouping Strategy 0 produced the highest response times and, thus, the worst performance (Figure 50b, first set of bars). Additionally, unlike Experiment 5, Grouping Strategy 0 had the highest mean response time for the uniform condition. This is because selection signals tended to be located directly on one of the circles due to the feature filter (see Figure 53, bottom row). Since Grouping Strategy 0 had no connections, such selection signal placement often resulted in only one circle, rather than the target pair, being selected and segmented. Thus, it seems that connections are required to facilitate performance on this task; grouping by selection alone is not sufficient.

The results of using Grouping Strategies 1, 2, and 3 (Figure 50b) indicate that performance is similar for a range of selection signal sizes provided that the target pair is connected. Thus, it seems that human observers are likely using a grouping strategy in which same-sized adjacent circles are connected along the target row, rather than the connection strategy implemented in Grouping Strategy 0 simulation in which none of the shapes are connected.

It is informative to compare in more detail the results of Grouping Strategy 1 (large selection signals; Figure 50b, second set of bars) with that of Grouping Strategy 3 (small selection signals; Figure 50b, fourth set of bars). Grouping Strategy 1 led to a lower mean response time for the within-group condition and a higher mean response time for the between-group condition. These differences arise from the selection signal size. In the within-group condition, the selection signal tends to be centered on or between the target pair since there is more horizontal signal in that region from the connections between the target pair and from the connected crosses below them (see the feature filter in Figure 53, top left plot). So, for within-group stimuli, a large selection signal can quickly segment and isolate the target pair and is unlikely to overlap with neighboring distractors. In contrast, a small selection signal, although also likely to be drawn to the location of the target pair, requires more time for it to segment and isolate the target pair. Due to the extra time required, a small selection signal that lands on a target may not have segmented enough of the target pair for identification before the next reset occurs. Thus, as a comparison of

the results of Grouping Strategies 1 and 3 demonstrates (Figure 50b), the use of a small signal comes at the cost of longer response times for the within-group condition. However, this small size strategy is beneficial for performance on the between-group condition. Compared with a large signal, a small selection signal is unable to overlap with neighboring distractors. In the between-group condition, the placement of the crosses results in a higher likelihood that a selection signal will be placed towards the gap between a target element and a distractor (see the feature filter in Figure 53, top right plot). So, there is a cost in performance on the between-group condition if using large selection signals since the signal is more likely to fall at a location where it overlaps with a neighboring distractor, but small selection signals, which can be placed more precisely, eliminate this cost and, thus, lead to a lower response time. The application of the model to this task provides a demonstration of the costs and benefits associated with using different selection signal sizes, which were identified in the “Selection Signal Size” section above.

In sum, although a somewhat different pattern of results was produced for each grouping condition by Grouping Strategies 1-3, which only differed in their selection signal size strategy, the general pattern is similar to that of Experiment 5 and the overall performance was very similar, i.e., the mean response time across all grouping conditions for Grouping Strategies 1-3 were 648 ms, 637 ms, and 648 ms, respectively. So, the connection strategy implemented by Grouping Strategies 1, 2, and 3, allows flexibility with regard to the selection signal size strategy.

If, as Vickery claims, reaction time in this experiment is a measure of induced grouping, then the model can be used to explain this phenomenon. For this task, stimuli, and simulated grouping strategies, induced grouping is caused by the selection signal being guided by horizontal contours from the entire scene (see the feature filters in Figure 53). Although the Connection Circuit is tuned to connect the target pairs, this strategy also causes horizontal connections between nearby crosses. The connections between crosses promote performance in the within-group condition, but hinder performance in the between-group condition due to how those signals influence the horizontal feature filter distribution. The model also indicates why performance for the uniform condition is similar to that of the within-group condition: with this connection strategy, all of the crosses form a single group, which neither increases nor reduces the likelihood of a selection signal falling on a target. Instead, the selection signal is guided effectively only by the horizontal signal in the row of circles. So, since there is more horizontal signal near targets in the row of circles due to their V2 connections, it is more likely that targets will be segmented out in

the uniform condition than in the between-group condition. Overall, there is more column-wise horizontal signal in both the within-group condition and the uniform condition near the target pair, which causes a much higher peak in the smoothed probability distributions for these conditions than for the between-group condition (Figure 53, top row).

In turn, the similar distributions that arise from choosing this grouping strategy are partially responsible for similar performance for the within-group and uniform conditions. Thus, induced grouping observed in this experiment is the result of an observer using a particular connection strategy and selection strategy with a feature filter that uses horizontal signals. Since the feature filter does not discriminate between horizontal signals from the circles and crosses and because the connection strategy used to identify target circle pairs also joins nearby crosses, horizontal signals from the edges of crosses and their V2 connections partially guide the selection signal toward or away from potential targets. Thus, the irrelevant horizontal signals from crosses that arise from a grouping strategy that generally promotes performance for this task and stimulus set, i.e., a grouping strategy that is designed to promote performance for two-thirds of the stimuli, can promote or hinder performance in the manner reported by Vickery.

Finally, comparing the results of the Palmer and Beck simulation with that of the Vickery simulation, mean response times tended to be higher for Vickery even though both sets of stimuli had nine elements per row. This mirrors the results from Experiments 1 and 5: mean response times for Experiment 5 (the Vickery replication with nine shapes) tended to be higher than for Experiment 1 (the Palmer and Beck replication). The reason for this difference in the simulations is as follows. In the Vickery simulations, the Spread Controller duration needed to connect the small circle was longer than the durations used for the majority of the Palmer and Beck stimuli. A longer duration causes more spreading, and it takes longer for the system to settle down to equilibrium. This causes there to be: a longer time when the spread from distractor circles may attract a selection signal to a non-target location, and a longer time period when selection signals that land on a target also happen to cover the retracting spread from a distractor and thereby select a distractor along with a target. Thus, the model indicates that the relatively high response time of Experiment 5 compared to Experiment 1 appears to be a result of the connection strategy used to promote performance on the stimulus sets.

TRICK AND ENNS (1997)

Trick and Enns (1997) attempt to tease apart two processes involved in grouping discussed by Koffka (1935/1963) that they argue have been conflated in recent literature. These processes are element clustering, which determines what elements belong together, and shape formation, which determines how grouped elements appear as a whole. They use enumeration tasks to argue that: these two processes are separable, and that clustering is preattentive while perceiving the shape of disconnected elements requires attention. In each trial, 1 to 8 targets are shown. These targets are always diamonds, but in one condition they are line figures and in another they are comprised of four dots with one dot at each corner of the diamond. In their experiment 2, there could be 0, 4, or 8 distractors, which were square line figures or square dot forms. Figure 54 shows the various stimuli used in a replication study, described below.

The task is to identify the number of diamonds. Participants were to press a key as soon as they have a target count, and then enter the number of targets. Line figures and dot forms were presented in separate blocks.

Trick and Enns compared the speed of responses when subjects may be able to subitize items, i.e., for conditions with 1 to 3 targets, and when subjects must count items, i.e., for conditions with 5 to 7 targets. Since enumeration is sensitive to attentional demands, they reason that subjects should be very fast and accurate for a small number of targets via subitizing if the targets can be detected preattentively. In turn, a plot of the mean reaction time against target number should have a relatively flat slope for items 1 to 3. If targets cannot be detected preattentively, then these slopes should be similar to those conditions in which counting is required, e.g., the slopes for conditions with 5 to 7 items.

Results from their experiment 2 are shown in the left column of Figure 55.

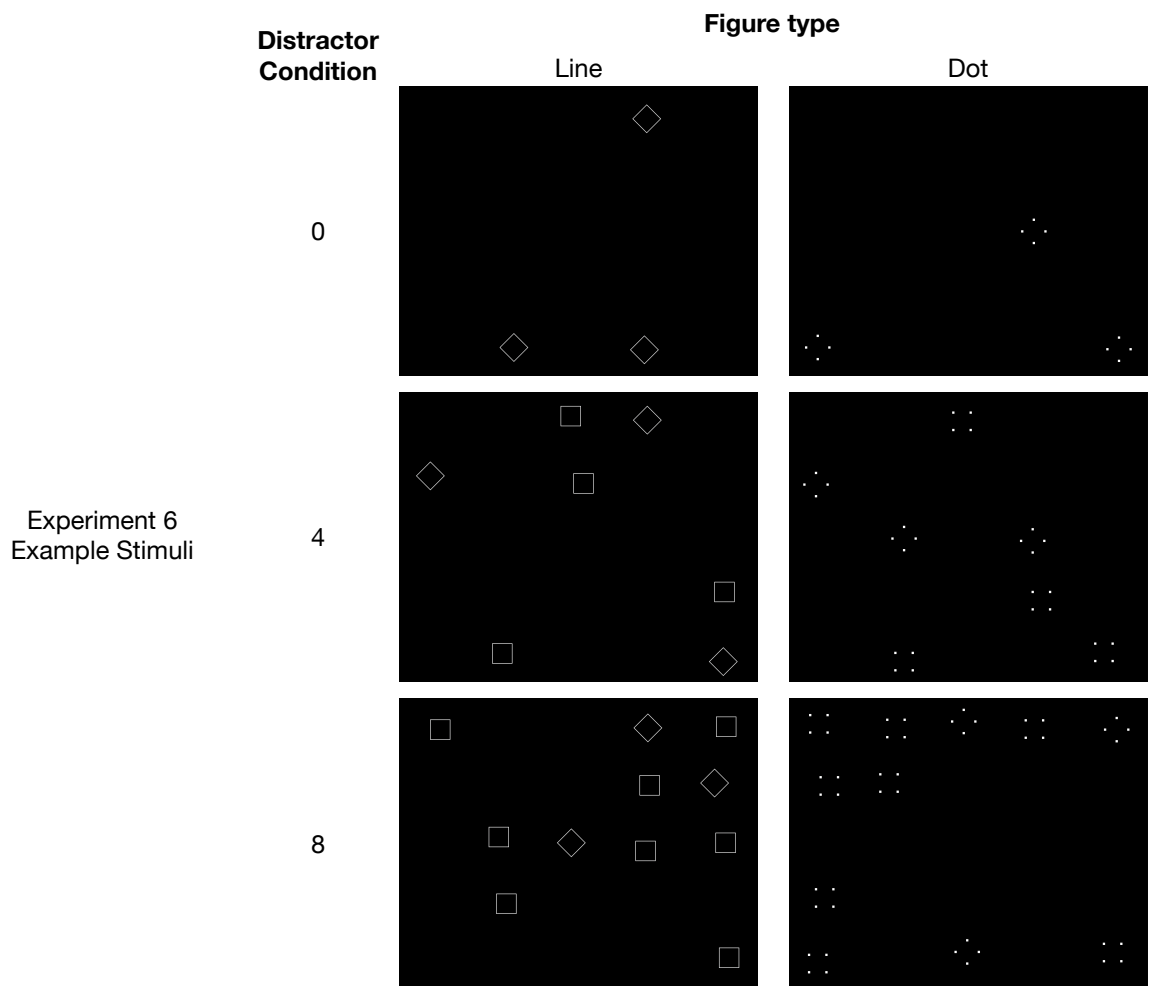


Figure 54. Screenshots of stimuli from Experiment 6 for the three distractor conditions and two figure conditions. Each stimulus shown here has three target diamonds.

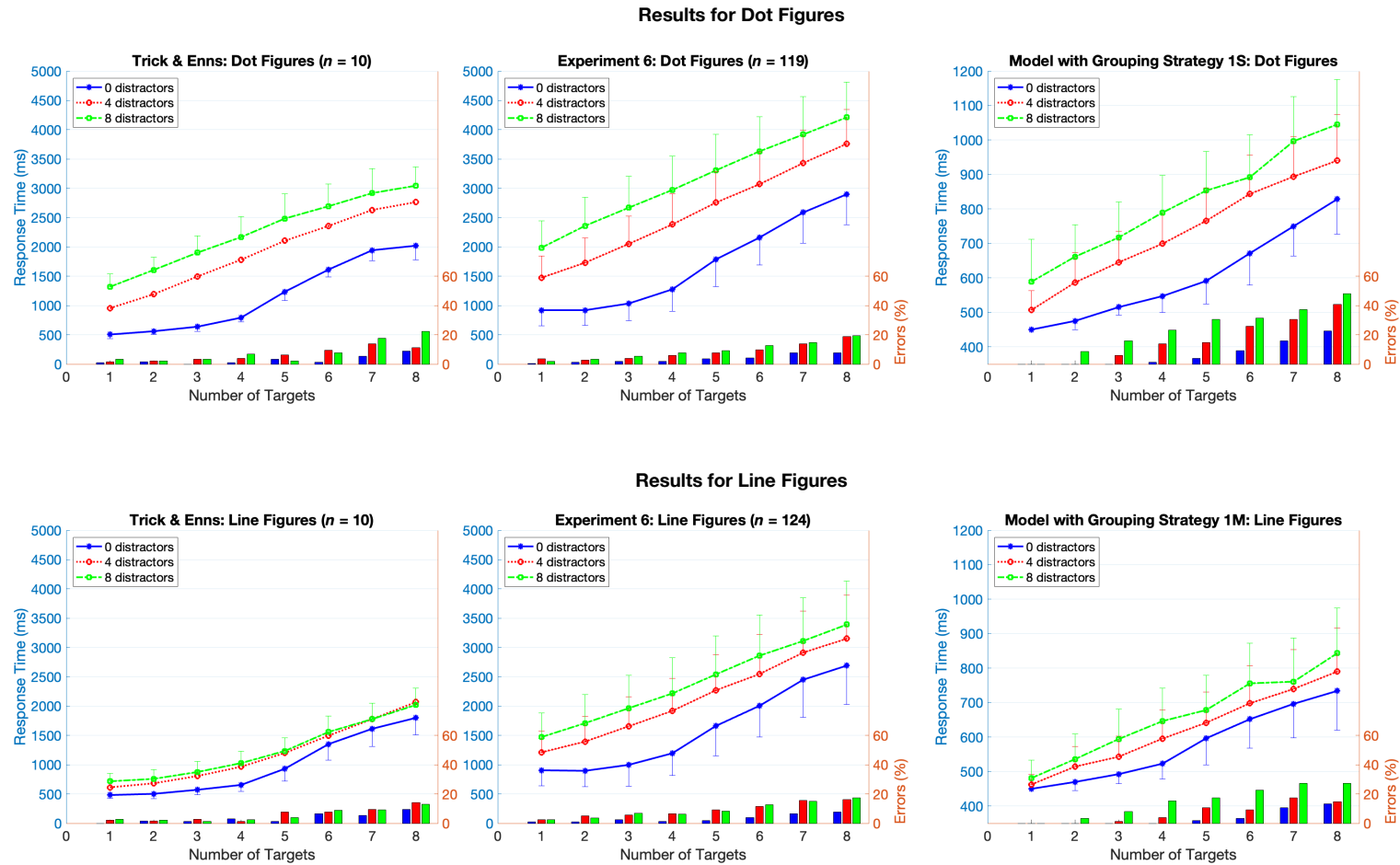


Figure 55. Top row shows results for the dot figure condition from the original experiment (with a single group of 10 participants seeing each figure condition in separate blocks), the replication, and the simulation when the model implements Grouping Strategy 1S. Bottom row provides results for line figure conditions.

Distractor condition is indicated by the color, and target condition is indicated by the x-axis. Bars represent percent incorrect, and points represent mean response times, which were calculated from correct trials only. Following Trick and Enns, error bars indicate standard deviation. For the simulations, 150 trials were simulated for each target-distractor condition.

Trick and Enns claim that all distractor conditions for line figures (Figure 55, bottom left plot) have a relatively flat slope when there are 1-3 targets, which is indicative that subitizing occurred, i.e., a preattentive process that allows for fast, accurate enumeration of a few items. In contrast, for the dot figures (Figure 55, top left plot) only the no distractor condition has a slope indicative of subitizing. From these results, Trick and Enns concluded that there are two separable grouping processes. Since, they claim, the element clustering process is preattentive, it can co-occur with subitizing, while the shape formation process, which requires attention, cannot. The preattentive element clustering process determines what elements belong together, e.g., that the four dots comprising a diamond are grouped together but are not grouped with a dot from a neighboring dot figure. The shape formation process determines how clustered elements appear as a whole. For example, it is needed to distinguish the dot diamonds from the dot squares when performing the task. Trick and Enns reason that since people can subitize in the dot condition when there are no distractors, only preattentive element clustering was involved. When there are distractors, shape formation is also needed to distinguish targets from distractors. Because this process requires attention, people can no longer subitize. So, the dot figure results suggest that there are two sorts of grouping processes at play: clustering only for conditions with no distractors, and likely both clustering and shape formation are required when there are distractors.

Experiment 6: Replication of Trick and Enns (1997)

Before modeling the findings of Trick and Enns (1997), I wanted to be confident in the experimental results, especially since the original study used a very small sample ($n = 10$). Experiment 6 is a replication of experiment 2 in Trick and Enns (1997). This replication deviates from the original in the following ways. First, the replication was an online experiment, which involved making the following modifications. Unlike the first experiment, in which all participants saw both figure conditions (dot or line) in different blocks, participants in Experiment 6 were randomly assigned to a group that only saw line figures or dot figures. This was partly due to wanting to keep the total session time to less than 30 minutes. Since the original experiment was blocked, this deviation shouldn't matter for present purposes. Another modification is that, due to the time participants in the dot figure condition tended to take to complete a session, the replication experiment had 8 trials per target-distractor condition per participant so that a session could fit within 30 minutes, while the original had 10. And, although the original experiment had

participants use the space bar to indicate that they had a count of the targets, I used the Y-key since the space bar has other default functions in some browsers, e.g., scrolling.

Second, feedback was given as the words ‘Correct’ and ‘Incorrect’ on the screen, instead of a plus sign and minus sign. Third, unlike the original experiment that instructed participants to use keys 0-8 to make a response in the second phase, Experiment 6 instructed participants to use the keys 0-9. Finally, for reasons related to model input, rather than using black stimuli on a white background, the replication stimuli had white figures on a black background.

Method

Participants. My general aim is to measure response time with a precision that would have a standard error of 10 milliseconds. I am most concerned with getting a precise measure of response times when there are 1 to 3 targets since the response time for these target conditions is regarded as being key to determining whether subitizing was involved. The only measure of variability provided by Trick and Enns were error bars in a plot. So, I used the error bars in the third figure of Trick and Enns (1997) to estimate that mean response times for line figure conditions with 1 to 3 targets had a standard deviation of around 117 ms (for the zero distractor condition only, standard deviation was approximately 76 ms), and those for dot figures was approximately 154 ms (for the zero distractor condition only, standard deviation was approximately 68 ms). So, if these estimates are correct, a sample size of 100 participants per group should suffice for measuring response times for conditions with 1 to 3 targets with a standard error of around 10 ms. There were a total of 124 participants in the dot group and 128 in the line group. Any participant with an error rate of 100% for at least one target-distractor condition was eliminated from analysis, which left $n = 119$ in the dot group and $n = 124$ in the line group.

To get a better measure of variation than my estimate of the length of error bars on somewhat distorted plots, I looked for other studies with a similar task and stimulus set. Enumeration studies in Trick and Pylyshyn (1993; 1994a,b) do not provide any measures of variability to compare these values with, and there are very few comparable enumeration studies with distractor conditions. Nevertheless, for an enumeration task conducted by Nan et al. (2006) where participants indicated whether the number of target rectangles, which could be among circle distractors, was even or odd, the results indicate that standard deviation was approximately 114

ms for stimuli with 1-3 targets, which is in line with the results reported by Trick and Enns (1997) and provides some support for the choice of sample size.

Apparatus. Participants used the Y-key on their keyboard to indicate that they had counted the target diamonds, and then used the number keys 0-9 to indicate their count. Participants were instructed to rest an index finger on the Y-key for the duration of the task.

Stimuli. Example stimuli used in the experiment are shown in Figure 54. There were two figure conditions (line or dot), eight target conditions (each image contained 1-8 diamond targets), and 3 distractor conditions (each image had 0, 4, or 8 distractor squares). The location of each element was randomized on each trial as follows. There were 25 possible locations for a given stimulus element as the black rectangle was divided into a 5x5 grid (this is a notational grid used in programming the experiment; the black rectangle was not visually divided into a grid). On a given trial, each stimulus element was randomly assigned to the center of one of these grid cells, and prior to adding random jitter, the figure's center was the center of its assigned grid cell. Random jitter was added to the x-coordinate and y-coordinate by drawing a sample from a uniform distribution over the range $\pm 0.03^\circ$. This random jitter was sampled independently for the x-coordinate and y-coordinate of each figure.

For the hypothetical participant described in the “General Experiment Methodology” section, the width of a diamond line figure subtended approximately 1.3° of the visual angle, and the width of a square line figure or dot figure (which was a rotated diamond) subtended approximately 0.9° . The width of the diamond dot figure subtended 1.1° (which is slightly less than the tip-to-tip width of the diamond line figure since the dots do not have left/right tips; the dot figures were designed such that they would fit tightly inside the line figures if they were superimposed). Each dot subtended 0.1° . The black rectangle in which the experiment took place had a constant width and height that subtended approximately 15.5° and 12.4° , respectively, and each cell of the notational grid was approximately 3.1° wide and 2.5° high.

Procedure. Participants were randomly assigned either to the line figure condition or the dot figure condition. After reading instructions that explained the task, participants scrolled to the bottom of the webpage, where the experiment took place. Figure 56 schematizes what occurred on a trial. A participant initiated a trial by pressing the N-key on their keyboard. A fixation cross appeared for 675 milliseconds after this key was pressed. Next the stimulus array was shown until the participant pressed the Y-key as quickly as possible once they had determined the number of targets. After this key press, the stimulus was replaced by a prompt to enter the number of targets, which participants were encouraged to do as accurately as possible. Feedback indicating whether the number entered was correct or not was shown for 500 ms before the prompt for initiating the next trial appeared.

After completing 24 practice trials in which all distractor and target conditions were experienced, the participant completed a total of 192 experimental trials. The experimental trials were presented in two blocks, and an onscreen notification after the practice trials and after block 1 encouraged participants to take a short break if needed. All distractor and target conditions were randomly interleaved in each block. If their response when pressing the Y-key was too fast (if the response time was less than 100 ms), or too slow (if the response time was greater than 6000 ms), participants were given additional feedback, which would appear at the same time as the correct/incorrect feedback. Incorrect trials and trials with reaction times lower than 100 ms or greater than 6000 ms were omitted from the analysis of response time.

Results and Discussion

Mean response times for correct trials and error rates for Experiment 1 are shown in the second column of Figure 55 and in Table 9.

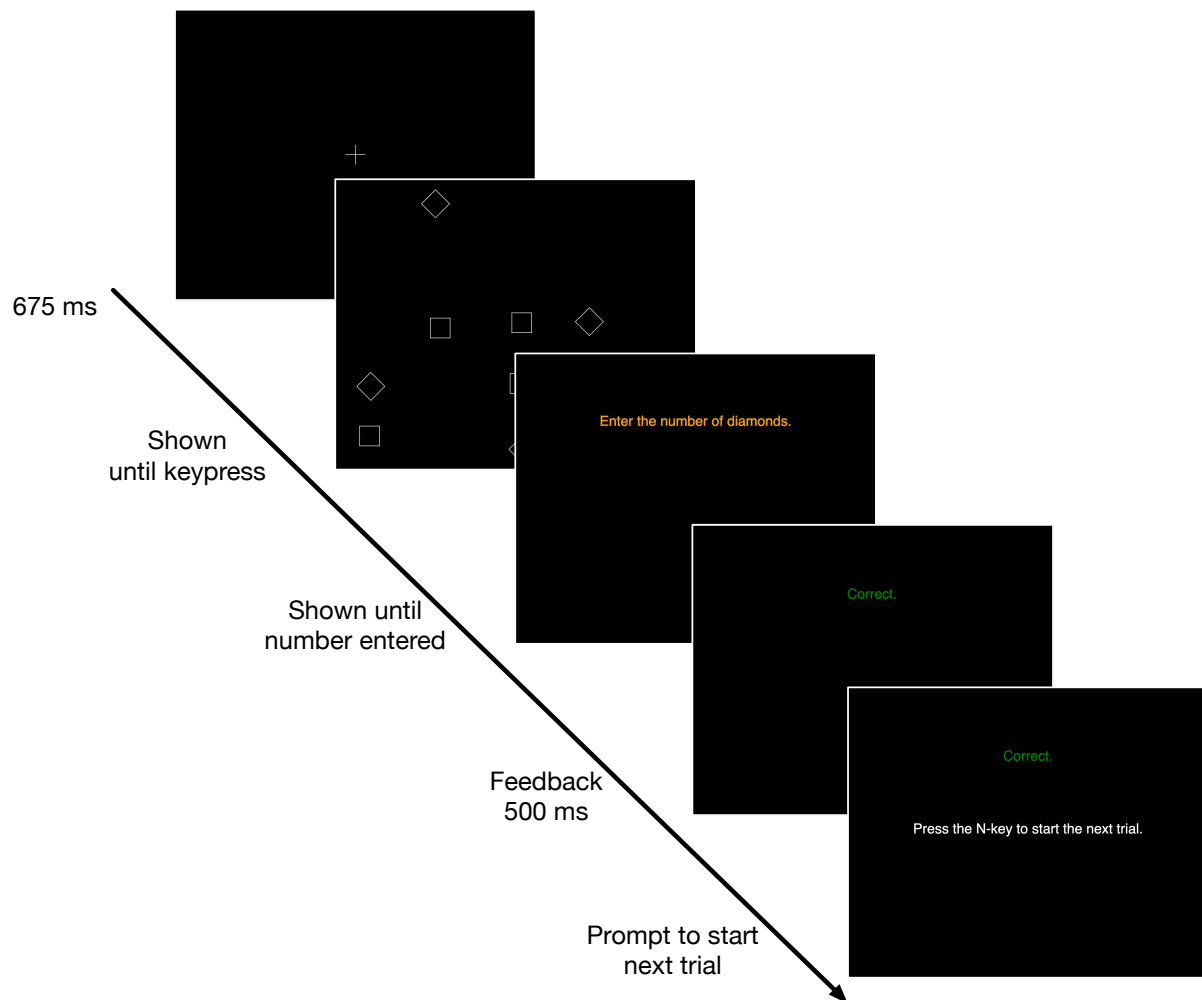


Figure 56. The time course for a trial of Experiment 6. For incorrect trials, the green ‘Correct’ was replaced with a red ‘Incorrect’.

Table 9. Experiment 6 Descriptive Statistics

		Dot Figures				Line Figures			
		Response Time (ms)		Percent Incorrect		Response Time (ms)		Percent Incorrect	
Targets		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
0 Distractors	1	920.61	265.54	0.53	2.52	905.70	269.01	1.11	4.22
	2	920.24	258.17	1.26	4.12	896.21	270.31	0.91	3.26
	3	1034.26	295.44	1.80	4.43	997.47	365.50	2.55	6.08
	4	1275.93	374.68	1.91	5.09	1195.27	377.53	1.21	3.71
	5	1787.34	464.24	3.69	7.71	1662.81	508.96	1.92	5.05
	6	2161.71	472.63	4.20	8.18	2006.41	533.60	3.83	8.16
	7	2587.79	522.98	7.68	10.19	2451.33	640.69	6.71	9.83
	8	2899.73	519.32	7.79	10.92	2692.32	665.49	7.82	11.73
4 Distractors	1	1474.79	373.17	3.57	6.74	1211.75	363.11	2.53	6.19
	2	1732.32	420.83	2.85	5.99	1395.34	428.26	5.07	9.08
	3	2052.29	471.09	3.89	7.76	1656.71	505.37	5.65	9.34
	4	2383.16	520.19	6.09	10.53	1919.16	553.84	6.60	9.70
	5	2756.13	526.82	7.82	10.95	2271.60	601.22	9.29	12.81
	6	3070.56	551.67	9.85	11.68	2548.38	675.24	11.66	14.94
	7	3432.61	557.72	14.03	15.07	2914.21	707.12	15.55	16.63
	8	3760.21	587.14	18.86	19.02	3150.75	744.41	16.17	19.27
8 Distractors	1	1987.23	455.75	1.91	5.58	1471.50	408.06	2.53	5.77
	2	2360.73	487.91	3.47	7.44	1709.39	489.05	3.73	7.12
	3	2668.87	538.81	5.52	9.05	1964.53	561.03	6.77	10.27
	4	2972.58	577.92	7.62	11.48	2215.78	607.16	6.18	10.53
	5	3307.12	612.61	9.24	12.95	2542.33	657.47	8.21	11.73
	6	3632.42	592.53	12.77	14.50	2862.37	695.09	12.82	16.65
	7	3918.42	643.14	14.66	15.44	3112.79	733.54	14.99	17.46
	8	4212.07	603.43	19.38	16.63	3392.19	741.47	17.49	18.95

Note. *M* represents mean, *SD* indicates standard deviation. The top third of the table provides statistics for conditions with no distractors and 1-8 targets, and likewise for the middle and bottom thirds. Correlations are not presented here for each target-distractor condition. Correlations for dot figure response times ranged from .45-.90, and for line figures correlations ranged from .59-.90 (no including values along the primary diagonal).

Apart from comparing the slope of each curve from 1-3 targets with the slope from 5-7 targets, it is not entirely clear what statistical analyses Trick and Enns performed on their data. To approximate something like their analysis for the reaction times from Experiment 6, I used a regression model with a two-way interaction (i.e., number of targets X whether the target number is in the range of a subitizable set (1-3 targets) or countable set (5-7 targets)), where reaction time was the dependent variable, number of targets was the independent variable, and whether the target number is in the subitizable set range was the moderating variable. This analysis was applied separately to each curve, and, following Trick and Enns, reaction times for conditions with 4 or 8 targets was dropped. If the interaction is significant, then the slope of the response time curve for stimuli with 1-3 targets significantly differs from that with 5-7 targets. Unlike Trick and Enns, I did not aggregate the distractor conditions since I do not have any theoretically based reason to believe that each distractor condition should be similar.

Table 10 provides the slopes of the curves when there are 1-3 targets and 5-7 targets, as well as the mean response times and their standard deviations for these conditions.

For zero distractors, the difference in slopes for the two sets of targets was significant for dot figures, $t(5516) = 17.96, p < .001$, and line figures, $t(5768) = 17.97, p < .001$. For 4 distractors, the difference in slopes was also significant for both dot figures, $t(5280) = 2.06, p = .039$, and line figures, $t(5431) = 4.85, p < .001$. For 8 distractors, the difference in slopes was not significant for the dot figures, $t(5196) = -1.70, p = .089$, but it was significant for the line figures, $t(5429) = 2.08, p = .038$.

The results of the statistical analysis for line figures mirrors that of Trick and Enns who found that there was evidence of subitizing in all line conditions. However, the results for dot figures only partially echoes Trick and Enns' findings since they did not report that there was significant difference in slope for the 4 distractor condition.

Table 10. Experiment 6 Slopes and Descriptive Statistics by Curve and Target Range

Dot Figures						
Distractors	Conditions with 1-3 Targets			Conditions with 5-7 Targets		
	<u>Response Time (ms)</u>			<u>Response Time (ms)</u>		
	<i>b</i>	<i>M</i>	<i>SD</i>	<i>b</i>	<i>M</i>	<i>SD</i>
0	58.27	957.46	401.60	398.70	2169.73	782.40
4	288.80	1751.84	645.35	336.10	3068.69	800.37
8	339.50	2333.89	718.08	296.60	3597.68	851.29
Line Figures						
Distractors	Conditions with 1-3 Targets			Conditions with 5-7 Targets		
	<u>Response Time (ms)</u>			<u>Response Time (ms)</u>		
	<i>b</i>	<i>M</i>	<i>SD</i>	<i>b</i>	<i>M</i>	<i>SD</i>
0	42.28	930.19	406.17	394.10	905.70	269.01
4	223.40	1418.08	613.77	338.60	2573.95	876.23
8	242.70	1711.52	677.89	295.40	2842.03	903.87

Note. *b* indicates the slope of the curve across conditions with 1-3 targets or 5-7 targets., *M* represents the mean response time for these conditions, and *SD* indicates standard deviation.

There is an obvious disconnect between visual inspection of slopes in the plots (Figure 55) for Experiment 6, which indicates only a large difference in slope for the zero distractor conditions, and this analysis, which indicates that each curve has a significant difference in slope except for the dot figure condition with 8 distractors. Although the zero distractor conditions produced curves with a flat slope indicative of subitizing if visually compared to the slopes of the other distractor conditions, it is not unsurprising that the statistical analysis did not distinguish these curves from the others. First, given the large sample size, small differences in slope can become significant. Second, Balakrishnan and Ashby's (1991, 1992) work on statistically detecting a difference in enumeration task response times indicated that there is no benchmark statistical measure that can be used to reliably distinguish response times for conditions with a count that is in the subitizing

range from other conditions. Instead, they suggest that subitizing is, rather than some distinct pre-attentive process, a phenomenon reflective of the limited capacity of visual attention. If this is true, then the model may be able to provide some account of these results in terms of connections, selection and segmentation, and without needing a separate subitizing process. I will address what the results of Experiment 6 mean for Trick and Enns' claim that there is a separable clustering process involved in grouping after presenting the simulations of this experiment.

Qualitatively, the pattern of results reported by Trick and Enns replicates for dot figures: only the curve for the zero distractor condition has a shape indicative of subitizing with a relatively flat slope at the beginning (Figure 55, bottom right plot). The Experiment 6 line figure results (Figure 55, bottom middle plot) nearly have the same pattern as the dot figure results (Figure 55, top middle plot), but with generally lower mean response times for the 4 and 8 distractor conditions. This does not match the line figure results reported by Trick and Enns, which has nearly overlapping curves for the non-zero distractor conditions that have a somewhat shallower slope at the beginning (Figure 55, bottom left plot). Thus, the results reported by Trick and Enns for 4 and 8 distractor conditions with line figures do not replicate qualitatively. Given that an observer will have to search through more items in the non-zero distractor conditions to find targets, which will take time, the Experiment 6 line figure results for non-zero distractor conditions with a linearly increasing slope as target items are added seems more plausible than those reported by Trick and Enns.

Finally, note that Experiment 6 had larger standard deviations than those reported by Trick and Enns. Supposing the measure of variability reported in Trick and Enns (1997) is accurate, this could be due to differences in equipment used, the possibility that their subjects were not naïve, and/or the total number of trials per participant. Assuming that Trick and Enns' subjects for their experiment 2 did not also participate in their first experiment, which had the same task (it is not made clear in Trick & Enns, 1997, whether a single group of 10 participants took both experiments or not), they experienced twice as many blocks since they saw both figure conditions, compared to participants in Experiment 6. So, Trick and Enns' participants may have a lower standard deviation from more practice with the task, even though the number of trials per target-distractor condition for each figure condition is similar (8 in Experiment 6, and 10 in the original experiment).

Model Simulations of Experiment 6

Simulated Grouping Strategies

Since stimuli with line figures and dot figures were presented in alternating blocks in the original experiment and to two different sets of participants in Experiment 6, it seems reasonable to assume that observers were using one grouping strategy for line figures and another grouping strategy for dot figures in order to promote performance for each stimulus type.

For dot figures one grouping strategy was simulated, which is Grouping Strategy 1S summarized in Figure 57b.

For line figures four grouping strategies were simulated as indicated in Figure 58b.

This difference in the number of explored grouping strategies was because a grouping strategy that would promote performance for line figures was not as obvious as for the dot figures. In the next two subsections, I present the components of the grouping strategies implemented by the model.

Connection strategies for dot figures. For the dot figures, it seems that it would be beneficial to connect at least the dots of the diamond figures. Such intra-shape connections would allow a selection signal to segment all of the dots constituting a diamond dot figure if the selection signal covers at least part of the cluster. This can be accomplished by turning on the Spread Controller for diagonally-tuned V2 cells. However, this also leads to diagonal connections between the dots of each square dot figure, which results in an X-shape (Figure 59).

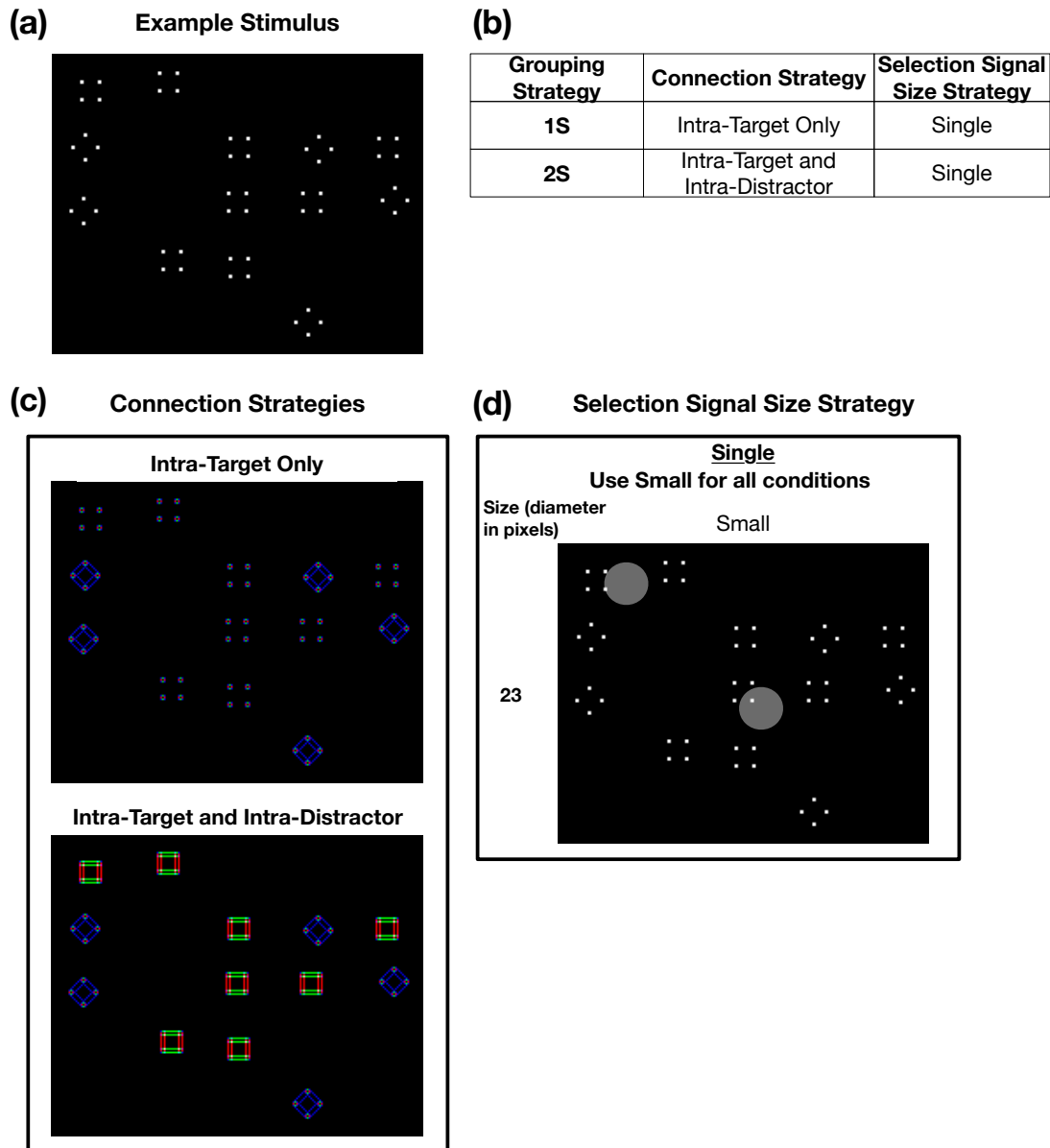


Figure 57. Grouping strategies for the dot figure condition. (a) shows an example stimulus image generated on a particular trial and input to the model. (b) provides a summary of two grouping strategies with reference to boxes (c) and (d). Only Grouping Strategy 1S was simulated for dot figures since pilot simulations indicated that they led to similar performance. Box (c) shows two connection strategies. For the Intra-Target Only strategy, the connection parameters for diagonally-tuned cells were: 1.5 ms Spread Controller duration, 50 ms Spread Controller onset, 1.0 Long Controller input, and 0.0 Short Controller input. The horizontal and vertical connection circuits were turned off. For the Intra-Target and Intra-Distractor strategy, the parameters for the diagonal connection circuit were the same as for the Intra-Target Only strategy. For the horizontal and vertical connect circuits, the parameters were: 1.0 ms Spread Controller duration, 50 ms Spread Controller onset, 1000.0 Long Controller input, and 0.0 Short Controller input. Box (d) illustrates the selection signal size strategy by overlaying a selection signal map with the stimulus to give some sense of the signal size relative to that of the stimulus elements.

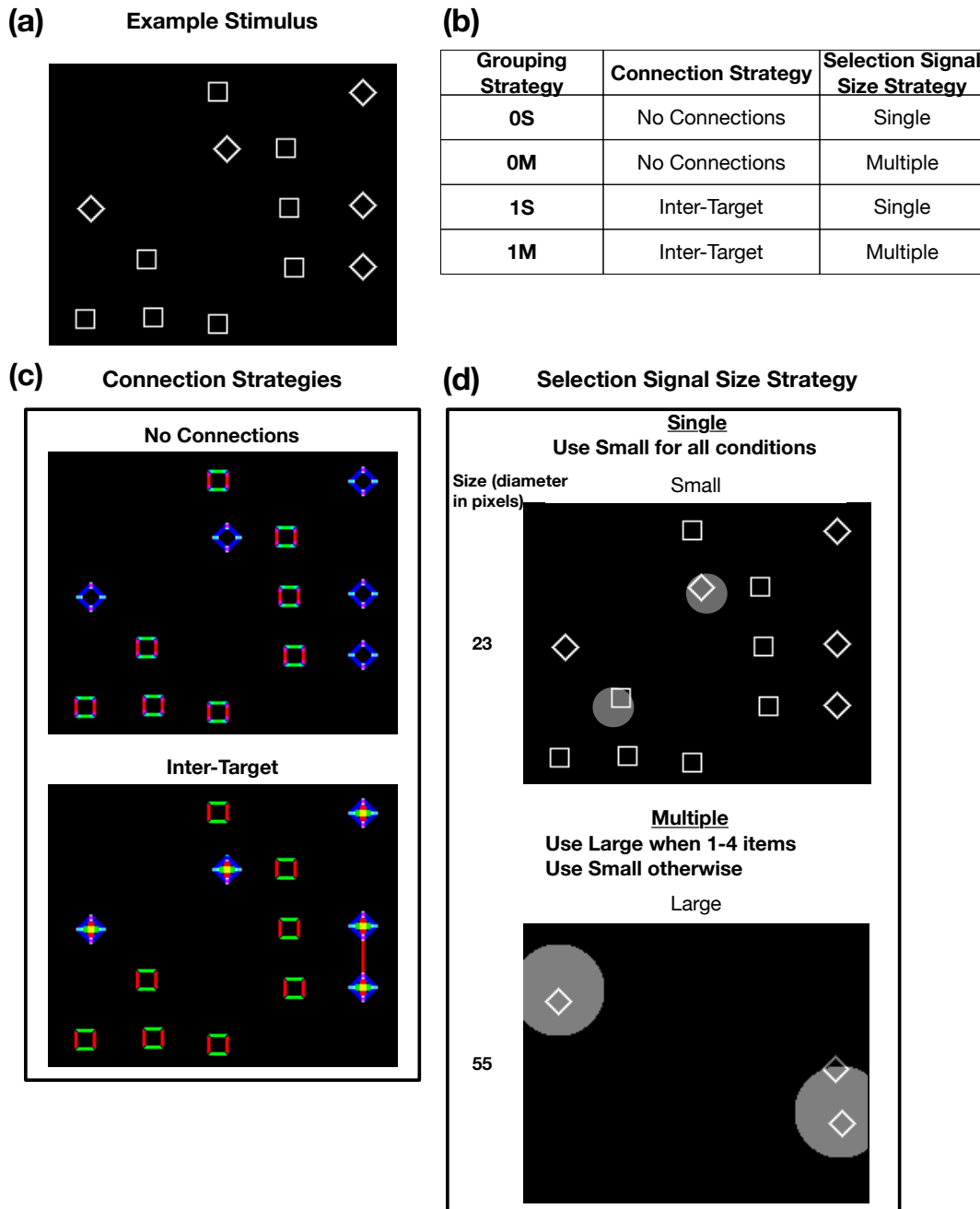


Figure 58. Grouping strategies for the line figure condition. (a) shows an example stimulus image generated on a particular trial and input to the model. (b) provides a summary of two grouping strategies with reference to boxes (c) and (d). Box (c) shows two connection strategies. For the No Connections strategy, all connection circuits were turned off with Spread Controller duration at 0 ms, Long Controller input at 1000.0, and Short Controller input at 0.0. For the Inter-Target strategy, the parameters for the diagonal connection circuit were: 50 ms Spread Controller duration with onset at 50 ms, 1000.0 Long Controller input, and 0.8 Short Controller input. For the horizontal and vertical connection circuits, the parameters were: 50 ms Spread Controller duration with onset at 50 ms, 0.1 Long Controller input, and 0.0 Short Controller input. Box (d) illustrates the selection signal size strategies by overlaying a selection signal map with the stimulus to give some sense of selection signal size relative to the stimulus elements.

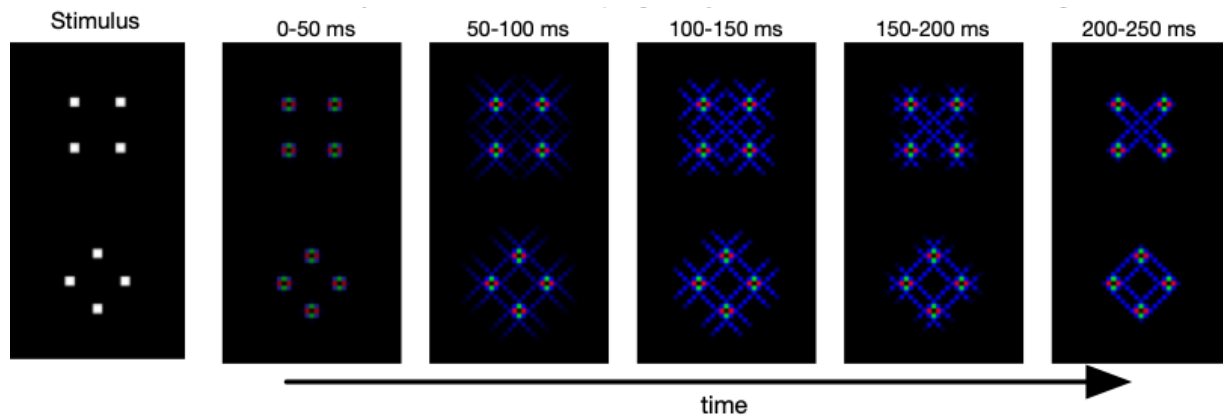


Figure 59. V2 output from a simulation using the stimulus image on the far left in which Spread Controller duration was 1.4143 ms, Spread Controller onset was 50 ms, and the Long Controller and Short Controller Circuits were both turned off. In contrast, for a simulation with a slightly shorter Spread Controller duration of 1.4142 ms (not pictured), there was no noticeable spread of diagonal signal from 50-100 ms and none of the dots connected.

This is undesirable because, looking ahead to the feature filter for the selection strategy, a connected diamond dot figure and a connected square dot figure would be relatively indistinguishable in terms of the amount of boundary signal: they both would have a lot of diagonal signal, which would make it difficult to devise a beneficial selection process using the model. For example, a diagonal feature filter would likely make it equiprobable that a selection signal would land on a target diamond or distractor square. Although diagonal connections for the diamond dots have to travel a shorter distance than those for square dots, a more precise setting of Spread Controller duration does not prevent the square dots from connecting (e.g., the simulation run for Figure 59). This is because of a combination of properties of the stimulus and Spread Controller. The diagonal signal of each dot is relatively weak, so a relatively high Spread Controller duration (it is high given that the dots are very close) is required to get each dot to spread, e.g., the image in Figure 59 under ‘50-100 ms’. But, once diagonal signal does spread from the dots, the dots of each figure will connect since the dots of each figure are extremely close to each other. In turn, even a very low and precisely set Spread Controller duration that causes the diamond dots to connect will also cause the square dots to connect. To prevent the square dots from forming diagonal connections, the Long Controller Circuit was turned on. It was effective since the diagonal connections of a square dot figure have to travel a greater distance than the diagonals across diamond dot figures. Additionally, an observer may or may not connect the dot squares via horizontal and vertical connections. In turn, there are two possible connection strategies for dot

figures: use intra-shape connections to group diamond dot figures only (e.g., Figure 57c, top image), or use intra-shape connections to group both diamond and square dot figures (e.g., Figure 57c, bottom image).

The results of pilot simulations indicated that performance is relatively similar for these two connection strategies. Due to the time it takes to run the full simulation (three weeks for 150 trials per condition) and since the pilot simulations indicate that the two beneficial connection strategies (i.e., intra-shape connections of dot diamond figures only, or inter-shape connections of the dots for all shapes) will result in very similar performance, I only ran simulations with connections among the dots of diamond dot figures (which is part of Grouping Strategy 1S summarized in Figure 57b; I did not run a full simulation of Grouping Strategy 2S).

Connection strategies for line figures. The connection strategy for line figures may involve inter-shape connections, i.e., connections between nearby diamonds. For example, in the line figure stimulus in Figure 58a, it seems plausible that an observer might connect the pairs of nearby diamonds, counting ‘two’ for the pair, ‘three’ for the triplet and adding the remaining diamond to get a total of five. This phenomenon, i.e., the enhancement of enumeration speed by grouping cues, has recently been coined ‘groupitizing’ by Starkey and McCandliss (2014) (see also van Oeffelen & Vos, 1982). In turn, it seems likely that an observer would connect diamond figures when possible, i.e., when they are nearby and appropriately aligned, in order to groupitize the diamonds and, thus, perform this enumeration task more rapidly.

So, a possible connection strategy for line figures is that the observer attempts to connect only diamonds by promoting the spread of V2 connections from the line diamonds but not the line squares (Figure 58c, bottom image). That way, diamonds that happen to be nearby and aligned may connect, which would allow them to be identified more quickly and perhaps counted as a pair or triplet. This connection strategy would likely not be used for the dot figures since, e.g., a dot from the corner of a square may connect with a dot from a diamond, which would make it difficult to segment out the dot diamond alone. Thus, it seems plausible that an observer might use a connection strategy for the line figures that encourages inter-shape V2 connections to form in order to more quickly segment out sets of targets. If the studies on groupitizing have merit, these connections may allow for faster enumeration of diamond line figures.

I have identified connection parameters that can facilitate this effect, i.e., the connection strategy for line figures is to only connect diamonds across the stimulus image when possible (see Figure 58 caption for details). Spread Controller duration and onset are the same for each orientation, but the Short Controller Circuit is set such that it prevents much spread of diagonal V2 connections for squares (because diagonal edges are detected only at the corners of squares) and the Long Controller prevents the spread of horizontal and vertical V2 connections from the long edges of squares. These same parameters allow the V2 connections of a diamond shape to spread freely because it has long diagonal edges and short horizontal and vertical edges at its corners. Figure 60 illustrates the V2 connections formed by this set of connection parameters for a simple stimulus.

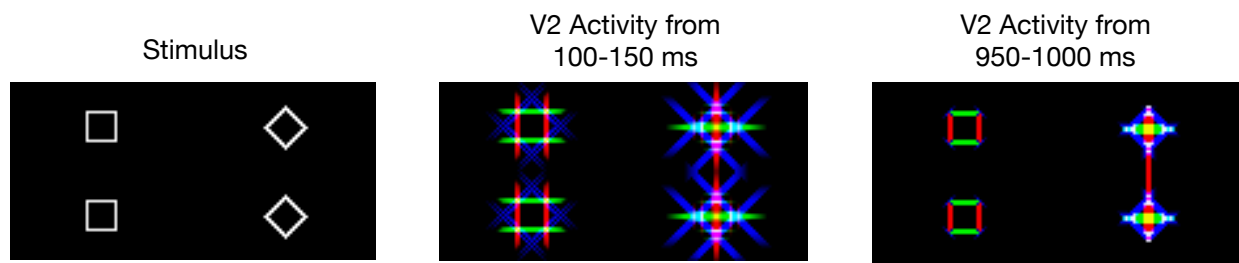


Figure 60. V2 activity at the beginning (middle image) and at a later time (right image) in a simulated trial with the set of connection parameters used in the Inter-Target connection strategy (Figure 58c, bottom image).

Initially as shown in the middle image of Figure 60, there is great spread of V2 connections of all orientations for the diamonds compared with those of the squares. As developed in the next subsection, a beneficial selection strategy involves a diagonal feature filter and, thereby, the selection signals are drawn to diagonals. But, due to the limited number of selection signals at any time, an observer would be unable to select all the targets in conditions with a large number of targets before the system settles down. However, some V2 connections may persist over time, as exemplified in the right image of Figure 60, that allow the selection process to be more efficient for conditions with a large number of targets and, in turn, emulate an observer's ability to select groups of targets in some images.

The Inter-Target connection strategy may have some empirical support if one considers it to underpin groupitizing, which has been the focus of several recent studies (e.g., Anobile et al.,

2020; Briggs et al., 2021; Maldonado Moscoso et al., 2020; Maldonado Moscoso et al., 2022; Pan et al., 2021; Schindler et al., 2020; Wege et al., 2021). However, the proposition that observers promote performance in enumeration tasks via groupitizing is relatively new (standardly being traced to Oeffelen & Vos, 1982) and is not widely accepted. Thus, it is worthwhile to compare the performance that results from implementing the Inter-Target connection strategy with the No Connections strategy in which the observer does not connect stimulus elements (Figure 58c, top image).

Selection strategies. Small selection signals, which were slightly larger (+6 pixels) than a diamond line figure (Figure 57d), were used for the dot figures to avoid selecting a distractor dot along with a target.

For line figures, these small selection signals are also reasonable to use in order to avoid selecting a distractor along with a target (the ‘Single’ selection signal size strategy in Figure 58d). However, for the line figures, there is an alternative selection signal size strategy given this stimulus set, which is called ‘Multiple’ in Figure 58d. When there are only 1 to 4 elements, there are never distractors and, thus, no risk of selecting one. So, the Multiple selection signal size strategy would be to use larger selection signals when there are 1 to 4 elements. I reason that an observer could use gist information from the scene to determine which size to use on a given trial. For example, on a trial where there is not much detected boundary signal, like the example of a trial with three items in Figure 58d, bottom image, an observer could use a larger selection signal to possibly select more items at a time since it is likely there will be only target items and, thus, no risk of selecting a distractor. In the example image, this strategy resulted in the selection of two targets with a single selection signal and, thus, resulted in a lower response time than if smaller selection signals were used. (The Multiple selection signal size strategy was not explored for dot figures since it is unlikely that the observer could quickly use gist information to determine that there were fewer than 5 shapes from an image with only dots, e.g., there would probably not be enough of a difference in boundary signal between an image with 16 dots compared to an image with 20 dots to reliably distinguish between the two conditions.)

For all simulated grouping strategies, the other components of the selection strategy were the same. A diagonal feature filter was used to guide placement of the selection signals, e.g., the first row of Figure 61a, b, and c.

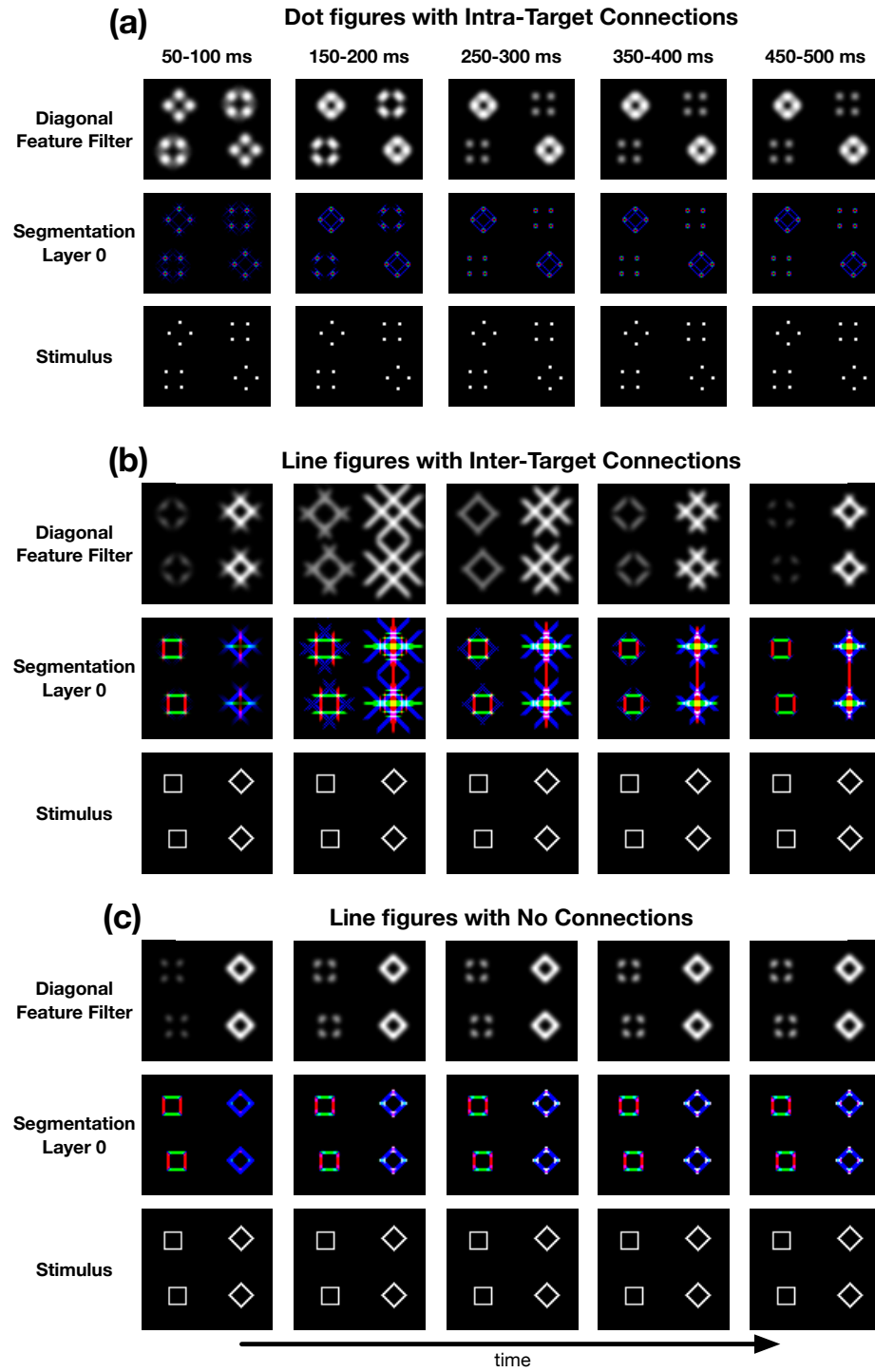


Figure 61. Examples of diagonal feature filters that result from each of the connection strategies simulated. These feature filters are shown over time since their temporal dynamics are important in my discussion of simulation results.

As explained in the “Application to Other Stimulus Images” section above, the 2D diagonal feature filters are generated by creating a probability distribution from smoothed diagonal activity in segmentation layer 0. The brighter a pixel in this feature filter, the more likely its coordinate will become the center of a selection signal. A diagonal feature filter should tend to guide selection signals to targets because the targets are diamonds, which will either have more diagonal signal (if a diamond line figure) than squares or have diagonal connections (if a diamond dot figure). On each time step of a trial beginning at 50 ms, two coordinates were sampled from the feature filter distribution. In an attempt to prevent selection signals from overlapping and thereby maximize the efficiency with which the model would find target items, the second coordinate was re-sampled up to two more times if the area of the second selection signal would overlap with the area of the first selection signal.

The remaining components of the selection strategy (timing of selection signals and reset of segmentation layers) are a bit complicated, partly due to the need to lower computational costs. Due to the large size of the stimulus, the model can only have a base segmentation layer and a single segmentation layer into which all selected elements are segmented (otherwise, with two non-base segmentation layers, it takes around 30 minutes to run 50 milliseconds of a single simulated trial). Even with this modification, it takes approximately 35 minutes (in real time) to run one trial. In view of the results of pilot simulations where I compared the results from using different numbers of selection signals, I chose to use a selection strategy with three pairs of selection signals staggered such that the first pair started at 150 ms, the second at 200 ms, and the third at 250 ms (Figure 62).

Segmentation layer 1 is reset periodically (every 150 ms beginning at 300 ms when the first pair of selection signals shifts) to clear it of distractors and, thus, segment out targets from distractors. Although this version of the model has only one non-base segmentation layer, this reset allows it to be more in line with the full version of the model that has a segmentation layer associated with each selection signal and, thereby, allows it to isolate target diamonds from distractor squares via the segmentation process. Finally, to keep selected targets in the non-base segmentation layer after reset, if a selection signal lands on an element that is identified as a diamond, then a selection signal is assigned to that location for the remainder of the trial. This acts as a sort of memory for locations of targets that have been counted in a trial.

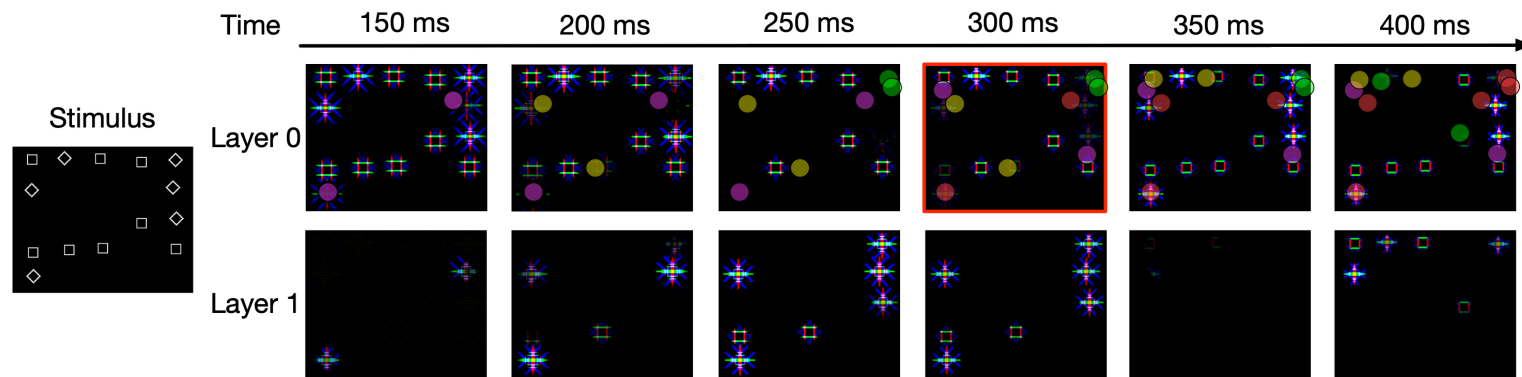


Figure 62. Illustration of the segmentation process for the Trick and Enns simulations. At 150 ms, the first pair of selection signals (pink circles) begin at some location and shift to a new location every 150 ms, e.g., at 300 ms. At 200 ms, the second pair of selection signals begin (yellow circles), and the final pair of selection signals (green circles) begins at 250 ms. Each pair of selection signals shifts to a new location after a 150 ms selection cycle. If a target is successfully segmented, a selection signal (red circles) at that location is added to the selection signal map for the rest of the trial. The reset signal occurs at the timestep at which the first pair of selection signals shift, e.g., at 300 ms indicated by the box with a red border.

Target identification algorithm (counting). Target identification is performed by dividing the non-base segmentation layer into a 5x5 notational grid and, in each grid cell, checking the distance between the highest and lowest row (and between the cell's most extreme columns) with activity in V4, layer 1. If these two most extreme rows with activity (and the two most extreme columns with activity) is a particular distance apart (i.e., greater than 12 pixels, see Figure 63), then the cell should contain a diamond.

After every 50 ms, the V4, layer 1, grid cells are checked for diamonds. The number of grid cells that have a diamond is tallied. If the same number of targets is found for five consecutive 50 ms timesteps, then the trial stops and the final count of targets is recorded. Five time steps, i.e., 250 ms, was chosen for the stopping rule since it takes some time for the non-base segmentation to recover following the reset; this was the shortest time for the vast majority of the targets selected by the staggered selection signals to eventually produce enough signal to be detected in V4, layer 1.

Simulation Stimuli, Method, and Procedure

For each figure type and target-distractor condition, 150 trials were run. Unlike the stimuli for the previous simulations, which are the same for each condition and target pair, each stimulus for this simulation was created at the start of a trial with each stimulus item placed in a randomly selected cell in a 5x5 notational grid and given a random position from the center of its cell, mirroring the procedure used to create the stimuli for Experiment 6. The sizes of the stimulus image and elements were proportionally kept as close as possible to those of Experiment 6 while reducing overall size as much as possible to shorten run time. The size of the stimulus image was 210 pixels wide and 170 pixels high. Line figure squares were 12 pixels wide made up of lines 1 pixel wide, and line figure diamonds were 17 pixels wide. Dots were 2x2 pixels. Dot squares were 12 pixels wide, and dot diamonds were 16 pixels wide. The jitter distance was sampled from a uniform distribution with a range of -4 to +4 pixels, and this value was added to one of the center coordinates of a particular grid cell. As in Experiment 6, for each stimulus element a different random sample was generated for its x-coordinate and its y-coordinated.

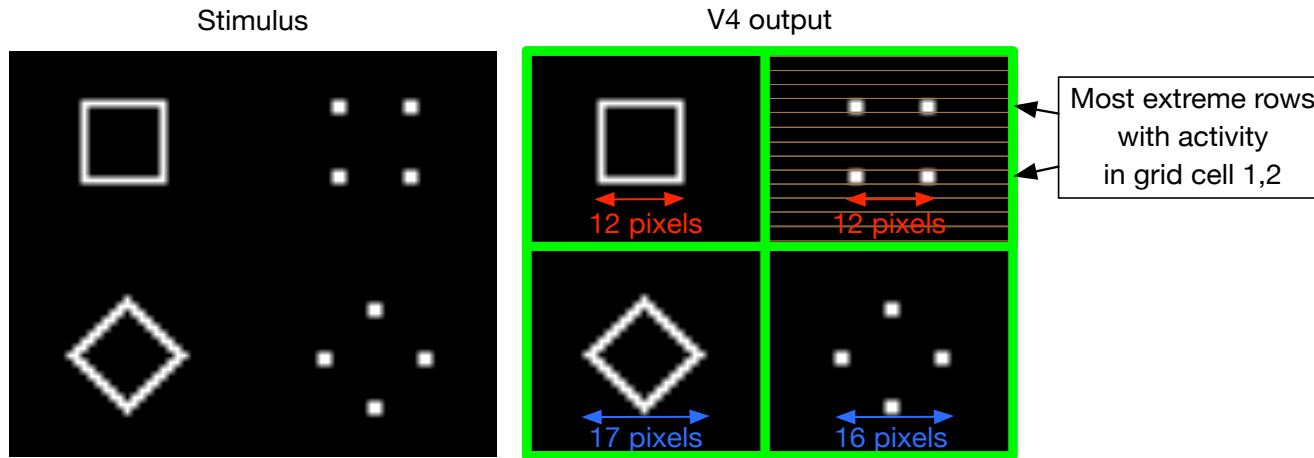


Figure 63. V4 output for an input image with each type of line figure and dot figure used in the simulation. Green lines indicate how this output is divided into a 2x2 grid consisting of four cells, where each cell may contain a single figure. Similarly, for the stimuli used for the Experiment 6 simulation, each stimulus was divided but into a 5x5 grid. Measurements in pixels are shown: the width and height of line squares and dot squares (12x12 pixels) are smaller than line diamonds (17x17 pixels) and dot diamonds (16x16 pixels). In each cell, the identification algorithm determines the row numbers in which there was V4 activity. To illustrate this, the top right cell (grid cell 1,2) is divided into rows by brown lines. (In the figure, there are two pixels per row to make the example visibly clearer, but the algorithm divides the V4 output into rows that are each one pixel tall.) The algorithm finds the most extreme row numbers that have activity in them, e.g., in grid cell 1,2, the row with the lowest row number with activity is the row that contains the top two dots, and the row with the greatest row number with activity is the row that contains the bottom two dots. The algorithm then checks whether the distance between these two rows is greater than 12 pixels. This process is also applied to columns. If the distance between the extreme rows with activity is greater than 12 pixels and if the distance between extreme columns with activity is greater than 12 pixels, then the grid cell must contain a diamond and the algorithm registers that there is a hit in this grid cell. Otherwise, as is the case in grid cell 1,2, the algorithm determines that the cell does not contain a diamond.

For line figures, four grouping strategies were simulated as summarized in Figure 58b. These are all possible combinations of the two connection strategies and selection signal size strategies discussed above and illustrated in boxes (c) and (d) of Figure 58.

Due to a memory leak that occurs with NEST 2.14.0 and the large size of the stimuli, simulations were conducted in batches of 5 trials. Additionally, some of pairs of grouping strategies for line figures, e.g., 0S (do not connect stimulus elements, and use only the small selection signal size) and 0M (do not connect stimulus elements, and use the small selection signal size for all conditions except those with 1-4 elements), had the same connection and selection strategy for all conditions except for those with 1-4 elements. Rather than running a full simulation for, e.g., Grouping Strategy 0M, I ran only conditions with 1-4 elements. These results replaced those for conditions Groupings Strategy 0S. So, other than the results for conditions with 1-4 targets, the plots of results for grouping strategies 0S and 0M are identical (and similarly for 1S and 1M). Not running repeated components of these simulations was done to reduce overall simulation time.

A trial would terminate if either the stopping rule or a maximum time (2000 ms) was reached. As with Experiment 6, only trials with the correct count were used to calculate mean response times, and the number of trials with the incorrect count (i.e., trials that (1) terminated due to either the stopping rule being satisfied and (2) ended with a target count that was not the same as the number of diamonds in the stimulus) per condition was used to calculate percent incorrect.

Model Results and Discussion

Simulation results for dot figures when Grouping Strategy S1 was implemented, i.e., connect dots of diamonds and use small selection signals for all conditions, are shown in Figure 55, top right plot. The simulated results for dot figures match that of Experiment 6 (Figure 55, top middle plot) in several key respects. First, the blue curve has a relatively flat slope for 1-3 targets compared with 5-7 targets. Second, the curves for the 4 and 8 distractor conditions are generally linearly increasing. Third, there is separation between each curve, with a larger separation between the curve for conditions with 0 distractors and the curve for conditions with 4 distractors. However, compared to Experiment 6 error rates, model error rates are higher. This could reflect, e.g., the model revisiting locations where a distractor was already detected, unlike a human observer who may avoid attending to areas already searched.

I also conducted a simulation for dot figures with the same selection signal size strategy but no connections between dots; however, this resulted in error rates that were near or at 100% for most of the 4 and 8 distractor conditions. I do not present the results here since, for some conditions, there were not any correct trials to calculate mean response time from. However, this indicates that connections between the dots of diamond targets are critical to good performance on this task.

For line figures, a total of four grouping strategies were implemented in the model, which are summarized in Figure 58b. Grouping Strategy 1M, which connected nearby diamonds if possible and used a large selection signal for conditions with 1-4 items, both had the best performance (lowest overall mean response times) and matched the pattern of results for line figures from Experiment 6. The results from this experiment and this simulation are shown side-by-side in Figure 55, bottom row. In both the experiment and simulated results, the blue curve has a relatively flat slope around 1-3 targets, which increases at 5-7 targets. Additionally, curves for the 4 and 8 distractor conditions are generally linearly increasing. However, simulation error rates are a bit higher for conditions with 8 distractors, and there is not as much separation between the simulation curves compared to the curves for Experiment 6.

Results for all four simulated grouping strategies are provided in Figure 64.

The left column shows results from grouping strategies with no connections, while the right column provides results from grouping strategies with connections between nearby targets only (i.e., the Inter-Target connection strategy in Figure 58c). The top row provides plots of model results when the Single selection signal size strategy was implemented, while the bottom row shows results for the Multiple size strategy. Since the difference in selection signal strategy only applied when there were 4 items (i.e., small selection signals were used for all conditions in the Single size strategy, while small selection signals were used in all conditions except when there were 1-4 items in the Multiple size strategy), the only difference between the top and bottom plots are the data points along the blue curve for 1-4 targets. In the bottom row, those points were calculated from simulations in which the large selection signal size was used (Figure 58d).

Compare the left plots (results from grouping strategies with No Connections) with the right plots (results from grouping strategies with Inter-Target connections) of Figure 64. It is obvious that Inter-Target connections led to lower overall response time and lower error rates, especially as the number of elements increases.

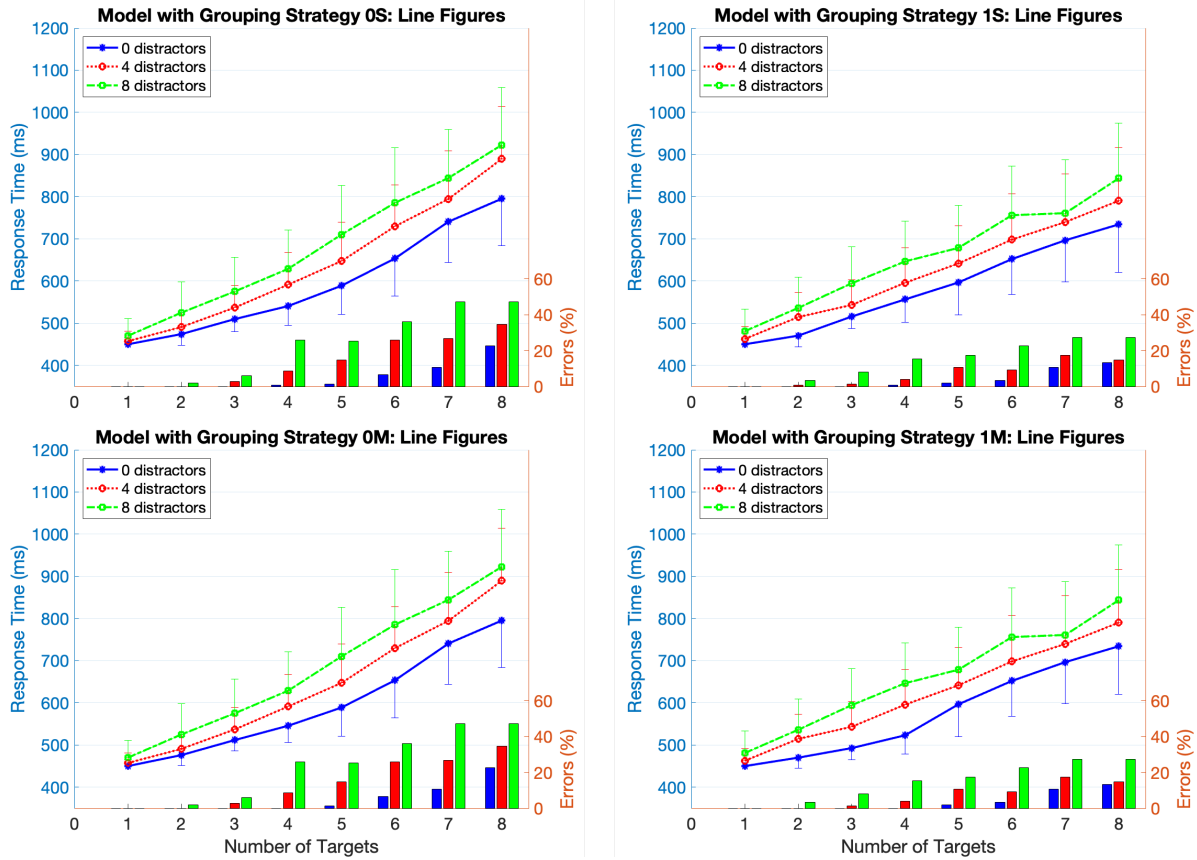


Figure 64. Plots of simulation results for all grouping strategies implemented for the line figure condition. Mean response times shown in each plot were calculated from the reaction times in trials in which the correct number of targets was reported. Error bars represent standard deviation.

With No Connections, the selection signal size strategy makes very little difference in the results; the data points for 1-4 targets and 0 distractors appear identical in the plots in the left column. With the Inter-Target connection strategy, however, the Multiple selection signal size strategy leads to much faster response times for these conditions. This occurs due to the difference in the spread of boundary signal early in the trial. Return to Figure 61b. With the Inter-Target connection strategy, there is a lot of spread in segmentation layer 0 from diamonds (Figure 61b, middle row). Since diagonal signal is used to generate the feature filter (Figure 61b, top row), the spread of boundary signal causes selection signals to be drawn to large areas around targets. In turn, small selection signals are less efficient at finding multiple targets early in a trial with this connection strategy, e.g., it is more likely that a selection signal misses a target completely or ends up a bit far away from the center of a target and, in turn, there may not be enough time for the

selection signal to spread and segment enough boundary signal for the target to be identified before the next reset. For this connection strategy, a large selection signal has two advantages over a smaller one. First, it is more likely that a selected target item will be selected and identified before the next reset even if it is placed a bit far from the center of the target. Second, the spread of diagonal signal makes it more likely that a large selection signal will fall on more than one target since such spread results in a probability distribution with an enlarged area where a selection signal center coordinate can be sampled from. In contrast, with no connections, there is no spread of boundary signal (Figure 61c, middle row). The feature filter generated from this boundary signal (Figure 61c, top row) is quite localized for the duration of a trial. In turn, a smaller selection signal doesn't incur much of a disadvantage: due to the relatively small regions that selection signals can be guided to, it is likely a selection signal will cover a target and segment it sufficiently for identification regardless of whether a small or large selection signal is used.

Why doesn't this issue for line figures with the Inter-Target connection strategy occur for dot figures with the Intra-Target connection strategy? This is due to the difference in Spread Controller duration used to implement these connection strategies. For connections between the dots of a dot figure, the duration was very low (1.5 ms), while it was quite high (50 ms) to connect diamond line figures with nearby diamond line figures. In turn, the spread from dot diamonds is just enough to connect the dots of a dot figure and, thus, does not cause much diagonal spread beyond the figure. This can be seen by comparing the feature filters produced by dot figures with the Intra-Target connection strategy (Figure 61a) with those produced by line figures with the Inter-Target connection strategy (Figure 61b). The feature filters produced by the diagonal activity from dot figures is quite localized throughout a trial, while those for line figures spreads quite far beyond the targets. So, unlike line figures with the Inter-Target connection strategy, the Single selection signal size strategy works well for dot figures with the Intra-Target connection strategy.

In sum, when the model implements Grouping Strategy 1S (Intra-Target connections and Single selection signal size strategy), it does a good job of accounting for the dot figure data from Experiment 6. Of the grouping strategies implemented for line figures, Grouping Strategy 1M (Inter-Target connections and Multiple selection signal size strategy) generally leads to the best performance in terms of lower overall response times compared to Grouping Strategy 1S and has a lower error rate than the grouping strategies without connections. Plus, the results from Grouping Strategy 1M best matches the pattern of mean response times from Experiment 6 for line figures

in terms of the shape of the curves: the blue curve is relatively flat when there are 1-3 targets, and the non-zero distractor curves are generally linearly increasing. However, error rates are a bit higher than those of the experiment, and the curves for each distractor condition produced by the simulated results have less separation between them.

Given that the model can produce similar patterns of results as Experiment 6, the model can be used to offer some explanation of performance on this task and stimulus set. Recall that Trick and Enns claim that the dot figure results indicate that there are two separate grouping processes: element clustering and shape formation. They claim that the flat slope of the zero distractor condition curve indicates that only the preattentive element clustering process was used. If the shape formation process, which requires attention, was used to enumerate the target items, then there would be no such flat slope. Setting aside the claim it is preattentive, element clustering, i.e., which elements belong together, could be regarded as the product of grouping the dots by V2 connections. According to the model, the connections between dots resulting from the Intra-Target connection strategy are involved for all conditions. So, assuming the clustering process is realized by connections, element clustering does not explain why the zero distractor condition has a relatively flat slope. Instead, the observer has very fast response times when there are 1-4 items because of the ratio of items to selection signals. One or two pairs of selection signals are usually sufficient to segment and identify all targets when there are only 1-4 elements in the image. In turn, model response times are very fast and similar when there are few items, and this results in the relatively flat slope at the beginning of the curve for the zero distractor condition. The model lacks a separate subitizing process; it just has cells that detect edges and manipulate connections, a selection process, and some algorithms for identifying and counting targets. Thus, the explanation offered by the model for dot figures arguably fleshes out the suggestion by Balakrishnan and Ashby (1991, 1992) that the pattern indicative of subitizing reflects the limited capacity of visual attention. According to the model, this pattern is due to the grouping strategy adopted for this task and stimulus set; the observer selects targets, which have their elements connected but do not connect with other targets, with a limited number of small selection signals that are guided by diagonal signals. On trials with no distractors and few targets, the target diamonds can be selected quite quickly. When there are more targets, it takes longer for the limited number of selection signals to segment all the targets. When there are distractors, some selection

signals are drawn to distractor dots and segment them rather than targets, which results in the model taking more time to find all the targets.

A similar explanation of the results for line figures can be offered. But, before providing it, it is informative to examine the notion of groupitizing in terms of the model. The Inter-Target connection strategy may be regarded as an implementation of groupitizing, i.e., the grouping of a subset(s) of targets to enhance performance on enumeration tasks. For this task, the model indicates that Inter-Target connections do improve performance (lower error rates and lower response times as element number increases) compared to using no connections. So, the model gives some support to claims made by advocates of groupitizing if it involves V2 connections between targets. However, for at least this task and stimulus set, groupitizing can come at a cost early in a trial if small selection signals are used due to the spread of boundary signal before the system settles down. For this stimulus set, it was reasonable to use larger selection signals when there were few items since there was no risk of selecting a distractor. This allowed the fast selection of targets when there were 1-4 items, and it was the Multiple selection size strategy that allowed the model to produce the pattern indicative of groupitizing in the zero distractor condition. Thus, again, the curve thought to be characteristic of subitizing is the result of a grouping strategy that promotes performance on the task, rather than the product of some dedicated subitizing process.

GENERAL DISCUSSION AND CONCLUSIONS

By rejecting the implicit assumption that there is only one way to group the elements of a given stimulus, we are better poised to investigate what grouping strategy(s) observers may be using to promote performance on a particular task and stimulus set. The model developed in this project was designed with the flexibility to run simulations of visual search tasks using a wide variety of grouping strategies.

Generally, the model demonstrates how bottom-up and top-down mechanisms are integrated in both the formation of perceptual groups and the implementation of grouping strategies in particular tasks, where a grouping strategy consists of a connection strategy and a selection strategy. Top-down connection control is driven by the observer's intention to promote small or large groups of elements via V2 connections. Bottom-up information, which can be regarded as "gist" information from the image itself, can be used to set connection parameters that are appropriate for the scene and can also be part of a strategy defined by the observer. Top-down selection control involves selection signals of a particular size that are guided by a feature filter. The selection signal size and type of feature filter are chosen as part of a selection strategy that promotes performance on a particular task and stimulus set. Feature filters can be tuned to contour information that is from the image itself and from V2 connections that are the result of the combination of top-down and bottom-up processes just described. In effect, model performance indicates that an observer uses a connection strategy in conjunction with a selection strategy that promotes performance on a given task and stimulus set.

Through simulations and experiments, I tested the possibility that observers are able to tune the Connection Circuit parameters and choose properties of the selection process for particular stimuli and tasks so as to connect, select and segment elements in a way that promotes performance. Within an experiment, the parameters can be changed to reflect previous failures to find the target. And within a trial, strategic top-down control of the Connection and Selection Circuits corresponds to different search strategies that change connections and selection signal properties in order to promote target search. The resulting grouping strategy is tailored to promote performance on a given task and stimulus set.

This research shows that when the model uses particular grouping strategies, simulated results generally match human performance in visual search tasks where perceptual grouping is

induced by proximity and shape similarity (Palmer & Beck, 2007), by the spacing of irrelevant distractors and size similarity (Vickery, 2008), or by the proximity of dots and the proximity and shape similarity of line figures (Trick & Enns, 1997). The explanations offered in each of these studies was in terms of Gestalt grouping principles: Palmer and Beck claimed that participants had lower response times for within-group and neutral conditions since the target pair was groupable by proximity and/or shape similarity, Vickery referred to grouping by size similarity for target circles and grouping by proximity of crosses, and Trick and Enns stated that grouping by proximity or similarity factors might be involved in element clustering.

Due to the dominance of Gestalt grouping principles in explanations offered by these researchers, it is imperative to consider the question: How do connections in the model relate to Gestalt grouping principles? Strictly speaking, the model neither obeys Gestalt grouping principles nor directly implements them by, e.g., recognizing that pairs of adjacent objects are of the same size or shape. The model does not know what shape or size are. Instead, it can only detect and manipulate edges. To produce connections between edges, the connection circuits use information about their arrangement, i.e., whether edges are (1) appropriately aligned, (2) an appropriate distance apart, (3) of a particular orientation, and (4) of a particular length. To see how the model uses these properties to produce groupings, consider some examples in Figure 30. In the grouping by proximity examples, some positive Spread Controller durations allow connections to spread between edges that are aligned and that are within a particular distance. In the grouping by orientation examples, some positive Spread Controller durations allow connections to spread between edges that are aligned and are of the same orientation, while the Long/Short Controller may also be in play to prevent aligned but relatively long edges, e.g., the horizontal bar in the row 3 example in Figure 30, or short edges, e.g., the top of the vertical line and the top of the diagonal line in the row 4 example, from connecting to other edges. In the grouping by similarity of size examples, some positive Spread Controller durations allow connections to spread between edges that are aligned, while the Short Controller prevents the shorter edges of the small shapes from connecting to the partially aligned edges of large shapes. In the grouping by shape examples, Spread Controller duration is again used to promote connections between aligned edges, while the Short and Long Controllers may be used to prevent the edges of particular shapes from connecting, e.g., the Short Controller prevents the short horizontal edges of the H and X from connecting, while the Long Controller prevents the long top edge of a square from connecting with the top

edge of a circle. Thus, fundamentally, the model uses information about properties (1)-(4) to connect particular elements.

Although not an instantiation of Gestalt principles, some of these properties are relatable to properties identified by the Gestalt grouping principles. Property (2) reflects the distances between the elements and, thus, proximity. Property (4) is, arguably, a function of shape for the similarity of shape examples set, e.g., the top edge of a square (circle) is the same length as that of the other square (circle) in the stimulus yet is much longer (shorter) than the top edge of a circle (square). And, size could be regarded as a function of (1) alignment of edges and (4) relative length, e.g., the pair of small circles have shorter edges and the tops and bottom edges of this pair are well aligned while the large circles have larger edges that are well aligned. However, it seems that properties (1), (2), and (3) are involved in all of these examples: creating and manipulating these connections require that the edges are appropriately aligned, be within a certain distance apart (which can be altered by top-down control depending on the Spread Controller duration and onset), and be of a particular orientation (which can be altered by top-down control depending on which orientations have a positive Spread Controller duration).

In sum, at least in view of the examples given above, the model indicates that there are several properties of the edges of stimulus elements that allow an observer to group them by connections. The elements that reflect expected groupings given classic grouping principles connect due to the combination of these edges having particular properties and a set of connection parameters that are tuned so as to connect particular edges. Although some of these properties relate to Gestalt grouping principles, there is not a clear one-to-one mapping between each property and a Gestalt principle. Thus, it seems that Gestalt grouping principles are high level descriptions of output (i.e., particular patterns of connected elements) that result from simpler underlying mechanisms, which only use information about the arrangement of edges, implemented in the model.

Further, these connections are only part of the account proposed here. It is not enough to, e.g., connect a target pair of elements. The target also needs to be selected, segmented, and isolated from the scene so that it can be easier to identify. Additionally, the selection process itself can be used to group objects, e.g., the grouping by selection of two nearby target diamonds that happened to both be selected by a single large selection signal. In turn, this work supports the claim that both connections and selection are components of grouping.

The model's interpretation of the empirical data highlights that perceptual grouping is neither a well-defined concept nor a process that has a single mechanism or even a series of mechanisms. Rather, I propose that what is referred to as "perceptual grouping" reflects many different model behaviors that together achieve a given task. What is described as a type of grouping may involve many different mechanisms depending on the task and stimulus set. The experiments simulated in this work provide examples of this. Table 11 summarizes the grouping strategy for each simulated experiment that resulted in the best performance on the task. In the table, the experiments are listed in order of increasing complexity of grouping strategy.

Table 11. Best Performing Grouping Strategy by Experiment

Experiment	Connection Strategy		Selection Strategy	
	One Strategy?	One Set of Parameters?	One Signal Size?	One Placement Strategy?
5. (Vickery)	Yes. (Connect targets by size.)	Yes.	Yes.	Yes.
6. Dot Figures (Trick & Enns)	No. (Connect the dots of each target or each shape.)	Yes.	Yes.	Yes.
6. Line Figures (Trick & Enns)	Yes. (Connect nearby targets when possible.)	Maybe. (May depend on the spread of elements.)	No.	Yes.
1. (Palmer & Beck)	Yes. (Connect target pairs when possible.)	No. (One for each separation between target pairs.)	No. (Depends on row width.)	Yes. (At least initially.)

Rather than relying on Gestalt grouping principles, the model offers an explanation of performance on these tasks in terms of low-level mechanisms such as those involved in the process that forms V2 connections and in the selection process. The complexity of involved mechanisms that operate in parallel make it challenging to empirically isolate one mechanism from others, a point that has been apparent in the empirical literature for quite some time (Wagemans, 2018). However, in the model top-down control of these mechanisms can be directly manipulated, thereby allowing for better understanding of how these mechanisms contribute to performing the task at hand.

REFERENCES

- Anobile, G., Castaldi, E., Maldonado Moscoso, P. A., Burr, D. C., & Arrighi, R. (2020). "Groupitizing": A strategy for numerosity estimation. *Scientific Reports*, 10(1), 1-9.
- Balakrishnan, J. D., & Ashby, F. G. (1991). Is subitizing a unique numerical ability?. *Perception & Psychophysics*, 50(6), 555-564.
- Balakrishnan, J. D., & Ashby, F. G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research*, 54(2), 80-90.
- Beck, D. M., & Palmer, S. E. (2002). Top-down influences on perceptual grouping. *Journal of Experimental Psychology: Human Perception & Performance*, 28, 1071-1084.
- Bevan, W. (1961). Perceptual learning: An overview. *The Journal of General Psychology*, 64(1), 69-99.
- Bevan, W., & Zener, K. (1952). Some influences of past experience upon the perceptual thresholds of visual form. *The American Journal of Psychology*, 65(3), 434-442.
- Braly, K. W. (1933). The influence of past experience in visual perception. *Journal of Experimental Psychology*, 16(5), 613-643.
- Briggs, G., Harner, H., & Khemlani, S. (2021). Preferences in the quantified description of visual groups. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43(43), 2910-2915.
- Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, 18, 515-578.
- Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 958-975.
- Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2, 324-353.
- Folk, C. L., Egeth, H., & Kwak, H. (1988). Subitizing: Direct apprehension or serial processing? *Perception and Psychophysics*, 44, 313-320.
- Francis, G., & Bornet, A. (2019). A model with top-down control of the range of perceptual grouping. *Journal of Vision*, 19, 151a.

- Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research*, 34, 1089-1104.
- Francis, G., Manassi, M., & Herzog, M. H. (2017). Neural dynamics of grouping and segmentation explain properties of visual crowding. *Psychological Review*, 124(4), 483-504.
- Gewaltig, M.-O., & Diesmann, M. (2007). NEST (Neural Simulation Tool). *Scholarpedia* 2(4), 1430.
- Gottschaldt, K. (1926/1950). Gestalt factors and repetition. In W. D. Ellis (Ed.), *A sourcebook of Gestalt psychology* (pp. 109-122). New York, NY: Humanities Press.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87, 1-51.
- Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92(2), 173-211.
- Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics*, 38(2), 141-171.
- Grossberg, S., & Raizada, R. D. S. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40, 1413-1432.
- Han, S., Humphreys, G. W., & Chen, L. (1999). Uniform connectedness and classical Gestalt grouping principles of perceptual grouping. *Perception & Psychophysics*, 61(4), 661-674.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194-203.
- Kimchi, R., & Hadad, B. S. (2002). Influence of past experience on perceptual grouping. *Psychological Science*, 13, 41-47.
- Kirchberger, L., Mukherjee, S., Schnabel, U. H., van Beest, E. H., Barsegyan, A., Levelt, C. N., Heimel, J. A., Lorteije, J. A. M., van der Togt, C., Self, M. W., & Roelfsema, P. R. (2021). The essential role of recurrent processing for figure-ground perception in mice. *Science Advances*, 7, eabe1833.
- Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences, second edition*. Belmont, CA: Brooks/Cole.

- Klein, A., Silvester, S., Tanbakuchi, A., Müller, P., Nunez-Iglesias, J., Harfouche, M., McCormick, M., Rai, A., OrganicIrradiation, Smith, T. D., Konowalczyk, M., rreilink, Nises, J., jackwalker64, Vaillant, G. A., Zulko, NiklasRosenstein, Hirsch, M., Schambach, M.,... Elliott, A. (2018, November 15). imageio/imageio: V2.4.1 (Version v2.4.1). Zenodo. <http://doi.org/10.5281/zenodo.1488562>
- Koffka, K. (1935/1963). *Principles of Gestalt psychology*. New York, NY: Harcourt, Brace & World.
- Köhler, W. (1929). *Gestalt psychology*. New York, NY: Horace Liveright.
- Kogo, N., Strecha, C., Van Gool, L., & Wagemans, J. (2010). Surface construction by a 2-D differentiation–integration process: A neurocomputational model for perceived border ownership, depth, and lightness in Kanizsa figures. *Psychological Review*, 117(2), 406-439.
- Lawrence, M. A. (2016). EZ: *Easy analysis and visualization of factorial experiments* (R Package Version 4.4-0) [Computer software]. Retrieved from <https://CRAN.R-project.org/package=eZ>
- Locher, P. J., & Wagemans, J. (1993). Effects of element type and spatial grouping on symmetry detection. *Perception*, 22, 565-587.
- Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, 9(12), 1-11.
- Moore, M. G. (1930). Gestalt vs. experience. *The American Journal of Psychology*, 42(3), 453-455.
- Maldonado Moscoso, P. A., Castaldi, E., Burr, D. C., Arrighi, R., & Anobile, G. (2020). Grouping strategies in number estimation extend the subitizing range. *Scientific Reports*, 10(1), 1-10.
- Maldonado Moscoso, P. A., Greenlee, M. W., Anobile, G., Arrighi, R., Burr, D. C., & Castaldi, E. (2022). Groupitizing modifies neural coding of numerosity. *Human Brain Mapping*, 43(3), 915-928.
- Nan, Y., Knösche, T. R., & Luo, Y. J. (2006). Counting in everyday life: Discrimination and enumeration. *Neuropsychologia*, 44(7), 1103-1113.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.
- Palmer, S. E. (2002). Perceptual organization in vision. In H. Pashler (Ed.), *Stevens' handbook of experimental psychology: Vol. 1. Sensation and perception* (3rd ed., pp. 177-234). New York, NY: Wiley.

- Palmer, S. E., & Beck, D. M. (2007). The repetition discrimination task: An objective method for studying perceptual grouping. *Attention, Perception & Psychophysics*, 69(1), 68-78.
- Pan, Y., Yang, H., Li, M., Zhang, J., & Cui, L. (2021). Grouping strategies in numerosity estimation between intrinsic and extrinsic grouping cues. *Scientific Reports*, 11, 17605. <https://doi.org/10.1038/s41598-021-96944-x>
- Peterson, M. A., & Kimchi, R. (2013). Perceptual organization in vision. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 9–31). New York, NY: Oxford University Press.
- Peyser, A., Sinha, A., Vennemo, S. B., Ippen, T., Jordan, J., Graber, S., Morrison, A., Trench, G., Fardet, T., Mørk, H., Hahne, J., Schuecker, J., Schmidt, M., Kunkel, S., Dahmen, D., Eppler, J. M., Diaz, S., Terhorst, D., Deepu, R.,... Plesser, H. E. (2017, October 20). NEST 2.14.0. Zenodo. <http://doi.org/10.5281/zenodo.882971>
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. [Computer software]. Retrieved from <https://www.R-project.org/>.
- Railo, H., Koivisto, M., Revonsuo, A., & Hannula, M. M. (2007). The role of attention in subitizing. *Cognition*, 107, 82-104.
- Raizada, R., & Grossberg, S. (2001). Context-sensitive bindings by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition*, 8, 431-466.
- Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, 29, 203-227.
- Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception, & Psychophysics*, 73, 2542-2572.
- Schindler, M., Schoenberg, V., & Schabmann, A. (2020). Enumeration processes of children with mathematical difficulties: an explorative eye-tracking study on subitizing, groupitizing, counting, and pattern recognition. *Learning Disabilities: A Contemporary Journal*, 18(2), 193-211.
- Starkey, G. S., & McCandliss, B. D. (2014). The emergence of “groupitizing” in children’s numerical cognition. *Journal of Experimental Child Psychology*, 126, 120-137.

- Stewart, C. A., Cockerill, T. M., Foster, I., Hancock, D., Merchant, N., Skidmore, E., Stanzione, D., Taylor, J., Tuecke, S., Turner, G., Vaughn, M., & Gaffney, N. I. (2015). Jetstream: a self-provisioned, scalable science and engineering cloud environment. *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, 1-8. <http://dx.doi.org/10.1145/2792745.2792774>
- Thórisson, K. R. (1994). Simulated perceptual grouping: An application to human-computer interaction. In A. Ram & K. Eiselt (Eds.), *Proceedings of the 16th annual conference of the Cognitive Science Society* (pp. 876-881). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, S. L., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., & Wilkins-Diehr, N. (2014). XSEDE: Accelerating scientific discovery. *Computing in Science & Engineering*, 16(5), 62-74. doi:10.1109/MCSE.2014.80
- Trick, L. M. (2008). More than superstition: Differential effects of featural heterogeneity and change on subitizing and counting. *Perception & Psychophysics*, 70(5), 743-760.
- Trick, L. M., & Enns, J. T. (1997). Clusters precede shapes in perceptual organization. *Psychological Science*, 8(2), 124-129.
- Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception & Performance*, 19, 331-351.
- Trick, L. M., & Pylyshyn, Z. W. (1994a). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, 101, 80-102.
- Trick, L. M., & Pylyshyn, Z. W. (1994b). Cueing and counting: Does the position of the attentional focus affect enumeration? *Visual Cognition*, 1, 67-100.
- van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition*, 10(4), 396-404.
- Vecera, S. P., & Behrmann, M. (2001). Attention and unit formation: A biased competition account of object-based attention. In T. F. Shipley & P. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 145-180). New York, NY: Elsevier.
- Vickery, T. J. (2008). Induced perceptual grouping. *Psychological Science*, 19(7), 693-701.
- Vickery, T. J., & Jiang, Y. V. (2009). Associative grouping: Perceptual grouping of shapes by association. *Attention, Perception, & Psychophysics*, 71(4), 896-909.

- Wagemans, J. (2018). Perceptual organization. In J. T. Wixted and J. Serences (Eds.), *The Stevens' handbook of experimental psychology and cognitive neuroscience: Vol. 2. Sensation, perception & attention* (pp. 803-872). Hoboken, NJ: John Wiley & Sons, Inc.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Hydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172-1217.
- Ward, E. J., & Chun, M. M. (2016). Neural discriminability of object features predicts perceptual organization. *Psychological Science*, 27(1), 3-11.
- Wege, T. E., Trezise, K., & Inglis, M. (2021). Finding the subitizing in groupitizing: Evidence for parallel subitizing of dots and groups in grouped arrays. *Psychonomic Bulletin & Review*, 1-9. <https://doi.org/10.3758/s13423-021-02015-7>
- Wertheimer, M. (1923/1950). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A sourcebook of Gestalt psychology* (pp. 71-81). New York, NY: Humanities Press.
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202-238.
- Zhang, S., Xu, M., Kamigaki, T., Do, J. P. H., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L. & Dan, Y. (2014). Long-range and local circuits for top-down modulation of visual cortex processing, *Science*, 345(6197), 660-665.
- Zemel, R. S., Behrmann, M., Mozer, M. C., & Bavelier, D. (2002). Experience-dependent perceptual grouping and object-based attention. *Journal of Experimental Psychology: Human Perception & Performance*, 28, 202-217.