Supplemental Information

Dynamic neural field (DNF) models are bi-stable attractor networks (Johnson, Spencer, & Schöner, 2009) that are composed of layers of neurons organized by functional topography along continuous, metrically organized dimensions (e.g., color). Through a local excitation/surround inhibition activation profile, local subpopulations within these fields can achieve a stable active state (i.e., a working memory state). The activation state within a layer depends on its intrinsic dynamics (e.g., the strength with which neighboring neurons are exciting or inhibiting each other) and external inputs.

 The DNF model used here consists of a 3-layered working memory system coupled to a decision system (see also Johnson et al., 2009a, 2009b). The 3-layer system consists of a contrast field (CF) which encodes visual stimuli and contrasts perceived and remembered inputs, a working memory (WM) field which maintains information about the sample array even in the absence of inputs, and a shared layer of inhibitory interneurons. The decision system consists of a “same” and a “different” node. These nodes strongly inhibit one another implementing winner-take-all dynamics and allowing the model to generate an active decision on every trial. The “same” node receives activation from the WM layer and the “different” node receives input from CF. If a new item is presented in the test array, then activation within CF will trigger a “different” response. If no new items are presented, sustained activation within WM will suppress activation in CF and trigger a “same” response. A final component of the architecture is a gate node that becomes activated when items are actively being maintained in WM and new inputs are detected (i.e., the test array is presented). Once activated, the gate node allows inputs from the 3-layer system to impact the decision nodes. Additionally, activation of the gate node turns off the projection from CF into WM serving to protect WM from updating based on inputs in the test array.

 Below we provide a full description of the model, including the model equations and details of the model-based fMRI approach. Table 1 provides a glossary of the mathematical notation used and Table 2 lists all parameter values. Parameter values were adapted from a previously reported model (Johnson, Simmering, & Buss, 2014). In general the model performed poorly at high set-sizes using these parameters; this is not surprising since previous work with this model has not simulated up to set-size 8. Several parameter adjustments were made to accommodate performance at the larger set-sizes. Since the equations are non-linear dynamical systems, each contribution to activation can depend on many other contributions. Thus, a series of parameter changes were required to maintain the balance of excitatory and inhibitory interactions within each component of the model. First, we modulated the strength of the input across set-sizes to better approximate the normalization that occurs in neural systems when multiple items are presented (see Jancke et al., 1999); Second, we increased the resting level and decreased excitation in the contrast layer to compensate for the decrease in input strength. Third, we adjusted the inhibitory projections into the excitatory layers to balance out the decrease in excitation. Fourth, the parameters for the gate and decision nodes needed to be adjusted to compensate for the changes in dynamics within the excitatory layers. Finally, colored noise was added to the resting level of the fields to impose additional realism in the stochastic fluctuations of neural systems.

**DNF Model Equations**

The contrast field (*u*; Equation 1) consists of self-excitatory neural units that receive stimulus inputs matching properties of the sample and test arrays. This layer is coupled to an inhibitory layer (*v*; Equation 4) and a working-memory layer (*w*; Equation 5), as well as to both decision nodes (Equations 6 and 7).

 Equation 1 specifies the rate of change of activation of the contrast field:

1

Activation in the contrast field evolves over the timescale determined by the *τ* parameter. The first term in Equation 1, *-u*(*x,t*), is a stabilizing term that serves to maintain activation around a stable attractor state. The second term, *h*, determines the neural resting level for the field (*h*<0) while the third term, *s*(*x,t*), specifies the stimulus input over the course of a simulation (i.e., the presentation of the colors values in the sample and test arrays). Next is local excitation, , which is defined as the convolution of a Gaussian interaction kernel, , with the gated output from the CF layer, *g*(*u*(*x’,t*)). The Gaussian interaction kernel (Equation 2) determines the spread of activation to neighboring units (see *σ*) with a strength determined by the amplitude parameter, *a*.

2

We used a sigmoidal gating function, *g* (see Equation 3). This function normalizes field activation such that activation levels lower than the threshold of 0 (*u0*) contribute relatively little to neural interactions, while activation levels higher than 0 contribute strongly to neural interactions. The *β* parameter determines the steepness of the sigmoidal function.

3

The next term in Equation 1 is the lateral inhibition component, , which is defined as the convolution of a Gaussian interaction kernel determining the spread of surround inhibition and the gated output from the inhibitory layer (*v*). There is also a global inhibitory contribution specified by, , which is applied homogenously across the field. The next term specifies spatially correlated noise, , which is defined as the convolution between a Gaussian kernel and a vector of white noise. This simulates a set of noisy inputs to the contrast field. The last two terms specify inputs from the decision nodes. Both of these inputs are gated by the gating function (*g*). The ‘different’ node (*d*) globally excites the *u* layer, , while the same or “match” node (*m*) globally inhibits the *u* layer, . These excitatory and inhibitory inputs help maintain peaks in CF if a difference is detected, and help suppress activation in CF if ‘sameness’ is detected.

 The rate of change of activation in the inhibitory layer (*v*) is given by Equation 4:

4

As with Equation 1, the first two terms correspond to the stabilization term and the neural resting level. The next two terms specify excitatory inputs from both the *u* and *w* layers. These two terms are defined by the convolution of a Gaussian kernel with a sigmoidal gating function. The final term corresponds to spatially correlated noise. This noise source is independent from the noise source in Equation 1. Note that activation in the inhibitory layer evolves over a slower timescale than in the *u* or *w* layers (*τi* > *τe*).

The WM layer (*w*) is given by Equation 5:

5

The first three terms correspond to the stabilization term, resting level, and stimulus input. Next, this layer is self-excitatory, with the spread of local excitation determined by the Gaussian kernel. The fifth term modulates input to the WM layer. In particular, the WM layer receives input from the *u* layer, , but only when the sample array is presented. At test, this input is shut down by activation of the gate or “outcome” node (*o*; see Equation 8) to protect the contents of WM during the decision phase. Note that in previous versions of the model, this term was not included; rather, we allowed WM to update when a change was detected.

The next term in Equation 5 specifies the lateral inhibition from the *v* layer followed by a term for global inhibition and then a term for spatially correlated noise. Note that this noise source is independent from the noise sources in Equations 1 and 4. The last two terms correspond to inputs from the decision nodes. The same or “match” node (*m*) globally boosts the *w* layer when ‘sameness’ is detected, while the ‘different’ node (*d*) globally inhibits the *w* layer when differences are detected.

 Equation 6 specifies the rate of change of activation of the “different” node.

6

The first two terms are stabilization and resting level terms. These are followed by a self-excitation term, , and inhibition from the same or “match” node, , both of which are gated by the sigmoidal function. These terms implement the winner-take-all interactions between the ‘different’ and ‘same’ nodes (see also, Equation 7). The ‘different’ node receives a boost of input at test from the activation of the gate or “outcome” node, , indicating that a decision or outcome is required. The activation of the outcome node also gates the input to the ‘different’ node from the CF layer, . The final term in Equation 6 specifies a random input to the ‘different’ node with strength, *ar*.

Equation 7 specifies the rate of change of activation of the same or “match” node:

7

This equation is comparable to Equation 6. In particular, the “match” node is self-excitatory, , and receives inhibition from the “different” node, . This node also receives a boost from the outcome node, , along with input from the WM layer, when an outcome is required, and inhibition from Inhib layer, . Note that tonic input from the inhibitory layer helps normalize the boost from the WM layer as the set size increases. The final term specifies an independent noise source.

 Finally, the gate or “outcome” node (*o*) is given by Equation 8:

8

The first two terms are the stabilization and resting level terms. Next, this node receives stimulus input, *s*(*t*), when a stimulus input is presented to the CF and WM layers. This node also receives a brief transient input at the onset of each stimulus, *strans*(*t*), to capture the strong phasic drive to early visual units at the onset of a stimulus. Next, this unit has self-excitatory input. It also receives excitatory input from the WM layer, . Thus, this unit becomes activated when peaks are present in the WM layer and a new stimulus onset is present. This co-occurrence happens when the test array is presented. The last term in the equation specifies a random noise source.

**LFP Equations**

To simulate hemodynamics directly from the DNF model, we adapted an approach proposed by Deco and Rolls(Deco, Rolls, & Horwitz, 2004). In particular, we generated an estimate of an LFP for each component of the model by summing the absolute value all of the terms in the above equations that contribute to the rate of change in neural activation, excluding the stability terms and resting level parameters. We also excluded the direct stimulus inputs since these were not modeled neurally (i.e., inputs were applied directly to the relevant components of the model and turned “on” and “off” by hand). The resultant LFPs were then convolved with an impulse response function that specifies the slow time course of blood-flow in response to neural activity (Ashby & Waldschmidt, 2008; Deco et al., 2004). The six LFPs we computed are defined below. Note that field activities were aggregated into a single LFP (representing a single neural region). Consequently, we normalized the field contributions over the dimension *x* by dividing by the number of units in each field, . In addition, we multiplied the inhibitory contributions by 0.3 based on data showing that direct stimulation of inhibitory interneurons yields a net positive BOLD signal with an amplitude that is weaker than direct stimulation of excitatory cells (Lee et al., 2010).

9

10

11

13

12

14

The impulse response function, *h*(*t*), used to simulate hemodynamics from the model is defined by Equation 15 (Deco et al., 2004) with the following parameters: n1=7.000, t1=0.875, n2=12.250, t2=1.000, a2=1.6x10-12.

15

**Simulation Methods**

Simulations were conducted in Matlab 7.5.0 (Mathworks, Inc.) on a PC with an Intel® i7 3.33 GHz quad-core processor (the Matlab code is available from the authors on request). For the purposes of mapping model dynamics to real-time, 1 time-step in the model was equal to 2 ms. For instance, to mimic the experimental paradigm reported in the main text, the model was given a set of Gaussian inputs (e.g., 3 colors = 3 Gaussian inputs centered over different hue values) corresponding to the sample array for a duration of 250 time-steps (500 ms). This was followed by a delay of 600 time-steps (1200 ms) during which no inputs were presented. Finally, the test array was presented for 900 time-steps (1800 ms). The response of the model was determined based on which decision node became activated during the test array. The model generated a real-time response on every trial. We simulated 60 same and 60 change trials at each set size for each simulation run. Data were averaged across a batch of 10 runs (i.e., 10 participants) for each task. Note that the resting level of the ‘different’ node was lowered from -5 to -9 in the simulations of the task used by Todd and Marois (2004), reflecting a reduced need to attend to differences given the single-item test array in their task.

 During these simulations, we tracked an LFP for each component of the model (see Equations 9-14). This time-course was then convolved with the impulse response function in Equation 15. The resultant hemodynamic signals were normalized to the maximum mean across components, calculated separately for the 3-layer components and the components of the decision system. Each trial in the event-related design was normalized to start at a value of 0 and an average hemodynamic response across trial types was calculated over a 14 s time-window.

**Varying the Hemodynamic Response Function**

The general patterns of differences in activation were the same between these fMRI studies and the DNF model. One noticeable difference, however, was that the double-humped shape of the hemodynamic response was not evident in the simulation results of Experiment 3 from Magen et al. To further explore how the shape of the hemodynamic response is related to the parameters of the canonical HRF function used to convert the simulated LFP to a hemodynamic response, we manipulated the parameters of the HRF function from Equation 15 above. Figure S1A compares this modified HRF (blue) to the original HRF used in the above simulations (red). Figure S1B-C illustrates the results with this modified HRF for the Magen data and the Todd and Marois data. Notably, the simulated hemodynamic response from the Magen et al. task now shows the characteristic double-humped shape.

**Code Availability**

Matlab code for the model and simulations are available at [www.dynamicfieldtheory.org](http://www.dynamicfieldtheory.org).

Ashby, F. G., & Waldschmidt, J. G. (2008). Fitting computational models to fMRI data. *Behavior Research Methods*, *40*(3), 713–721.

Deco, G., Rolls, E. T., & Horwitz, B. (2004). “What” and “where” in visual working memory: A computational neurodynamical perspective for integrating FMRI and single-neuron data. *Journal of Cognitive Neuroscience*, *16*(4), 683–701.

Jancke, D., Erlhagen, W., Dinse, H. R., Akhavan, A. C., Giese, M., Steinhage, A., & Schoner, G. (1999). Parametric population representation of retinal location: Neuronal interaction dynamics in cat primary visual cortex. *The Journal of Neuroscience*, *19*(20), 9016–9028.

Johnson, J. S., Simmering, V. R., & Buss, A. T. (2014). Beyond slots and resources: grounding cognitive concepts in neural dynamics. *Attention, Perception & Psychophysics*, *76*(6), 1630–54. http://doi.org/10.3758/s13414-013-0596-9

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science*, *20*, 568–577.

Johnson, J. S., Spencer, J. P., & Schöner, G. (2009). A layered neural architecture for the consolidation, maintenance, and updating of representations in visual working memory. *Brain Research*, *1299*, 17–32.

Lee, J. H., Durand, R., Gradinaru, V., Zhang, F., Goshen, I., Kim, D.-S., … Deisseroth, K. (2010). Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature*, *465*(7299), 788–92.

Supplementary Data Tables

**Table S1.** Glossary of mathematical notation used in the DNF model equations.

|  |  |
| --- | --- |
| Letter | Definition |
| *a* | amplitude/strength parameter |
| *x* | dimension (color) |
| *u* | activation variable for CF |
| *v* | activation variable for Inhibitory layer |
| *w* | activation variable for WM field |
| *d* | activation variable for “different” node |
| *m* | activation variable for same or “match” node |
| *o* | activation variable for the gate or “outcome” node |
| *c* | connection weight function |
| *g* | gating function (sigmoidal function) |
| *t* | Time |
| *τ* | timescale parameter |
| *h* | neuronal resting level parameter |
| *η* | number of units along a dimension |
| *r* | random contribution |
| *ξ* | noise variable |

**Table S2.** Parameter values used for the DNF simulations.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Self-Excitation | Inhibition | Interactions | h, τ |
| u | *auu =* 1.5*, σuu =* 3.0 | *auv =* 2.15*, σuv =* 22.0, *auv\_global =* 0.01 | *aud =* 4.0, *aum =* 0.15 | -5, 80 |
| v |  |  | *avu =* 1.7, *σvu =* 10.0,*avw =* 1.95, *σvw =* 5.0 | -12, 10 |
| w | *aww =* 3.3*, σww* = 3.0 | *awv =* 0.325*, σwv =* 35.0, *awv\_global =* 0.148 | *awu =* 1.6, *σwu =* 5.0 *awd =* 1.0, *awm =* 1.0 | -5, 80 |
| d | *add =* 3.7 | *adm =* 100.0 | *adu =* 2.8, *ado =* 4.75 | -5, 80 |
| m | *amm =* 3.7 | *amd =* 100.0, *amv =* 0.2 | *amw =* 0.5, *amo =* 4.75 | -5, 80 |
| o | *aoo =* 4.8 |  | *aow =* 0.03 | -4.92, 80 |

*Note:* β = 5.0 for all terms. Gaussian inputs to the CF layer had a strength of (where SS equals the number of items in the array) and a width of 3. Inputs to the WM layer were weaker, scaled by a factor of 0.2. The external input to the outcome node (o) was the sum of the input to the CF layer scaled by a factor of 0.01. The transient input to the outcome node was present for a duration of 30 ms starting from the onset of a sample or test stimulus. The strength of this input was 18.



Figure S1: Panel A shows the modified hemodynamic response function. Panel B shows the hemodynamic responses from Todd & Marois (2004) on the left and the DF model on the right. Panel C shows the hemodynamic responses from Magen et al. (2009) on the left and the DF model on the right.