**A dynamic field theory of visual exploration and memory formation in infancy**

John P. Spencer1, Sammy Perone2 & Stacey Shulberg1

1School of Psychology, University of East Anglia, Norwich UK

2Department of Human Development, Washington State University, Pullman, WA USA

**Abstract**

The study of how attention, learning and memory interact in infancy has a long history. Advances in neural modelling have recently pushed this work to a new level, allowing researchers to explore how to conceptualise this interaction in mechanistic terms. Here, we focus on one particular answer to the question of how attention, learning, and memory are integrated mechanistically, that we have developed using the framework of dynamic field theory. We summarise a 20-year journey using a model of visual exploration in infancy to simulate a host of findings from different paradigms, including findings that other models fail to explain. In section 1, we present an overview of this model and the framework of dynamic field theory. In section 2, we focus on how we have used this model to understand changes in visual cognition over development. In section 3, we conclude by looking to the future and the possibility of capturing developmental changes in dyadic visual cognition in the wild. This presents concrete challenges, including the challenge of working with complex models and trying to retain a tight relationship between theoretical and experimental work.

**Keywords (5-10)**:

Infant cognition, attention, working memory, learning, memory, dynamic field theory, neural networks, habituation, speed of processing

**Introduction**

Where we look and attend affects what we see, which in turn constrains what we learn and remember. And learning and memory formation processes, of course, feed back on and influence subsequent looks and bouts of attention. The dynamic interplay among attention, learning and memory have been a central topic in the field of psychology for over a century (James, 1890). Critically, these topics have formed the backbone of infant cognitive research as much of our understanding about infant cognition comes from studies of looking behaviour. Despite this long history, the topic of how attention, learning and memory ***interact*** and ***how to conceptualise this interaction*** ***in mechanistic terms*** is quite current. Indeed, advances in neural modelling have taken these topics to a new level, moving beyond the verbal / conceptual theories that dominate the history of thought in the field (e.g.,Cohen, 1972; Hunter & Ames, 1988).The goal of our chapter is to focus on one particular answer to the question of how attention, learning, and memory are integrated mechanistically that we have developed using the framework of dynamic field theory (Schöner et al., 2016).

The work we present here synthesises a 20-year history of ideas (2003-2023) that has been distributed across more than 30 papers (see References). At the core of this journey is a neural process model of autonomous visual exploration where looking is an action that shapes and is shaped by cognition, that is, by processes that extend beyond the sensory and motor ‘surfaces’ (see Perone & Spencer, 2013). While this may not seem controversial, this simple commitment is not found in many other models of infant looking. For instance, in connectionist models of learning (see French et al., 2004), there is no real-time unfolding of a look because each look is computed as ‘one-shot’ (i.e., in one computation step). By contrast, evidence suggests that looking and learning are fundamentally intertwined. For example, experimental manipulations of how infants distribute their looks on the second-to-second time scale affect what infants learn and remember (Jankowski et al., 2001). Similarly, how infants distribute their looks when exploring objects is influenced by how other agents manipulate the world (Parrinello & Ruff, 1988). This, in turn, impacts how they explore the world and ultimately the development of cognitive and social abilities (Landry et al., 2008; Landry & Chapieski, 1988). Thus, we need formal theories that explain these links between looking and learning.

The dynamic field model of infant visual exploration does exactly this: our autonomous model explores a visual world, looks from location to location, and dwells on some locations longer than others. By treating looking as an action that shapes and is shaped by cognition, our theory explains how measures of looking are related to measures of learning and memory formation in infancy, including empirical effects that no other models explain. We will unpack the central ideas of our model in section 1 of this chapter by providing an overview of DFT, an overview of the concepts captured by the visual exploratory model, and then pointing to key findings we have explained using this framework.

Another central innovation during our 20-year journey has been a focus on development – how does the visual exploratory system change over time, particularly in the first few years of life? There are well-known changes in infant attention and memory formation, going back to early studies of habituation and preferential looking, including increases in speed of processing and enhanced discrimination (Cohen et al., 1971; Colombo & Mitchell, 1990; Rose et al., 2001; 2002). Here, we proposed a central hypothesis – the spatial precision hypothesis – which posits that neural interactions (both excitatory and inhibitory) become stronger and sharper over development (Schutte et al., 2003). When implemented in our model, this captures a wide array of data from habituation (Rose et al., 2002) and visual paired comparison tasks including faster speed of processing and enhanced discrimination (Brannon et al., 2007). Indeed, subsequent studies showed that these developmental changes are linked within individuals (Perone & Spencer, 2014). In section 2, we describe this body of work. We also describe how this strengthening of neural interactions might arise from the accumulation of long-term generalised experience.

A final goal of our 20-year journey has been to think about how autonomous visual exploration and memory formation link with other aspects of development (see, e.g., Bhat et al., 2020). In section 3, we highlight one example which shows how visual exploration is related to social cognition, focusing on dyadic parent-child interactions. We conclude by discussing what this body of work offers, as well as challenges ahead including the use and accessibility of models in developmental science.

**1. A neural process model of visual exploration, attention, and memory formation in infancy**

Here we describe the central elements of our model of autonomous visual exploration in infancy.

**1.1: Introduction to Dynamic Field Theory (DFT)**

DFT proposes that cognition arises from activation within dynamic cortical fields simulated using dynamic fields (DFs). DFs are collections or ‘populations’ of neurons, which are wired up in a way that mimics how neural populations in the brain are wired based on the perceptual, cognitive, or motor dimensions over which they are distributed. For instance, we might have a DF with neurons that represent retinal spatial position (see Figure 1A; e.g., Markounikau et al., 2010). Here, neurons on the left side of the field would have receptive fields that ‘prefer’ inputs at ‘left’ spatial positions, while neurons on the right side of the field would have receptive fields that ‘prefer’ inputs at ‘right’ spatial positions. With this setup, ‘neighbouring’ neurons in the field (i.e., neurons that ‘code for’ similar spatial positions) would excite one another (local excitation), while neurons far apart in the field would inhibit one another (surround inhibition; see Figure 1G-I). This allows activation ‘peaks’ in dynamic fields to represent the metric details of the input pattern. Peaks are defined as activation patterns that rise above an activation threshold of 0. Once activation builds above this threshold, neurons in the field start interacting, both with one another and with neurons in other cortical fields.

For instance, one could present a visual stimulus 20° to the left of the midline . This would activate neurons that ‘prefer’ inputs on the left side of the retina, particularly those that really like the 20° location. The excited neurons would then activate their local neighbours, further driving up activation. At the same time, the excited neurons would inhibit neurons ‘far away’ in the field, preventing excitation from growing at other sites. The result will be a ‘peak’ of activation – a local above-zero ‘bump’ of activity centred at 20° in the field – which faithfully represents the presence of the visual stimulus at this location (see Figure 1A).

By changing the strength of connections in a dynamic field (e.g., by changing the strength of excitatory parameters ‘by hand’), one can create different types of activation patterns through time. For instance, if local excitation and inhibition are moderate, peaks will be stable when inputs are present (e.g., visible), but decay back to a resting level (i.e., no peak) when the input disappears. We call such fields ‘*input-driven*’ (see Figure 1A; Schöner et al., 2016). If, however, excitation and inhibition are strong, peaks can be *self-sustaining* and remain active even if input is removed, acting as a form of working memory to maintain information even when inputs are no longer available (Figure 1B; Schöner et al., 2016; Spencer, 2020). Finally, with the addition of strong global inhibition (see Figure 1I), DFs can be *selective or ‘winner-take-all’*, forming only one peak at a time. Consequently, if two inputs are present, the field will pick one, forming a peak at one input location and suppressing activation at the other input location (Figure 1C).

DF architectures can be constructed by coupling fields together, that is, by having one DF pass activation to another DF and vice versa. This must be done carefully as each DF has receptive fields ‘tuned’ to a particular dimension. For instance, we might have one DF which is sensitive to retinal spatial position and another that is sensitive to colour (i.e., hue). In this case, how would we pass activation from the spatial field to the colour field as these dimensions don’t have any a priori relationship (i.e., blue objects can be on the left or right)? One solution is that the spatial and colour fields can be joined up into a two-dimensional DF where we have neurons ‘tuned’ to all possible combinations of space and colour (see the ‘colour-space’ field in Figure 2). Here, the presentation of a blue object on the left would lead to a peak in the ‘blue-left’ region of a colour-space field while the presentation of a red object on the right would lead to a peak in the ‘red-right’ region of the colour-space field (Figure 2). Interestingly, we can also pass activation to and from this two-dimensional field to separate retinal space and colour fields as shown in Figure 2 (see bi-directional blue arrows). This enables the neural architecture to represent that the object is on the ‘left’ (in the space field), it is ‘blue’ (in the colour field), and the ‘blue object is on the left’ (in the colour-space field). Note that we don’t always have to represent information in multi-dimensional DFs; indeed, there are good reasons not to do this as very high-dimensional DFs have a lot of neurons (more than in the human brain). Thus, we have proposed some rules for how to join dimensions up to create large neural architectures using special ‘binding’ dimensions like space.

So far, we have discussed how DFs capture patterns of neural activation from second-to-second through time. But we also need a way for the patterns of activity to be carried forward over longer periods of time. For instance, how can a pattern of activity on trial 1 impact a future pattern of activity on trial 2 or trial 12? Or how can a pattern of activity be learned over days, weeks, months, and years of experience? The work we describe here uses a variant of Hebbian learning to capture such effects. In particular, ‘memory traces’ can form in DFs when above-threshold peaks build. In this case, the peak boosts activation in a memory trace which feeds back on the field activity, strengthening local excitation in that region of the field. For instance, if the model were asked to encode and respond to the blue item on the left, a memory trace might form which makes the model faster to respond to blue things on the left on future trials (because the memory trace boosts local excitation in this region of the field). This results in a ‘pre-shaping’ effect, facilitating recognition of familiar inputs. By building up memory traces over long periods of time, we can simulate changes over learning and development.

**1.2. Central concepts of the theory**

Here, we introduce the central concepts of our model of visual exploration. This model was first introduced in Perone and Spencer (2013b, 2013a). At the time, we borrowed heavily from two other models: a model of the development of spatial working memory (Schutte & Spencer, 2009; Simmering & Spencer, 2008) and a model of visual working memory (Johnson et al., 2008; Johnson, Spencer, & Schöner, 2009; Johnson, Spencer, Luck, et al., 2009; Simmering, 2008) that is, working memory for object features. Since the initial publications, we have expanded these ideas, incorporating them into a more general theory of visual exploration and visual working memory (Schneegans et al., 2014) Schneegans, Spencer, et al., 2016) as well as a recent model of word learning (Bhat et al., 2021). This latter model is called WOLVES which stands for Word Object Learning via Visual Exploration in Space. The ‘visual exploration in space’ (‘VES’) part of this model is the model from (Schneegans, Lins, et al., 2016; Schneegans, Spencer, et al., 2016). Thus, we’re going to present the central concepts of the visual exploration model using these more recent models.

Figure 3A shows the model architecture. Each box is a dynamic field (except for the box in the top left labelled ‘looking time’). Green arrows show excitatory connections among fields; red arrows show inhibitory connections. We’ll start our tour on the right side of the figure with the ‘visual fields’. There are two visual fields: the top field is a colour-space field; the bottom is a shape-space field. Sample feature values appear along the y-axis of each field. The x-axis is space, in this case, retinal space. Activation in each field is captured by the colour scale, with ‘hotter’ colours showing more intense activation. Looking at the figure, there is a bit of activation at the cyan colour on the left of the colour-space field and activation at the red colour value on the right. In the shape visual field, there is activation at the triangle shape value on the left, and at the star value on the right. This reflects the stimuli the model is exposed to which are shown just above the visual fields – a cyan triangle on the left and a red star on the right. In this figure, we have just started the simulation and exposed the model to the stimulus display on trial one. To preview where we are going, we’re going to embed the model in a ‘speed of processing’ task (Rose et al., 2002). On each trial, the model will see a repeating object over and over (the cyan triangle). This will always be paired with a novel item.

Just above the visual fields is a spatial attention field. The model binds visual features by virtue of their shared spatial positions, implementing a variant of Feature Integration theory (Treisman & Gelade, 1980). Spatial attention plays a key role here. The spatial attention field is a ‘winner-take-all’ field with strong global inhibition. Thus, there can only be one peak in this field at a time (i.e., the model can only attend to one spatial location at a time). As can be seen with the coupling (green arrows), both visual fields pass activation to the spatial attention field and spatial attention passes activation back. To the left of the visual fields are two feature attention fields (‘ATN’). These are the feature analogs to the spatial attention field, that is, they are ‘winner-take-all’ fields and can attend to only one feature at a time. As we will discuss below, the feature attention and spatial attention fields work together to attend to one object at a time – either the cyan triangle or the red star.

Moving leftward along the feature (i.e., ventral) pathways, one can see a contrast field (‘CON’) next with an associated memory trace (‘CON MT’). The contrast fields receive input from the visual fields as well as from the attention fields. Their role is to identify visual features on the retina that are novel, where novelty is defined as ‘not currently in working memory’. In Figure 3A, we’ve highlighted four peaks in blue. As there are multiple peaks, it is hopefully evident that the interactions in this field are not ‘winner-take-all’; rather, this field operates in an ‘input-driven’ mode. The four peaks identify all the features in the visual fields as novel because this is a start of a new simulation and the model has never seen these objects before. These peaks in the contrast fields will start to build associated memory traces in the CON MT layers. This will ‘prime’ these features on subsequent trials, opening the door to a type of familiarity effect we discuss later in the chapter.

The final set of boxes in the model architecture along each feature pathway are the visual working memory fields (‘VWM’) and their associated memory traces (‘VWM MT’). The VWM fields receive weak input from the visual fields and stronger input from the attention fields. These fields have strong local excitation and strong surround inhibition. Consequently, with sufficiently strong input, they can enter a ‘self-sustaining’ mode where peaks are maintained, even when the input is removed. This will be central to our account of visual habituation below. Note that we purposely set the neural interactions in VWM just at the cusp of this self-sustaining mode. This allows the memory traces to play an important role. To preview what is to come, the additional excitatory input to VWM from the memory traces – built by the model’s own experience – will move VWM from the ‘input-driven’ mode to the ‘self-sustaining’ mode. Note also that VWM is inhibitorily connected to the contrast fields; thus, once specific feature peaks are consolidated in VWM, they will suppress the associated feature values in the contrast fields, making these features no longer ‘novel’.

Continuing leftward in Figure 3A, one can see two green dots which are connected to two small circles in the inhibition of return (‘IOR’) field. The small circles represent dynamical nodes in the architecture that detect the presence of peaks in the VWM fields. The idea here is that the model will continue to attend to an object until peaks are consolidated in VWM, that is, until the model has actively represented the features of the object it is attending to. Once consolidation is complete (i.e., peaks are robust in VWM), the model detects this state and the nodes in the IOR field both go above-threshold. Once this happens, this drives the activation of a third node (see third small circle in the IOR field) which globally boosts the IOR field. This boost amplifies the input from spatial attention, building a peak at the current spatially attended location. This IOR peak, in turn, inhibits the associated location in spatial attention, releasing attention from the current object.

In the figures that follow, we walk through the full cycle of identifying which features are novel, attending to one of the novel objects, consolidating the object’s features in VWM, and releasing fixation to look somewhere else. This cycle underlies how this model autonomously visually explores its virtual world. In addition, we highlight how this visual exploratory cycle changes as the model builds stronger memory traces and stronger working memory peaks. This leads to changes in how the model explores the items and the phenomenon of visual habituation.

Figure 3B shows the model about 400ms into the simulation. Note that the box labelled ‘looking time’ in the top left corner shows the model’s looking history, with time shown along the x axis. As highlighted in blue circles, the model is attending to the red star on the right, yielding a peak at the right location in the spatial attention field, and at the red and star values in the feature attention fields. Initially, attention is random – here, the red star happened to win the competition for attention due simply to stronger neural noise around the red star values. Later, however, we’ll see that once the model gets going, cycles of attention are influenced by multiple factors.

Figure 4A shows the model about 400ms later. As can be seen in the looking time plot, the model has been attending to the right location (see positive blue line) for about 400ms. Note that in subsequent figures, there will be a positive red line in the looking time plot when the model looks to the left location. The blue circles highlight that peaks have emerged in the VWM fields. Thus, the model has started to actively represent the red and star feature values in working memory. In Figure 4B, the nodes in the IOR field have gone above threshold with positive activation values. This has boosted the IOR field, forming a peak in this field at the right location (see blue circle). Thus, the model is about to release fixation from the red star. Note the small bumps in the memory trace layers at the red and star feature values. These bumps capture what the model has learned about the right object from this initial bout of visual exploration.

Figure 5A shows the model a few hundred milliseconds later. Here, the model is attending to the cyan triangle on the left (see blue circles in the feature attention fields). Why did the model look at the cyan triangle? Notice that the contrast fields have suppressed activation at the red and star feature values. This suppression is coming from the VWM peaks which are sustaining for a short period of time. With ‘red’ and ‘star’ suppressed, the model has determined that ‘cyan’ and ‘triangle’ are the only novel features in the visual field, so these features quickly capture attention and the model looks to the left. This is more clearly evident in Figure 5B. Here, the model is building robust peaks in VWM at the cyan and triangle feature values, consolidating these items in working memory. Note the red activation timecourse in the looking time plot showing that the model is looking to the left.

In Figure 6A, the model has released fixation from the cyan triangle and has re-attended to the red star (see blue circles in the attention fields). Note that, once again, this happens because the VWM peaks suppress the cyan and triangle features in the contrast fields (see blue circles), so the red and star features are the only novel features the model has detected. Additionally, note that the look to the left (red timecourse in the looking time panel) was a bit longer (~1600 ms) than the look to the right (blue timecourse; ~1200 ms). This difference in duration is driven primarily by neural noise: it look longer for the peaks to consolidate in VWM for the second look due to fluctuations in the strength of activation in working memory. On subsequent trials, more deterministic forces will start to shape these durations.

Figure 6B shows the simulation a few seconds later. At this point, trial 1 has ended and trial 2 has started (see marked time points in the looking time plot). On trial 2, the model is presented with a cyan triangle and a yellow triangle with a slightly different shape. The model is attending to the yellow triangle on the right and starting to consolidate these new features in working memory.

Figure 7A jumps even further into the simulation to the start of trial 4. On this trial, the model is presented with a cyan triangle and a purple square. Note that the last look on trial 3 was to the cyan triangle (see red timecourse in the looking time plot). There are two things notable here. First, the look to the cyan triangle was shorter than the first look to this object – the duration decreased from ~1600ms to ~1200ms. Second, the working memory peaks have survived the inter-trial interval (see blue circles in the VWM fields), supported by the emerging strong memory traces (see blue ovals in the VWM MT layers). Consequently, the model will initially look at the purple square at the start of trial 4. This shows how sustained activation in VWM can influence the initial selection of a look, even on a new trial. Figure 7B show the simulation a bit later on trial 4 when the model is attending to the cyan triangle. Here, the model has just released fixation from the cyan triangle after only 800ms. This is caused by faster consolidation in VWM due to the emerging strong memory traces for the cyan and triangle features.

The final simulation figure in this sequence (Figure 8A) shows the model on trial 6. Here, the red star is presented to the left while the cyan triangle has been moved to the right. We inserted this display to facilitate comparisons with the first trial which is shown in Figure 8B for comparison (typically, the red star would not be re-presented in the speed of processing task). In both panels, the model is looking to the red star, consolidating these features in working memory. Notice, however, that on trial 6, there is lingering activation associated with the initial look to the cyan triangle. This again shows that the working memory for this repeated item is becoming stronger and stronger over trials, supported by the strong memory traces (see VWM MT layers). The simulation in Figure 8 also emphasises that the model is learning about the features of the objects and not about the spatial locations of the objects. That type of learning – learning which object is where – is part of our more recent models (e.g., Bhat et al., 2021, 2023), but does not play a large role in the work discussed in the remainder of this chapter.

**1.3. Linking to empirical evidence**

Initial work with these concepts was done in the simplest possible setting – visual habituation. Here, there is only one thing to look at, and the goal is simply to model the pattern of looking and looking away over time. The left panels of Figure 9 (labelled ‘habituation’) show the basic pattern of looking changes over time in the model. These data come from averages across 200 simulations. On each trial, an object was presented for 20 s with an interstimulus interval of 5 s. We tracked the model’s looking and looking away across 10 trials and then presented a test stimulus. The first test stimulus was a ‘close’ test – an object similar in features to the habituation item. For instance, if the habituation stimulus was a cyan triangle, the ‘close’ test might be a blue triangle. The second test stimulus was a ‘far’ test – an object with very different stimulus features (e.g., a red triangle). We’ll focus on the black lines for now – the behaviour of the young infant model – and discuss development in the next section.

As can be seen in Figure 9A, looking time to the cyan triangle decreased systematically over trials with low levels of looking by trial 10, that is, the model habituated to the stimulus. When we presented the ‘close’ test, the model generalised it’s looking, showing a low-level of looking to the new stimulus. On the ‘far’ test, however, the model dis-habituated, showing long looking to the new item. Figure 9B shows that as looking time decreased, the number of looks per trial increased. Finally, Figure 9C shows that the duration of each look also decreased over trials. These data capture patterns of looking that have been reported across numerous experiments (Colombo & Mitchell, 1990).

What mechanisms in the model lead to these looking patterns? As discussed in the previous section, the decrease in look durations happens as the model builds a memory for the habituation stimulus in the VWM memory trace. This produces faster consolidation in VWM (due to the extra excitation) and shorter look durations. The excitation boost in VWM also leads to more self-sustaining activation patterns in VWM. Consequently, the model detects the stimulus as ‘novel’ less often over time (recall that VWM suppresses activation in the contrast fields), and the model spends more time ‘looking away’ (i.e., looking to some other random location in space – at the edges of the TV monitor, at their feet, and so on).

Figure 10 shows that the model’s entry into the self-sustaining VWM state gives learning a non-linear flavour. The left panels in Figure 10 show the same three measures of looking behaviour over trials as in Figure 9, but here we are highlighting individual simulations. Notice that all the simulations, for instance, decrease looking time over trials, but they do this at different rates, with lots of ‘jumps’ in behaviour over trials (much like infants). Interestingly, we can ‘smooth’ out these individual differences across simulations by re-plotting the same data, not by trial, but by the trial on which peaks in VWM first sustained across the intertrial interval (i.e., 5 s). This is shown in the right panels in Figure 10 (averaged across the full set of 200 simulations). Here, changes over time are more gradual, with still a jump in behaviour around the first trial where peaks in VWM sustained for the first time (see ‘WM’ trial along the x-axis). This highlights that habituation is both a gradual process – VWM consolidation gradually gets faster and faster with the build-up of excitation – and a non-linear process as VWM moves more and more robustly into the self-sustaining state. Note that the linear relationship between looking and memory formation is a cornerstone of early theories of attention in infancy (e.g., Hunter & Ames, 1988). Our model is the first to capture non-linearities in familiarity and novelty preferences (Colombo et al., 1990; Roder et al., 2000), positing that it reflects a shift in the attractor state of the VWM system caused by the build-up of longer-term memory.

One of the key innovations of this visual exploratory model is that looking and learning are mechanistically integrated. This allowed us to explore a central question in the infant literature – how does looking shape learning? For instance, Jankowski and colleagues (2001) showed that experimental manipulations of infant looking changed what infants learned. These researchers did a pretest assessment of infant’s individual learning patterns: short-looking infants initially showed a novelty preference at test, while long-looking infants showed no preference. They then did a second bout of learning. For half of the long-looking infants, they manipulated their looking pattern by presenting a brief stimulus in a quadrant different from where they were looking, driving these infants to show short-looking patterns. At test, the experimentally-manipulated short-looking infants showed a robust novelty preference – manipulating looking affected learning.

Figure 11 shows simulations demonstrating this point with our model. We embedded the model in a visual habituation task with stimuli varying from trial to trial across a small metric range (e.g., a variety of blue-ish / purple-ish items). After 28 trials of exposure, we tested performance on a ‘within’ category test (another blue-ish item) and an ‘outside’ category test (e.g., a red item). Critically, half the models had a small ‘away’ bias – these models got a bit of extra spatial boost to look away, while the other models had a small ‘look’ bias – these models got a spatial boost to look to the stimulus. As can be seen in Figure 11, models with the ‘look’ bias showed a decrease in looking time and look durations over learning, along with an increase in the number of looks. Looking performance in the look ‘away’ bias model remained flat over the learning phase. Critically, at test, we removed the bias for all models. Results in Figure 11 show that the ‘look’ bias models showed low looking to the within-category test and greater looking to the outside-category test, while the ‘away’ bias models failed to distinguish these test items. Thus, a small ‘look’ bias in our model was sufficient to drive better learning. Note that additional analyses showed that this was caused by an acceleration in working memory formation: the extra looking to the stimulus boosted the formation of the VWM memory trace which led to more self-sustaining working memory peaks earlier in learning (Perone & Spencer, 2013b).We know of no other models of visual cognition in infancy that have explained how manipulations of looking impact learning.

**2. Development**

In this section, we describe our efforts to use the visual exploratory model from section 1 to explain developmental changes in attention and memory formation in infancy.

**2.1. Developmental changes in visual cognition: The Spatial Precision Hypothesis**

There are well-known changes in visual attention in infancy. For instance, habituation rates increase during the first year with older infants showing a faster decline in looking time over trials with a faster increase in the number of looks (Colombo & Mitchell, 1990). In addition, older infants show enhanced discrimination abilities at test. For instance, while young infants might generalise low levels of looking to a ‘close’ test item (e.g., showing habituation to a blue item at test when habituated to a cyan item), older infants often show dishabituation (Brannon et al., 2007). These patterns of looking are captured by the light lines in Figure 9 which show simulations of our ‘older’ infant model. Note in particular the differential response to the ‘close’ test item across panels A-C comparing the young and older models.

To capture development in the model, we implemented the spatial precision hypothesis (SPH; Schutte et al., 2003; Schutte & Spencer, 2009). The SPH posits that exposure to and learning about metric dimensions (e.g., space or feature values) results in a strengthening of excitatory and inhibitory neural interactions within cortical fields across early development. We initially implemented this hypothesis ‘by hand’ in our models by changing the strength of excitation and inhibition parameters. For instance, we increased the strength of the locally excitatory interactions in the VWM and contrast fields, also increasing the strength of the laterally inhibitory interactions. This speeds up the dynamic interactions between, for instance, the VWM field and the memory trace. VWM peaks are initially more robust leading to a more robust memory trace. This boosts excitation in VWM further, leading to an early entry into the self-sustaining state. Once the model enters this self-sustaining state, learning continues to accelerate. The result is a faster habituating (i.e., faster learning) model (see Figure 9). But why the enhanced discrimination? Here, we discovered that stronger peaks are also sharper due to the increase in surround inhibition. Consequently, the model can distinguish a working memory of the cyan hue value from a new input at the blue hue value nearby.

It is important to note that the SPH mechanistically links speed of processing and discrimination, that is, the increase in the strength of excitatory and inhibitory interactions over development give rise to both effects. This suggests that we should see correlated changes in speed of processing and discrimination within individuals in infancy. Perone & Spencer (2014) tested this prediction in a visual paired comparison task with 5-, 7-, and 10-month-old infants. Infants were shown pairs of identical stimuli during a familiarisation phase, followed by a test phase where the original item was paired with a similar or ‘close’ test item and with a dissimilar or ‘far’ test item. Results showed shorter look durations with age, faster shift rates with age, and better discrimination with age. Most critically, individual differences in looking predicted discrimination performance on the harder similar test trials in both infants and models, that is, infants and models with faster shift rates showed better discrimination. We know of no other models of visual cognition in infancy that have captured this relationship between speed of processing and discrimination within individuals.

Another notable empirical finding we have explained with the SPH digs into one origin of familiarity effects in early development. Familiarity preferences have long been observed, but the mechanism that underlies such preferences is still poorly understood. Familiarity preferences are more pronounced early in learning and early in development (Gilmore & Thomas, 2002; for a review, see Rose et al., 2004, 2008). Familiarity preferences are typically studied in the visual paired comparison procedure, however, sometimes infants will look longer at a stimulus even when comparison is not directly possible. For instance, infants will sometimes exhibit relatively ***less*** looking to a novel stimulus on one trial relative to a familiar item on a previous trial. Such effects are very difficult to explain. For instance, if longer looking is thought to reflect the error between the current stimulus and the represented stimulus as in connectionist models (French et al., 2004) and predictive coding models (Köster et al., 2020), then why would infants show longer looking to a familiar stimulus over a novel one?

This pattern of longer looking to a familiar stimulus was reported empirically by Wetherford & Cohen (1973). They placed 6*-* to 12-week-old infants in a habituation task, showing them a novel stimulus on trials 2, 9, and 16 (see Figure 12A). Interestingly, 6- and 8-week-old infants looked longer at a familiar stimulus than a novel stimulus over learning with a slight increase in looking times over blocks (Figure 12B,C). We simulated these effects with the visual exploratory model and implemented the SPH to capture the developmental changes. Thus, the younger model had very weak excitatory and inhibitory interactions within the VWM and contrast fields, while older infants had stronger interactions. We also destablized the fixation dynamics of the model with weak excitatory and inhibitory interactions; thus, the model had a hard time robustly fixating an item. Figure 12C and 12D show that the model did a good job capturing the developmental shift from a familiarity bias early in development to a novelty bias later in development.

Why does the model show stronger familiarity early in development? This effect comes from the contrast memory trace. With weak fixation dynamics, the model has a hard time holding fixation on each object. The contrast memory trace helps here as it increases the robustness of peaks in the contrast layer which initially drive attention. Thus, when the young infant model sees a familiar item, it looks longer to this item relative to when it looks at a novel item. We know of no other models that have captured these effects. We note that our visual exploratory model shows other types of familiarity biases in other contexts later in development (see Bhat et al., 2023).Thus, this is an excellent modelling framework for thinking about familiarity and novelty biases more generally, a topic that is receiving renewed attention (see, e.g., Many Babies 5 at <https://osf.io/preprints/psyarxiv/ck3vd>).

We conclude this section by noting that the SPH has also been used to explain developmental changes in other areas of research. For instance, Perone et al. (2011) showed that models with stronger neural interactions in VWM can support the maintenance of more peaks simultaneously (see, e.g., the multiple peaks in VWM in Figure 8). As peaks become stronger, they are able to resist the inhibitory impact of other peaks in the same field. This provides to our knowledge the only developmental account for how a neural system can grow its own capacity over time (see Spencer, 2020) More recent work has extended this account to explain developmental changes in the canonical change detection VWM task used with older children and adults, as well as explaining developmental relationships between the preferential looking and change detection tasks (Simmering, 2016).

We have also used the SPH to explain developmental changes in spatial attention in infancy. With stronger excitatory and inhibitory interactions in spatial attention, older models build activation more quickly (Ross-Sheehy et al., 2015). This leads to faster reaction times in spatial orienting tasks over development, but also more errors when there are conflicting orienting cues and more competition in cases where there is competition between an on-going fixation state and the appearance of a lateral target (Spencer et al., 2022).

**2.2. How does the SPH arise? Creating autonomous changes in visual exploration via long-term generalised experience**

The SPH has captured developmental changes in looking dynamics in the habituation, visual paired comparison, change preference, and spatial orienting paradigms. These developmental changes were captured by hand tuning model parameters to implement the SPH. But what causes these changes in excitatory and inhibitory interactions within the neural system? One hypothesis is these changes arise from the accumulation of memory traces along continuous dimensions. Such accumulation could lead to a homogenous boost of excitation in VWM which, in turn, would yield faster formation of robust, sharper WM peaks, effectively capturing the changes implemented ‘by hand’ in prior work.

Perone and Spencer (2013b) tested this hypothesis in a relatively simple way – by running the visual exploratory model for a very long time. Concretely, we ran the model for 300,000 time steps (divided into 30 episodes of 10,000 time steps). In between each episode, we reset the state of the model (a bit like a sleep episode), carrying forward the memory traces. The model was put in a virtual world with four objects to look at. Over time, we swapped in new objects when the model was looking at another item so it had a good variation of objects to explore. This exposed the model to the entire metric dimension (e.g., all colours).

Periodically, the model was situated in a lab-based speed of processing task designed by Rose et al. (2002) to measure visual processing speed in infancy (see Figures 3-8). The task involves presenting infants with pairs of items, one that does not change (familiar item) and one that does (novel item). Once infants look at the novel item for a greater proportion of time across three consecutive trials, they are said to have processed the familiar item. With age, infants process the familiar stimulus across fewer trials. Our goal was to see if the model would reproduce these longitudinal patterns. This was indeed the case.

What drove this developmental change in visual processing speed? Over development, the model accumulated memories across a continuous dimension that grew in strength with increasingly more experience. This, in turn, led the model to more quickly form robust VWM peaks. When Perone & Spencer (2013b) examined the shape of the WM peaks the model was acquiring over development, they found those peaks had stronger local excitatory and lateral inhibitory components, precisely what was produced when implementing the SPH by hand in prior work. Thus, experience on a daily basis of looking and learning about objects that vary along continuous dimensions appears to be sufficient to produce the mechanistic changes that we hypothesise underlie changes in speed of processing over development.

These model simulations have implications for understanding atypical developmental trajectories and identifying methods that may curtail them. For example, preterm infants showed a developmentally delayed pattern of behaviour in Rose et al.’s (2002) visual processing speed task – they meet the criterion in fewer trials with age but in more trials than term infants of the same age. What accounts for these group differences? Perone and Spencer (2013b) hypothesized this pattern of results could be captured by starting models with initially weaker neural interactions, that is, creating a developmentally less mature model at the start. Aside from this one parameter difference, they then let the preterm infant model develop in an autonomous fashion just as the other, term infant model. Results showed that the initially weaker neural interactions resulted in developmentally delayed visual processing speed. Interestingly, the memory traces over the continuous dimension the model acquired were quite strong, just like the term infant model. However, those memory traces were insufficient to overcome the weaker neural interactions governing consolidation in VWM.

This raised a key question: could an intervention somehow overcome these slightly weaker initial neural interactions? Interventions with preterm infants often involve training parents to ‘follow-in’ on their infants’ visual attention, such as looking at and interacting with the object the infant is visually exploring. This has been shown to promote social, cognitive, and language development (Landry et al., 2006, 2008). It is known that when other agents (e.g., parents) manipulate an object, it influences infants’ looking behaviour (Landry & Chapieski, 1988; Parrinello & Ruff, 1988). Perone & Spencer (2013b) hypothesized that this extra boost of attention – provided by caregivers – might be sufficient to strengthen neural interactions over development, effectively overcoming the initially weaker neural dynamics of the preterm infant model. To explore this intuition, they created a virtual parent who occasionally boosted the spatial input of the object the infant was exploring. Conceptually, this is like a parent manipulating an object to draw or maintain infants’ gaze on an object. Simulation results showed that this simple manipulation led to a stronger homogenous boost in the memory traces the model accumulated which, in turn, led the intervention model to process visual stimuli at a rate comparable to the term infant model – the parent intervention worked.

In summary, this body of work shows how the stronger neural interactions captured by the spatial precision hypothesis can arise via generalised long-term experience. It also suggests that the details of this long-term experience – for instance, starting with an immature cortical system or interacting with a caregiver – matter.

**3. Conclusions: Looking to the future**

The work we have summarised here offers one solution to how attention, learning and memory are mechanistically integrated. This work places looking and learning at the heart of infant visual cognition, revealing the subtle ways in which these systems interact over multiple timescales. Across a 20-year journey, we have unpacked these relationships, anchoring the theory to empirical evidence. In some cases, this evidence includes findings which other models fail to explain. In this sense, the model makes unique theoretical contributions across the fields of visual attention, spatial attention, and visual working memory in infancy. We have also embedded the visual exploratory account in recent models that focus on other aspects of cognitive development such as early word learning (Bhat et al., 2021, 2023). This shows that the theory generalises well.

What does the future hold? One direction within reach is work examining dyadic interactions in infancy. For instance, the virtual parent model simulations described above show developmental trajectories in looking behaviour are influenced by the presence of others. This was a good first step in capturing the role of parents in neurodevelopment, but it does not capture important aspects of infants’ social experience when interacting with caregivers in the wild. For example, across the first year of life, parents and their infants develop mutually synchronous behaviour, a temporal coordination between parent and infant through time (Feldman, 2012; Feldman et al., 1999). Initially, parents tend to follow into attention episodes with infants leading the interactions, but over time, each member of the dyad learns to repair their state to match that of the other, creating what is often referred to as a dance between parent and infant (Feldman, 2007).

Can the visual exploratory model shed light on such observations? For instance, could we create an autonomous parent-infant model that learns about the world and each other? Such a model should be within reach because our visual exploratory model has previously been used to capture visual cognitive dynamics in infants, children, and adults (Perone & Simmering, 2019; Simmering, 2016; Johnson et al., 2022). To explore this, we created an infant model that looked at a virtual world consisting of five locations with object features sampled from a continuous dimension as before (see Figure 13). The architecture of the parent model was identical to the infant model with one exception – it was equipped with an action system. The action system was coupled to the looking system such that when the parent model looked at an object in the virtual world, this would activate the action system to act on the fixated object with some probability (e.g., shaking a toy). This, in turn, would create extra input at a specific location in the virtual world, biasing both models to fixate that location. This is a great way for infants to follow the lead of their parent.

But how do the parent and infant learn about each other to develop a mutually synchronous style of interaction? To accomplish this, we let the direction of gaze of each member be input to the other, such that when the infant was looking to the left, the parent was slightly biased to look at the left (via stronger input at this location in the shared virtual world). In addition, when they shared gaze at the same location, we increased the strength of a bias node (see Figure 13). The bias node did not track anything about specific spatial locations. Rather, it simply increased the strength of input from the spatial dimension. Consequently, each time the dyad shared joint attention, the joint model became a little more biased to re-enter a joint attention state in the future.

We situated the dyadic model in a virtual world and let it freely explore objects together across 30 episodes of 10,000 timesteps as before. In addition, we created two types of parent models based on evidence that parental responsiveness impacts early visual cognition. Parental responsiveness captures the degree to which a parent will provide prompt and supportive responses to their infant’s behaviour (Ainsworth et al., 1978). To implement different levels of responsiveness, we manipulated how strongly the infant model’s gaze direction was input to the parent model. In our responsive parent model, the parent model had a slightly stronger strength of this infant input; in the unresponsive parent model, the parent model had a weaker strength of this input from the infant.

To quantify parent-infant interactions, we used a metric called recurrence rate. Higher values indicate that models looked to the same location frequently; lower values indicate a lower frequency of joint looking to the same location. As can be seen in Figure 14A, the recurrence rate for the the dyadic model with the responsive parent (grey line) quickly rose to high levels. For the unresponsive model (black line), recurrence was low and did not increase dramatically over time.

Another key result centres on the lead-follow dynamics. The lead-follow dynamics of the dyadic model can be quantified by the percentage of looks where the infant model looked at an object first and then the parent followed. Values greater than .5 indicate the proportion of infant-led interactions; values below .5 indicate the proportion of parent-led interactions (or infant-follow). Values around .5 indicate mutual synchrony (i.e., sometimes the parent leads and sometimes the infant leads). For the responsive parent model (grey line), the infant initially led most of the bouts, but then slowly the dyad began to develop mutually synchronous behaviour. This happens because both the infant and the parent are learning about each other through the accumulation of their experience sharing gaze over development (via the bias node in Figure 13). The responsive parent is initially biased to follow the infant, and the infant learns to follow the parent over time. For the dyad with the unresponsive parent (black line), the parent leads and the infant follows early in the simulation. This pattern becomes even more extreme over time as the infant model learns to follow the direction of the parent’s gaze.

Finally, we asked a central question: does sharing a virtual world with a responsive parent have consequences for infant learning? Our simulations indicate the answer is ‘yes’. Figure 15 shows the VWM memory traces acquired by the infant model over time. The memory traces acquired by the infant model paired with the responsive parent (Figure 15A) were stronger than those acquired by the infant paired with the unresponsive parent (Figure 15B).

These dyadic simulations are presented here as a proof of concept, to demonstrate what is possible with the visual exploratory model. We think the results are compelling and suggest that visual cognition ‘in the wild’ might be within reach of this framework. Nevertheless, to achieve this goal, we need a tighter relationship to real-world visual cognition in infancy. Recent work has moved in this direction by using audio-visual recordings from infants in naturalistic contexts combined with a powerful (and complex) deep-learning network (Vong et al., 2024). It may be possible to do something similar with DFT.

Although exciting, there are many challenges to moving in this direction. We will undoubtedly need a more complex theory. For instance, our colleagues have developed an elaborated dynamic field theory of visual attention that captures a host of findings from the field of adult visual attention (Grieben et al., 2020). This theory does an excellent job quantitatively capturing key findings that other theories have difficulty explaining. It’s also a more complex theory relative to the simple visual exploratory model presented here. This makes it more complicated to understand and harder to simulate, which creates barriers to uptake, particularly for researchers who find working with models already daunting. The field of infant cognition will have to find ways through these barriers if we want to use more complex theories and still keep theoretical and empirical work integrated step-by-step as our science progresses.

**References**

Ainsworth, M., Blehar, M., Waters, E., & Wall, S. (1978). *Patterns of Attachment*. Lawrence Eribaum Associates. Inc., Publishers.

Ambrose, J., Schneegans, S., Schoner, G., & Spencer, J. P. (2016). A ‘how-to’ guide to modeling with dynamic field theory. In G. Schoner, J. P. Spencer, & The DFT Research Group (Eds.), *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford University Press.

Bhat, A. A., Samuelson, L. K., & Spencer, J. P. (2023). Formal theories clarify the complex: Generalizing a neural process account of the interaction of visual exploration and word learning in infancy. *Child Development*, *94*(6), 1491–1510. https://doi.org/10.1111/cdev.14023

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2020). WOLVES: A Model of Children’s Response to Novelty and Word Learning. *Manuscript in Preparation*.

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2021). Word–Object Learning via Visual Exploration in Space (WOLVES): A Neural Process Model of Cross-Situational Word Learning. *Psychological Review*. https://doi.org/https://doi.org/10.1037/rev0000313

Brannon, E. M., Suanda, S., & Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Developmental Science*, *10*(6), 770–777. https://doi.org/10.1111/j.1467-7687.2007.00635.x

Cohen, L. B. (n.d.). *A Two Process Model of Infant Visual Attention*.

Cohen, L. B., Gelber, E. R., & Lazar, M. A. (1971). Infant habituation and generalization to differing degrees of stimulus novelty. *Journal of Experimental Child Psychology, 11*, 379–389.

Colombo, J., & Mitchell, D. W. (1990). Individual differences in early visual attention: Fixation time and information processing. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, and prediction* (pp. 193–227). Hillsdale, NJ: Lawrence Erlbaum.

Colombo, J., Mitchell, D. W., Coldren, J. T., Atwater, J. D., Etzel, B. C., Horowtiz, F. D., Morris, E. K., & Rovee, C. K. (1990). Discrimination Learning During the First Year: Stimulus and Positional Cues. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 16, Issue 1).

Colombo, J., & Mitchell, Wayne. D. (1990). Individual differences in early visual attention: Fixation time and information processing. In J. Colombo & J. Fagen (Eds.), *individual differences in infancy: Reliability, Stability, and Prediction* (pp. 193–227). Lawrence Erlbaum.

Feldman, R. (n.d.). *PARENT-INFANT SYNCHRONY: A BIOBEHAVIORAL MODEL OF MUTUAL INFLUENCES IN THE FORMATION OF AFFILIATIVE BONDS*.

Feldman R. (2007). Parent-infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *Journal of child psychology and psychiatry, and allied disciplines*, *48*(3-4), 329–354. <https://doi.org/10.1111/j.1469-7610.2006.01701.x>

Feldman, R., Greenbaum, C. W., & Yirmiya, N. (1999). Mother-infant affect synchrony as an antecedent of the emergence of self-control. *Developmental Psychology*, *35*(1), 223–231. https://doi.org/10.1037/0012-1649.35.1.223

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: General*, *133*(3), 382–397. https://doi.org/10.1037/0096-3445.133.3.382

Gilmore, R. O., & Thomas, H. (2002). Examining individual differences in infants’ habituation patterns using objective quantitative techniques. *Infant Behavior and Development, 25*, 399–412.

Grieben, R., Tekülve, J., Zibner, S. K. U., Lins, J., Schneegans, S., & Schöner, G. (2020). Scene memory and spatial inhibition in visual search. *Attention, Perception, and Psychophysics*, *82*, 775–798.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. O. Lipsitt (Eds.), *Advances in infancy research* (pp. 69–80). Albex.

James, W. (1890). *The Principles of Psychology*. Holt.

Jankowski, J. J., Rose, S. A., & Feldman, J. F. (2001). Modifying the distribution of attention in infants. *Child Development*, *72*(2), 339–351. http://www.ncbi.nlm.nih.gov/pubmed/11333070

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*(5), 568–577. https://doi.org/10.1111/j.1467-9280.2009.02329.x

Johnson, J. S., Spencer, J. P., & Schöner, G. (2008). Moving to a higher ground: the dynamic field theory and the dynamics of visual cognition. *New Ideas in Psychology*, *26*, 227–251.

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. https://doi.org/10.1016/j.brainres.2009.07.008

Johnson, J. S., van Lamsweerde, A. E., Dineva, E., & Spencer, J. P. (2022). Neural interactions in working memory explain decreased recall precision and similarity-based feature repulsion. *Scientific Reports*, *12*(1), 1–18. https://doi.org/10.1038/s41598-022-22328-4

Köster, M., Kayhan, E., Langeloh, M., & Hoehl, S. (2020). Making Sense of the World: Infant Learning From a Predictive Processing Perspective. *Perspectives on Psychological Science*, *15*(3), 562–571. https://doi.org/10.1177/1745691619895071

Landry, S. H., & Chapieski, M. L. (1988). Visual Attention During Toy Exploration in Preterm Infants: Effects of Medical Risk and Maternal Interactions. In *INFANT BEHAVIOR AND DEVELOPMENT* (Vol. 11).

Landry, S. H., Smith, K. E., & Swank, P. R. (2006). *Responsive Parenting: Establishing Early Foundations for Social, Communication, and Independent Problem-Solving Skills*. https://doi.org/10.1037/0012

Landry, S. H., Smith, K. E., Swank, P. R., & Guttentag, C. (2008). A Responsive Parenting Intervention: The Optimal Timing Across Early Childhood for Impacting Maternal Behaviors and Child Outcomes. *Developmental Psychology*, *44*(5), 1335–1353. https://doi.org/10.1037/a0013030

Markounikau, V., Igel, C., Grinvald, A., & Jancke, D. (2010). A dynamic neural field model of mesoscopic cortical activity captured with voltage-sensitive dye imaging. *PLoS Comput Biol*, *6*(9). https://doi.org/10.1371/journal.pcbi.1000919

Parrinello, R. M., & Ruff, H. A. (1988). The Influence of Adult Intervention on Infants’ Level of Attention. In *The Infiuence of Adult Intervention on Infants’ Level of Attention CHILD DEVELOPMENT* (Vol. 59).

Perone, S., & Simmering, V. R. (2019). Connecting the dots: Finding continuity across visuospatial tasks and development. *Frontiers in Psychology*, *10*(JULY). https://doi.org/10.3389/fpsyg.2019.01685

Perone, S., Simmering, V. R., & Spencer, J. P. (2011). Stronger neural dynamics capture changes in infants’ visual working memory capacity over development. *Developmental Science*, *14*(6), 1379–1392. https://doi.org/10.1111/j.1467-7687.2011.01083.x

Perone, S., & Spencer, J. P. (2013a). Autonomous visual exploration creates developmental change in familiarity and novelty seeking behaviors. *Frontiers in Psychology*, *4*(648), 648. https://doi.org/10.3389/fpsyg.2013.00648

Perone, S., & Spencer, J. P. (2013b). Autonomy in Action: Linking the Act of Looking to Memory Formation in Infancy via Dynamic Neural Fields. *Cognitive Science*, *37*(1), 1–60. https://doi.org/10.1111/cogs.12010

Perone, S., & Spencer, J. P. (2014). The Co-Development of Looking Dynamics and Discrimination Performance. *Developmental Psychology*, *50*(3), 837–852. https://doi.org/10.1037/a0034137

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants’ Preferences for Familiarity and Novelty During the Course of Visual Processing. In *INFANCY* (Vol. 1, Issue 4).

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Attention and recognition memory in the 1st year of life: A longitudinal study of preterm and full-term infants. *Developmental Psychology, 37*, 135–151.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2002). Processing speed in the 1st year of life: a longitudinal study of preterm and full-term infants. *Developmental Psychology*, *38*(6), 895–902. https://doi.org/10.1037/0012-1649.38.6.895

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2004). Infant visual recognition memory. *Developmental Review*, *24*(1), 74–100. https://doi.org/10.1016/j.dr.2003.09.004

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Rossem, R. Van. (n.d.). *A Cognitive Cascade in Infancy: Pathways from Prematurity to Later Mental Development*.

Ross-Sheehy, S., Schneegans, S., & Spencer, J. P. (2015). The Infant Orienting With Attention Task: Assessing the Neural Basis of Spatial Attention in Infancy. *Infancy*, *20*(5), 467–506. https://doi.org/10.1111/infa.12087

Schneegans, S., Lins, J., & Spencer, J. P. (2016). Integration and selection in dynamic fields: Moving beyond a single dimension. In G. Schoner, J. P. Spencer, & the D. F. T. R. Group (Eds.), *Dynamic Thinking: A Primer on Dynamic Field Theory* (pp. 121–149). Oxford University Press.

Schneegans, S., Spencer, J. P., & Schöner, G. (2016). Integrating ‘ what ’ and ‘ where ’: Visual working memory for objects in a scene. In G. Schoner, J. P. Spencer, & The DFT Research Group (Eds.), *Dynamic Thinking: A Primer on Dynamic Field Theory* (pp. 197–226). Oxford University Press.

Schneegans, S., Spencer, J. P., Schoner, G., Hwang, S., Hollingworth, A., Schöner, G., Hwang, S., & Hollingworth, A. (2014). Dynamic interactions between visual working memory and saccade target selection. *Journal of Vision*, *14(11)*(9), 1–23. https://doi.org/10.1167/14.11.9

Schöner, G., Spencer, J. P., & The DFT Research Group. (2016). *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford University Press.

Schutte, A. R., & Spencer, J. P. (2009). Tests of the dynamic field theory and the spatial precision hypothesis: capturing a qualitative developmental transition in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1698–1725. https://doi.org/10.1037/a0015794

Schutte, A. R., Spencer, J. P., & Schöner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development*, *74*(5), 1393–1417. https://doi.org/10.1111/1467-8624.00614

Simmering, V. R. (2008). *Developing a magic number: The dynamic field theory reveals why visual working memory capacity estimates differ across tasks and development* [PhD]. University of Iowa.

Simmering, V. R. (2016). Working memory capacity in context: Modeling dynamic processes of behavior, memory, and development. *Monographs of the Society for Research in Child Development*, *81*(3), 7–24. https://doi.org/10.1111/mono.12249

Simmering, V. R., & Spencer, J. P. (2008). Generality with specificity: The dynamic field theory generalizes across tasks and time scales. *Developmental Science*, *11*(4), 541–555. https://doi.org/10.1111/j.1467-7687.2008.00700.x

Spencer, J. P. (2020). The Development of Working Memory. *Current Directions in Psychological Science*, *29*(6), 545–553. https://doi.org/10.1177/0963721420959835

Spencer, J. P., Ross-Sheehy, S., & Eschman, B. (2022). Testing predictions of a neural process model of visual attention in infancy across competitive and non-competitive contexts. *Infancy*, *27*(2), 389–411. https://doi.org/10.1111/infa.12457

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Vong, W. K., Wang, W., Orhan, A. E., & Lake, B. M. (2024). Grounded language acquisition through the eyes and ears of a single child. *Science*, *383*(6682), 504–511. https://doi.org/10.1126/science.adi1374

Wetherford, M. J., & Cohen, L. B. (1973). Developmental Changes in Infant Visual Preferences for Novelty and Familiarity. *Child Development*, *44*(3), 416. https://doi.org/10.2307/1127994

**Acknowledgements**. Support for the writing of this chapter was provided by grant RPG-2021-350 from the Leverhulme Trust awarded to JPS.

**Figure Captions**

**Figure 1.** The graphs in the top row (a, b, c) show how activation (z-axis) evolves through time (y-axis) in the dynamic field across locations in retinal space (x-axis) as inputs are turned on and off. The blue arrow in each panel shows when input is turned on; the red arrow shows when input is turned off. The panels in the middle row (d, e, f) show the state of the field activity at the last time step. The dark blue line shows the activation level (y-axis) over retinal space (x-axis), the red line shows which neurons are engaged in neural interactions (i.e., above zero activity), and the cyan line shows the strength of the memory trace. The graphs in the bottom row (g, h, i) show the rule governing how neurons talk to one another, with local excitation around the activated site (0) and surround inhibition to the left and right (at farther distances in retinal space). The simulation in the left column shows an ‘input-driven’ DF with moderate excitation and inhibition. The simulation in the middle column shows a ‘self-sustaining’ DF with strong excitation/inhibition where activity is sustained even though the input is turned off. The simulation in the right column shows a ‘selective’ DF with strong excitation and global inhibition. Here, the field ’selects’ one input even though two inputs are presented (see panel c). Figure from Bhat et al. (2023). Used with permission under a CC Attribution-NonCommercial-NoDerivs license.

**Figure 2.** An example of a DF model with a two-dimensional color-space field which shares activation with a retinal space field (top) and a color field (left). Input from the visual display is passed directly into the two-dimensional field, mapping the location of the input along both the spatial dimension (15 degrees; x-axis) and the color dimension (hue value 120 = ’blue’; y-axis). The intensity of activation in the 2D field is captured by the color value with ‘hotter’ colors showing more intense activation. Activation in the color-space field then projects activation to the retinal space field (top) and the color field (left), building a peak in those fields (see activation profile in blue; y-axis). The red curve in the space and color fields shows which neurons are above 0 activation (i.e., ‘active’ or ‘on’). These peaks then pass activation back to the 2D field, further amplifying the activation pattern in the color-space field. Figure from Bhat et al. (2023). Used with permission under a CC Attribution-NonCommercial-NoDerivs license.

**Figure 3**. The top panel (a) shows the model architecture as the model is attending to novelty (see contrast fields; ‘CON’). Green arrows indicate excitatory connections; red arrows indicate inhibitory connections. Connectivity for the shape dimension is the same as for the colour dimension. Blue circles highlight key aspects of the activation patterns (see text for details). Looking time plot in the top left shows the time series of looking during the simulation. Bottom panel (b) shows activation as the model is attending to the red star.

**Figure 4.** The top panel (a) shows the model consolidating visual features in visual working memory (see text for details). Bottom panel (b) shows the model releasing fixation via activation in the inhibition of return (IOR) field.

**Figure 5**. Top panel (a) shows the model attending to the cyan triangle because the features of the red star are suppressed by activation in visual working memory. Bottom panel (b) shows the model consolidating the features of the cyan triangle in visual working memory.

**Figure 6**. Top panel (a) shows the model attending to the red star because the features of the cyan triangle are no longer novel. Bottom panel (b) shows the model attending to the yellow triangle at the start of trial 2.

**Figure 7.** Top panel (a) shows the working memory for the cyan triangle sustaining at the end of trial 3. Bottom panel (b) shows a short look to the cyan triangle due to fast consolidation in visual working memory.

**Figure 8.** Top panel (a) shows the model at the start of trial 6 with a sustained working memory for the cyan triangle. Bottom panel (b) is a reproduction of Figure 4A to facilitate comparisons late in learning (a) versus early in learning (b).

**Figure 9**. The looking behavior of the young infant model (black lines) and old infant model (gray lines) during the habituation and test phases averaged across simulations. The young infant model is slower to exhibit a decline in looking (a), exhibits fewer looks (b), and exhibits longer look durations (c) across trials than the old infant model. The young infant model exhibits an increase in looking time to the ‘far’ test, but not the ‘close’ test, relative to the last habituation trial. The old infant model exhibits an increase in looking time to both the close and far tests.

**Figure 10**. A sample of the looking time (a), number of looks (b), and look duration during the habituation phase from five simulations of the old infant model. Each line is the behavior

of a different simulation. The looking behavior of the model fluctuates across trials and simulations. Panels to the right highlight the linkage between looking and the time course of stable WM peak formation. Here, looking measures are anchored to the trial on which a stable WM peak was formed (see ‘WM” along the x-axis). Black lines show models that acquired a WM peak on trial 2; light lines show models that acquired a WM peak on trial 5. Anchoring looking measures to the trial of working memory formation removes the simulation to simulation variability evident in panels a-c. There is also a clear non-linear shift in panels d-f during the trial after working memory is first formed (see ‘WM’ trial).

**Figure 11**. The looking behavior of ‘look bias’ (black) and ‘away bias’ (grey) models. The look bias model exhibited more looking time across the learning phase than the away bias model (a), as well as an increase in the number of looks (b) and a decrease in look durations (c). The look bias model recognized the within stimulus test item and renewed looking to the outside stimulus (better discrimination), whereas the away bias renewed looking to both the within and outside test items.

**Figure 12**. (a) Experimental design from Wetherford and Cohen (1973). Infants were habituated to a single stimulus (star) across seventeen 15-s trials with ISIs of 8 s. On trials

2, 9, and 16, a different novel stimulus was presented (e.g., heart). A preference to look at the novel stimulus was calculated by subtracting looking time to the novel stimulus from looking time on the preceding trial with the familiar, habituation stimulus. (b) Looking time across seven blocks of two trials from 6- (light grey solid line, circles), 8- (light grey dotted line, triangles), 10- (dark grey dotted line, triangles), and 12-week-old (black lines, circles) infants. Six- and 8-week-old infants exhibited no evidence of habituation, while 10- and 12-week old infants show a decline in looking. (c) Six-week-olds showed a trend toward a familiarity preference on the third novel test, and 8-week-olds exhibited a familiarity preference on the second and third novel tests. Ten- and 12-week-olds exhibited a novelty preference on the second and third novel tests. The model produced the same looking behavior across trials (D) and preferences at test (E).

**Figure 13**. Cartoon depicting the parent-infant model interactions over the shared space. Models were designed as in prior figures with a visual field (see ‘virtual world’) and contrast and VWM fields; however, we modified how they interacted over space. (a) When the infant model looked at a location (see location 3), this boosted the spatial input at that location for the parent (see thicker green arrow pointing to location 3 for the parent). Similarly, when the parent looked at a location (see location 1), this boosted the spatial input at that location for the infant (see thicker input arrow at location 1 for the infant). The parent also had an action system (see ‘act’). When the parent looked to a location (see ‘fix’), the parent might also jiggle the object with some probability which further boosted the spatial input at that location for the infant. (b) We also introduced a bias node (‘B’). The bias node increased whenever the parent and infant looked to the same location. This further amplified the spatial inputs back and forth over learning. See text for additional details.

**Figure 14**. (a) Dyadic models with a responsive (grey lines) parent showed a steep increase in recurrence rate over learning (looking to the same object within a short window of time); dyadic models with an unresponsive (black lines) parent showed a moderate increase in recurrence rate. (b) Dyadic models with a responsive parent showed initially infant-led interactions, moving toward synchronous (infant lead = .5) interactions over time; dyadic models with an unresponsive parent showed initially parent-led interactions, moving toward more parent-led interactions over time.

**Figure 15**. Panels show the memory trace strength (y-axis) from the visual working memory layer for the infant models after different amounts of autonomous dyadic visual exploration: early in blue (after 10 episodes of 10,000 time steps), middle in red (after 20 episodes), and late in cyan (after 30 episodes). The infant model paired with the responsive parent has stronger memory traces, with more detailed structure. Thus, the infant from the responsive dyad has learned more.