UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
A Dynamic Field Theory of Visual Recognition in Infant Looking Tasks

Permalink
https://escholarship.org/uc/item/47853579

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Perone, Sammy
Spencer, John P.
Schoner, Gregor

Publication Date
2007

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

A Dynamic Field Theory of Visual Recognition in Infant Looking Tasks

Sammy Perone (sammy-perone@uiowa.edu) and John P. Spencer (john-spencer@uiowa.edu)
Department of Psychology, 11 Seashore Hall East
Iowa City, IA 52242 USA

Gregor Schöner
Institut für Neroinformatick
Ruhr-Universität-Bochum, Germany

Abstract

Infant looking tasks are fruitful sources of data on infant cognition. Researchers using these tasks measure differential looking times as infants look at multiple stimuli. These looks are used to infer stimulus recognition and visual processing of familiar and novel stimuli. We propose here that looking behavior in two of these tasks, infant categorization and visual paired comparison, arise from a common visual recognition mechanism. We propose a dynamic field theory (DFT) of visual recognition in infant looking tasks and describe simulations using a dynamic neural network that capture looking behavior associated with central phenomena in the infant categorization and infant recognition memory literatures. Critically, these phenomena are supported by a common recognition mechanism.

Keywords: Dynamic Field Theory; Infant Categorization; Visual Recognition Memory; Looking Behavior.

Infant Visual Recognition Tasks

A variety of tasks used to examine infant categorization and infant visual memory can be classified more generally as visual recognition tasks. In a typical categorization task, infants are familiarized with a collection of category exemplars (e.g., land animals), to which their looking typically decreases with the number of exposures. The scope of infants' categories is probed by evaluating looking time to a novel within category exemplar (e.g., tiger) and an out-of-category exemplar (e.g., seal). If infants sustain low looking to the within category exemplar after the familiarization period, they are said to have generalized category membership or recognized the novel exemplar as belonging to the familiar category. Infants are said to have learned an exclusive category if they increase their looking to an out-of-category exemplar, but they are said to have learned an inclusive category if they look at low levels to both a within and out-of-category exemplar.

Visual recognition memory is typically examined in the visual paired comparison task in which infants are typically presented with pairs of identical stimuli for a fixed number of trials or until a looking criterion is met. To test recognition of the familiar stimulus, infants are simultaneously presented with the familiar stimulus paired with a novel stimulus. Preferential looking to one stimulus - often low looking to the familiar stimulus and preferential looking to the novel stimulus - is taken as evidence of recognition of the familiar stimulus.

One important commonality among these tasks is that differential looking behavior is taken as evidence of visual recognition. Although this is the case, the infant categorization and infant visual recognition memory literatures remain isolated, and there has been little discussion of the common processes that might underlie looking in these tasks. We propose here that a common visual recognition mechanism supports looking behaviors in both tasks. Specifically, infant categorization tasks can be classified as tasks that assess whether novel stimuli are recognized as belonging to a familiar category, and visual paired comparison tasks can be classified as tasks that assess whether identical stimuli are recognized as familiar.

We sketch a dynamic field theory (DFT) of visual recognition in infant looking tasks and present simulations that demonstrate looking behaviors commonly observed in infant categorization and visual paired comparison tasks. Moreover, we provide a unified account of a collection of diverse findings obtained from both paradigms. We focus here on two central findings obtained from each task. Specifically, in infant categorization tasks, we focus on the role of perceptual similarity among category exemplars in learning inclusive and exclusive categories, and in visual paired comparison tasks, we focus on preferential looking to a novel stimulus in the context of simultaneously present familiar and novel stimuli.

A Dynamic Field Theory of Visual Recognition in Infant Looking Tasks

The DFT is a process-based theory of visuo-motor decision-making. Thelen, Schöner, Scheier, and Smith (2001) originally proposed the DFT to explain the memory processes that support reaching behavior in the Piagetian A-not-B task. The DFT has been extended by Spencer and colleagues to capture spatial cognition processes and their development (e.g., Schutte, Spencer, & Schöner, 2003). The DFT is formalized in a dynamic neural network that simulates real-time behavior, integrated with the longer time-scale of task-specific learning.

The network described here recognizes a visual stimulus, indexed by looks away from the stimulus toward something more novel, if the network has previously perceptually represented the stimulus and is actively maintaining or can rapidly regenerate a working memory representation of the stimulus. The network consists of five layers of neurons tuned to respond to continuous feature dimensions (e.g., the

color of an object, the object's orientation, and so on). In particular, the network consists of a perceptual layer, $u(x)$, working memory layer, $w(x)$, and a shared layer of (inhibitory) interneurons, $v(x)$. The perceptual and working memory layers are reciprocally coupled to associated long-term memory fields, $u_{ltm}(x)$ and $w_{ltm}(x)$. A bistable looking node is also reciprocally coupled to the perceptual layer.

Figure 1 shows the architecture of the network. We describe here how the network processes and remembers a stimulus and how this process maps onto looking behavior. At the onset of stimulus presentation, the looking node receives transient, excitatory input via an attention-getting stimulus, turning the node from a look away state (i.e., below threshold or "off" state) to a fixation state (i.e., in a self-stabilized "on" state). When the node is in the fixation state, the node passes global excitation to the perceptual layer. Moreover, the look node acts as a perceptual "gate": when this node is fixating the stimulus, excitatory input enters the perceptual field and excites neurons tuned to the stimulus properties present at the fixated location. The perceptual layer, in turn, passes excitatory input back to the looking node, supporting the fixation state while neurons in the perceptual layer are active.

Excited neurons in the perceptual layer activate associated neurons in the perceptual long-term memory layer. The long-term memory traces grow according to a slower time scale; thus, activation in $u_{ltm}(x)$ builds up gradually at stimulated sites and decays slowly when competing inputs are present. Activation in $u_{ltm}(x)$ feeds back onto the perceptual layer, enhancing activation at remembered feature values. Excited neurons in the perceptual layer also activate similarly tuned neurons in the working memory layer, which, in turn, activate associated neurons in the long-term layer associated with working memory.

The final type of cross-layer interaction in the network centers on the shared layer of (inhibitory) interneurons: the perceptual and working memory layers activate similarly tuned interneurons. Reciprocal feedback to and from the inhibitory layer creates a locally-excitatory / laterally-inhibitory type of interaction within the perceptual and working memory layers. If activation is sufficiently strong and localized, these neural interactions can lead to the formation of self-sustaining peaks of activation—the basis for strong working memory representations in the network. Critically, such working memory peaks effectively suppress activation at associated sites in the perceptual layer due to the shared layer of interneurons. This is the basis for visual recognition.

The ability of the network to recognize a stimulus emerges as a long-term memory representation of the input is generated. During each stimulus presentation, activation in the working memory layer is strengthened by activation at associated sites in $w_{ltm}(x)$. This increase in excitation in the working memory layer strengthens activation in the inhibitory layer, which in turn strengthens the inhibitory contribution to the perceptual layer. If the inhibitory contribution is strong enough, it will inhibit further perceptual processing of the input (i.e., suppress activation in the perceptual layer). Consequently, the perceptual layer will no longer support activation of the looking node, the looking node will enter the "off" state, and the network will look away from the stimulus.

To summarize, as a stimulus is repeatedly presented to the network, activation in the working memory layer will become more energetic via facilitation from the accumulating long-term memory trace. Ultimately, this will lead to an increasingly more rapid destabilization of perceptual processing of the repeated input. Consequently, the perceptual layer will support the fixation state for increasingly brief durations as the stimulus or a collection of similar stimuli are repeatedly presented. The low looking behavior of the network to a familiar stimulus—the recognition of the familiar stimulus—is attributable to this dynamic interplay among the multiple layers in the network. In the sections that follow, we demonstrate how these dynamic interactions lead to the looking behaviors infants demonstrate in two canonical visual recognition paradigms.

Recognition of Novel Stimuli

Input Distribution One central finding in the infant categorization literature is that when infants are familiarized with a collection of perceptually similar category exemplars, they learn an exclusive category, but when they are familiarized with a collection of perceptually variable category exemplars, they learn an inclusive category (e.g., French, Mareschal, Mermillod, & Quinn, 2004). In other words, when infants are familiarized with a collection of perceptually similar category exemplars, they recognize a constricted range of within category exemplars as belonging to a familiar category, but when infants are familiarized with a collection of perceptually variable category exemplars, they recognize a broader range of exemplars as belonging to a familiar category. Here, we simulate looking behavior in a categorization context in which exemplars are successively presented in isolation. In this context, exclusive categorization is indexed by an increase in looking to an out-of-category exemplar and inclusive categorization is indexed by sustained low levels of looking to within and out-of-category exemplars (e.g., Oakes, Coppage, & Dingel, 1997).

In our network, recognition of novel stimuli as belonging to a familiar category depends on the distribution of activation in long-term memory. Any long-term memory trace at sites tuned to novel stimuli strengthens activation in the working memory layer, which leads to the rapid generation of a working memory peak and ultimately, via strong inhibition of activation in the perceptual layer, looking away. The more precise the long-term memory trace, the more likely that novel inputs will *fail* to overlap with long-term memory traces leading to longer fixations. The more distributed the pattern of long-term memory activation, the more likely that novel inputs will overlap with long-term memory traces leading to shorter fixations.

To illustrate (1) that the distribution of input contributes to the range of excited neurons in long-term memory, and (2) that the range of neurons excited in long-term memory affects category decisions (i.e., looking behaviors), we

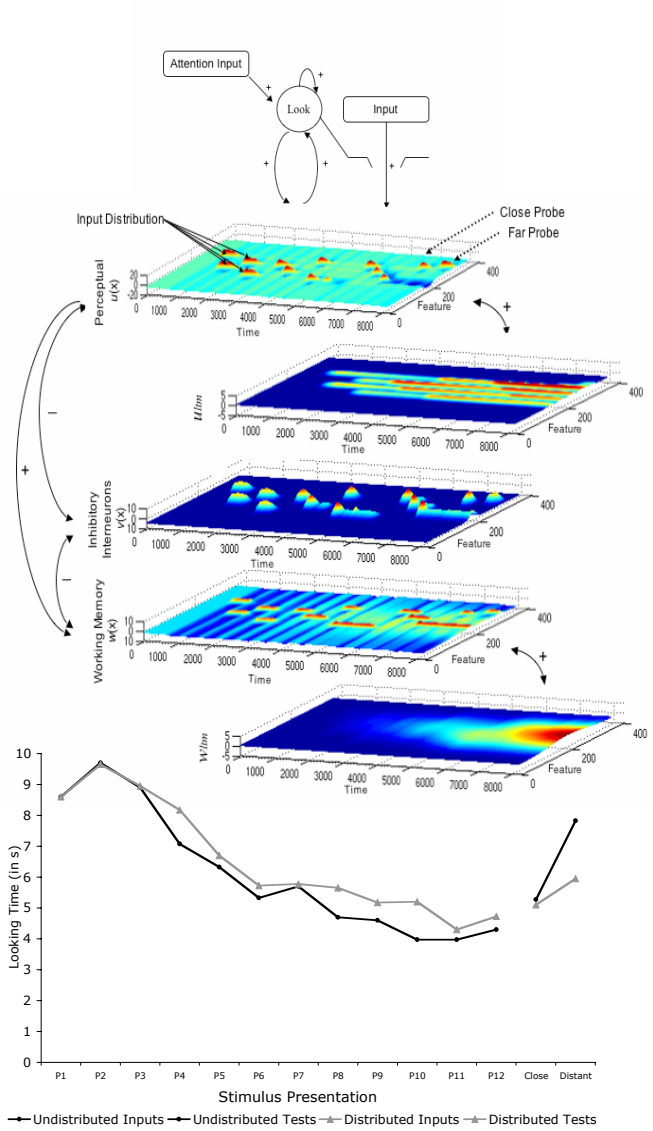


Figure 1. Simulation of one run through an infant categorization task with a distributed set of inputs. Top shows looking node architecture. Middle section shows the 5-layered structure of the DFT and between-layer connections. Note that strongly activated neurons are shown in red and inhibited neurons shown in blue. Bottom panel shows looking time results from the model when familiarized with distributed and undistributed sets of inputs.

simulated familiarization with perceptually similar and perceptually variable stimulus sets. For the perceptually similar set, we presented the network with four relatively undistributed inputs separated by 30 neurons, and for the perceptually variable set, we presented the network with four relatively distributed inputs separated by 40 neurons. In both simulations, every input was presented once in random order across three blocks of four trials.

Figure 2 shows the input distribution (solid lines) and test probes (dotted lines). Notice that the input closest to the test probes in both distributions is identical; thus, any low looking to the test probes cannot be attributed to the distance between familiarization inputs and test probes. In both simulations, recognition was tested with a metrically close test probe 20 neurons from the input shared by both distributions, which we selected to correspond to a within category test, and with a metrically distant test probe 40 neurons from the input shared by both distributions, which we selected to correspond to an out-of-category test.

Figure 1 shows a sample simulation presented with the distributed set of input. The solid lines directed at the perceptual layer show the distribution of inputs presented during familiarization. As can be seen in the figure, a peak of activation is associated with each input and stimulus presentation. As inputs are repeatedly presented, activation in the perceptual layer is sustained for increasingly brief durations. The source of this decrease in activation is the increasing energy of activation in the working memory layer and associated long-term memory.

The dotted lines directed at the perceptual layer toward the right of this simulation show the close and far test probes. As can be seen in the figure, this network shows relatively small perceptual peaks (i.e., low levels of looking) when looking at the close and far test probes, looking behavior consistent with inclusive categorization or recognition of both inputs as familiar. Activation in the perceptual layer cannot be sustained for extended durations because the distributed set of inputs generated excitation across a broad range in $w_{ltm}(x)$, facilitating the rapid generation of a working memory peak and inhibition of perceptual processing.

The bottom panel of Figure 1 shows the looking times accumulated across familiarization trials and test probes for the distributed and undistributed set of inputs. The looking times are means calculated from eight simulations for each distribution; on four of the simulations the metrically close test probe was presented first and on four of the simulations the metrically distant test probe was presented first. Looking was computed as the number of time steps the looking node was above threshold on each trial. Note that fifty time steps equal one second. This time step to seconds mapping was chosen to effectively capture the hypothesized memory processes and keep simulation time reasonable.

The looking plots show that each network exhibited a decrease in looking across stimulus presentations, a typical observation with infants. As discussed above, when the model was presented with the distributed set, looking remained low across both the close and far probes. The network presented with the undistributed set of inputs, in

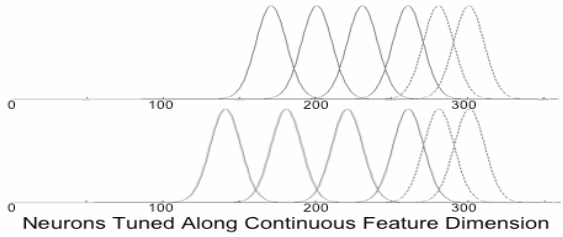


Figure *2*. Solid lines of top panel show the undistributed familiarization set; solid lines of bottom panel show the distributed familiarization set. Dotted lines show test probes.

contrast, exhibited low looking to the close test probe and an increase in looking to the far test probe, looking behavior consistent with exclusive categorization. This looking behavior occurred because the distant test probe excited neurons in the perceptual layer not associated with a strong long-term memory representation. Consequently, the model was not able to build a robust working memory of the far test probe and it continued looking at this novel stimulus. It is important to emphasize that the same network produced the two patterns of looking times shown in Figure 1. The only difference across simulations was in the distribution of the inputs.

Mareschal and colleagues (e.g., Mareschal, French, & Quinn, 2000; French et al., 2004) have extensively examined the role of input distribution in asymmetric categorization and have tested predictions of an autoencoder network. They have shown that the probability that young infants will generalize category membership to out-of-category exemplars depends on whether the distribution of familiar feature values (e.g., ear length, nose width, etc.) is broad and encompasses feature values associated with out-of-category exemplars. The autoencoder recognizes a novel category exemplar as familiar if it can reproduce its inputs, and similarly, the network described here recognizes a novel category exemplar if it can rapidly build a working memory representation of the stimulus. In both networks, recognition depends on the distribution of feature values presented during familiarization. Here, we show input distributed only on one dimension, which approximates the small number of potentially diagnostic dimensions in complex, multidimensional exemplars in infant experiments (see Mareshal et al, 2000). One interesting hypothesis generated by our network is that developments in the precision in which feature values are processed and remembered enhances the likelihood that infants will hone in on the relevant dimensions that differentiate similar visual categories. Older infants have indeed been shown to differentiate categories that young infants treat asymmetrically (Furrer & Younger, 2005).

Input Frequency When infants are presented with a typical category exemplar more frequently during familiarization than a collection of non-typical exemplars, they exhibit evidence of learning an exclusive category. When infants are familiarized with a non-typical category exemplar frequently, however, they exhibit evidence of learning an inclusive category (see Oakes & Spalding, 1997). Here, we illustrate how the metric properties of the input interact with the frequency of stimulus presentation to produce this pattern of looking behavior.

We presented the same network described above with the distributed stimulus set, but in a *central frequent* distribution, the two center inputs were presented three times as often as the end inputs, and in an *end frequent* distribution, the two end inputs were presented three times as often. Note that Oakes and Spalding (1997) used the same frequent and infrequent ratio in experiment. The four inputs were distributed randomly across twelve stimulus presentations, followed by the presentation of a single critical test probe 40 neurons from the closet familiar input.

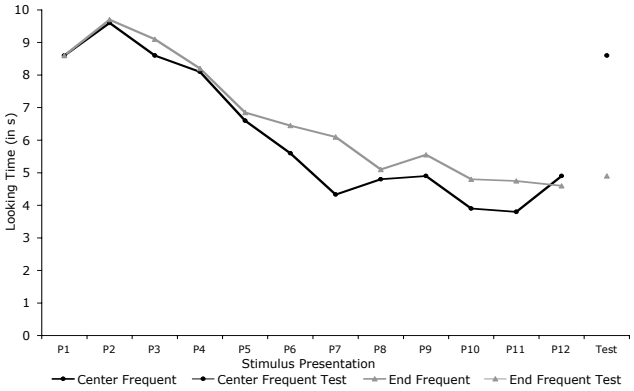


Figure 3

Figure 3 shows the mean looking times of four simulations with central inputs frequent and the mean looking times of four simulations with end inputs frequent. As can be seen, the network presented with the central inputs more frequently increased looking to the test probe, looking behavior consistent with exclusive categorization. The network presented with the end inputs more frequently did not increase looking to the test probe, looking behavior consistent with inclusive categorization. These looking patterns emerge because the metric properties of frequent and non-frequent inputs contributes to the range of active neurons in the long-term memory layers, which influences the probability that the network will recognize a novel stimulus as belonging to a familiar category.

Recognition of Identical Stimuli

Visual recognition of identical stimuli in the visual paired comparison task typically involves familiarization with pairs of identical stimuli followed by a test event in which the familiar stimulus is paired with a novel stimulus. To examine visual recognition in the visual paired comparison task, we implemented a simple, multi-location fixation system (Goldberg, 2005), and to take a first step toward representing multiple feature dimensions, we coupled the fixation system to two identical five-layer networks that each process and remember feature values along a single dimension. Importantly, the network is identical to the network described above but simply engages processing along an additional dimension when a multidimensional stimulus is present. This architecture is consistent with neurophysiological evidence from monkeys and humans that suggest that object properties (e.g., color, size, etc.) are encoded in a distributed manner across different neural populations (Felleman & Van Essen, 1991; Tootle, Tsao,& Vanduffel, 2003).

The fixation system consists of a collection of bi-stable and mutually inhibitory nodes that gate featural input at the fixated spatial locations to the perceptual layers. The top portion of Figure 4 shows a schematic of the fixation system. The dotted line in the schematic shows a fixated two-

dimensional stimulus (i.e., small, dark-grey rectangle) located at a right location. The line projecting from the right node shows that when the right node is fixated, the perceptual gate allows "small" input to enter the first 5-layer network and "dark-grey" input to enter the second 5-layer network. The left bottom portion of Figure 4 shows the real-time behavior of the looking nodes when fixating a left location, followed by fixation at a right location, and finally a look away. This panel also shows the corresponding activation in one of the perceptual layers. Notice that when the left object is fixated, excitation at one feature value is presented to the network, and when the right object is fixated, a different feature value is presented to the network. Also notice that no input is presented to the network when the look away node is fixated and the activation produced by the just-previous look at the right object gradually decays away. To capture the stochastic nature of infant looking among multiple spatial locations, the left and right nodes were presented with noisy inputs associated with the presence of a stimulus at a particular spatial location. Noisy, low-level input was constantly presented to the away node to reflect the continuously available input in the task space (e.g., the edges of a computer monitor, the infant's toes, etc). Noise facilitates spontaneous sampling of objects in the task space. A below threshold node may spontaneously pierce threshold, for instance, and compete for fixation via mutual inhibition. To facilitate competition among the nodes, we also algorithmically implemented a drop in the resting level of the fixated node, which allows the node to be sustained by the perceptual layer but also gives a competitive advantage to the non-fixated nodes (Goldberg, 2005).

Looking behavior in visual paired comparison tasks is clearly more complex than in tasks in which a single stimulus is presented in isolation, and there is a large literature

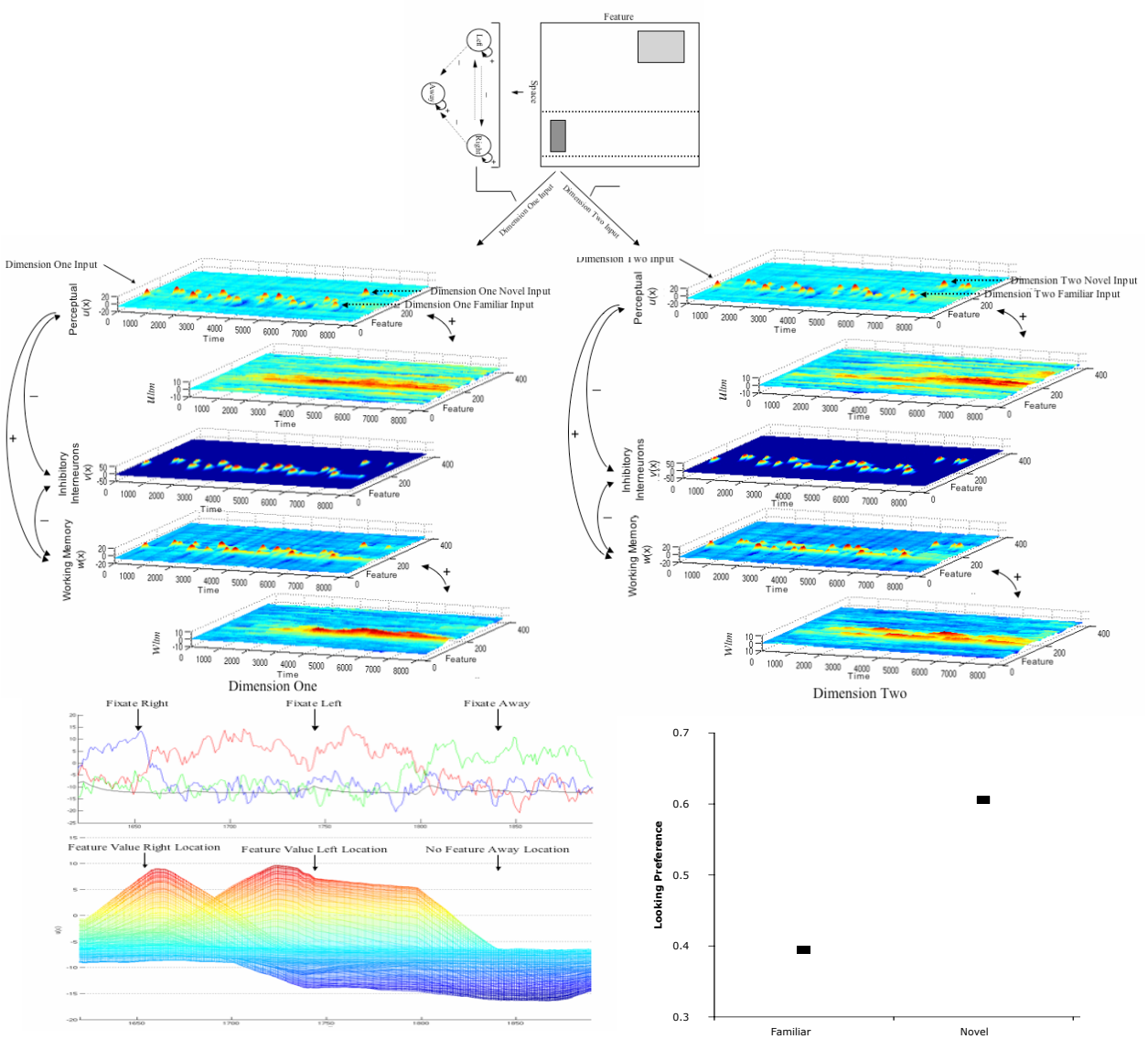


Figure 4. Schematic of fixation system and method of gating input into two 5-layer neural networks. Bottom left panel shows node behavior in real-time and corresponding activation in one perceptual layer. Bottom right panel shows mean preference to look at a familiar and a novel stimulus during test.

examining development in processing speed, novelty detection, and looking dynamics in the visual paired comparison procedure (e.g., see Rose, Feldman, & Jankowski, 2007 for a review). We are currently exploring the ramifications of this complexity for processing and remembering visual stimuli. As a first step in this direction, we simulated a particular pattern of looking behavior in the visual paired comparison task: low levels of looking to a familiar stimulus and preferential looking to a novel stimulus following familiarization. We presented the model with two two-dimensional inputs separated by forty neurons along each dimension. Specifically, when the network fixated one location, neurons tuned to similar values along each dimension were excited, and when the network fixated another location, neurons tuned to different similar values along each dimension were excited. As in real tasks, the spatial location of the inputs occasionally alternated positions, which altered the input gated by each node.

To probe the network for recognition, we presented the network with one familiar stimulus paired with one novel stimulus. The novel stimulus was separated by 80 neurons from the absent stimulus presented during familiarization and 120 neurons from the simultaneously available familiar stimulus. The dotted lines directed at the perceptual layer in Figure 4 show activation peaks generated by the familiar and novel test inputs. The bottom right panel of Figure 4 shows mean familiarity and novelty preferences. A novelty preference was calculated by dividing the total amount of looking time to the novel stimulus on each test trial by the total amount of looking time accumulated to the familiar and novel stimulus on each test trial. On four of the simulations the novel stimulus was presented on the left first, and on four of the simulations the novel stimulus was presented on the right first. As can be seen in the figure, the network exhibited longer looking to a novel stimulus than to a simultaneously presented familiar stimulus.

Why does a novelty preference emerge? When the network fixates the familiar stimulus, a strong peak of activation emerges in working memory supported by long-term memories of the familiar stimulus. Consequently, associated neurons in the perceptual layer are inhibited, de-stabilizing the currently active looking node and generating a look to the novel object. When the network fixates the novel stimulus, in contrast, the activated looking node can sustain fixation because activation in the perceptual layers lingers.

Conclusions

The present report outlines a DFT of visual recognition in infant looking tasks, which posits that a stimulus will be recognized if a working memory representation of the stimulus is actively maintained or can rapidly be regenerated. The DFT further posits that recognition critically depends on the distribution of activation in long-term memory built up during familiarization. Visual recognition in the DFT is indexed by low looking to familiar stimuli and increased or preferential looking to novel stimuli, which we demonstrated can arise from the same recognition mechanism in the categorization and visual paired comparison tasks. Finally, this work highlights how brain and behavior are integrated across the multiple time-scales involved in looking, building perceptual and working memory representations, and generating long-term memory representations in a specific task context.

Acknowledgments

This work supported by NIMH RO1 MH62480 and NSF HSD 0527698 awarded to JPS.

References

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology, 133,* 382-397.

Furrer, S. S. & Younger, B. A. (2005). Beyond the distributional input? A developmental investigation of asymmetry in infants' categorization of cats and dogs. *Developmental Science, 8,* 544-550.

Goldberg, J. (June, 2005). Dynamic systems modeling of infant preferential looking. Poster presented at *Connectionist and Dynamic Systems Approaches to Development,* Iowa City, IA.

Mareschal, D., French, R., & Quinn, P.C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology, 36,* 635-645.

Oakes, L. M., Coppage, D. J., & Dingel, A. (1997). By land or by sea: The role of perceptual similarity in infants' categorization of animals. *Developmental Psychology, 33,* 396-407.

Oakes, L. M. & Spalding, T. (1997). The role exemplar distribution in infants' differentiation of categories. *Infant Behavior and Development, 20, 457-475.*

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2007). Developmental aspects of visual recognition memory in infancy. In L. M. Oakes & P. J. Bauer (Eds) *Short- and Long-term Memory in Infancy and Early Childhood* (pp. 153-178). New York, NY: Oxford University Press.

Schutte, A. R., Spencer, J. P., & Schoner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development, 74,* 1393-1417.

Thelen, E., Schoner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences, 24,* 1-86.