Models provide specificity: Testing a proposed mechanism of visual working memory capacity development

Vanessa R. Simmering a,b,*, Rebecca Patterson a

a Psychology Department and Waisman Center, University of Wisconsin – Madison, United States
b Eye Research Institute, University of Wisconsin – Madison, United States

A R T I C L E   I N F O

Keywords:
Capacity
Visual working memory
Neural field model
Developmental mechanism
Memory development
Inhibition
Computational model

A B S T R A C T

Numerous studies have established that visual working memory has a limited capacity that increases during childhood. However, debate continues over the source of capacity limits and its developmental increase. Simmering (2008) adapted a computational model of spatial cognitive development, the Dynamic Field Theory, to explain not only the source of capacity limitations but also the developmental mechanism. Capacity is limited by the balance between excitation and inhibition that maintains multiple neural representations simultaneously in the model. Development occurs according to the Spatial Precision Hypothesis, which proposes that excitatory and inhibitory connections strengthen throughout early childhood. These changes in connectivity result in increasing precision and stability of neural representations over development. Here we test this developmental mechanism by probing children’s memory in a single-item change detection task. Results confirmed the model’s predictions, providing further support for this account of visual working memory capacity development.

Published by Elsevier Inc.

Visual working memory (VWM) provides a critical foundation for our understanding of the visual world around us. Without the ability to represent visual information as we move our eyes around the world, our experience would be a series of disjointed snapshots. Decades of research on VWM have revealed its severely limited capacity, just 3–5 simple items in young adults (Cowan, 2010), as measured in the change detection task. In this task, a memory array containing a small number of
simple objects—for example, colored squares—is presented briefly (100–500 ms), followed by a short delay (500–1000 ms). After the delay, a test array is presented in which all the colors are the same (on no-change trials) or one has changed to a new color (on change trials). Capacity is estimated based on how performance declines as the number of items increases (Pashler, 1988). Despite how well established capacity limits are empirically, there remains an active debate over the source of such limits in adults (Fukuda, Awh, & Vogel, 2010).

Multiple studies have demonstrated that VWM capacity increases early in development, but the mechanism(s) underlying this developmental improvement is also a source of debate (Simmering, submitted for publication; Simmering & Perone, submitted for publication). Simmering (2008) addressed the source of capacity limits and their developmental improvement by adapting Dynamic Field Theory (DFT; Spencer, Simmering, Schutte, & Schöner, 2007), a model of spatial cognition and development, to capture change detection performance in a neurally-grounded computational model. The core architecture of the model consists of three layers of neurons tuned along a continuous color dimension. Items are represented as localized “peaks” of activation in excitatory layers, which are supported through the local excitatory and lateral inhibitory connections within and between layers. Simmering demonstrated that the model could capture change detection performance and capacity limits from early childhood into adulthood through an established neuro-developmental mechanism, the Spatial Precision Hypothesis (SPH; first proposed by Schutte, Spencer, & Schöner, 2003).

A primary advantage of using computational models to explain behavior is the ability to generate and test novel predictions. Our goal here is to test Simmering’s (2008) account for developmental changes in VWM through such predictions. In the sections that follow, we first describe how Simmering’s model captures change detection performance through early childhood. Next, we discuss the implications of the DFT and the SPH and use the model to generate novel predictions for a color discrimination task. Then we present results from a new task we developed to test how memory representations change from early childhood to adulthood. We conclude by considering further questions raised by our results, how the DFT may address these in the future, and the implications for our understanding of developmental processes in general.

1. Modeling change detection performance over development

The three-layer architecture of the DFT was developed to account for performance across a number of spatial memory tasks (Spencer et al., 2007), and was recently extended to capture some characteristics of change detection performance (Johnson, Spencer, & Schöner, 2009; see Schöner & Spencer, in press, for details on development of this architecture and additional applications). Simmering (2008) built on these previous instantiations to test whether the DFT could provide a source of capacity limits in VWM (see also Johnson, Simmering, Buss, & Spencer, in preparation). The three layers of the model contribute different cognitive functions to the task. The first excitatory layer, the Perceptual Field (PF), serves as an encoding field; inputs are presented to the model as Gaussian distributions of activation centered at the relevant color values (e.g., red, blue, and green, along a continuous color dimension). When these inputs are “on” (i.e., projecting activation into PF) localized peaks of activation form in PF; when the inputs turn “off” (i.e., the visual items disappear from the display and no activation is projected), the neurons in PF quickly return to their resting level. In this way, neurons in PF are tuned to respond only when visual stimuli are present in the array.

The second excitatory layer of the DFT, the Working Memory (WM) field, also receives weak input from the environment and strong input from PF. Thus, when visual stimuli are presented in the array, the peaks in PF and direct input to WM combine to form localized peaks in WM. Once peaks are established in WM, the items from the array have been encoded into memory. Although both PF and WM are excitatory layers, the excitatory connections within WM are tuned to be stronger than in PF. This allows WM to serve a maintenance function; when inputs are removed, the peaks in WM enter a self-sustaining state and are maintained in the absence of input, unlike activation in PF. Critically, this

---

1 This 360° color dimension is an approximation of the CIELAB 1976 perceptually-uniform color space (following Johnson, Spencer, Luck, & Schöner, 2009; Johnson, Spencer, & Schöner, 2009).
maintenance depends not only on self-excitation within WM but also on projections from the third layer in the model, the Inhibitory field (Inhib). Inhib projects inhibition back into both PF and WM. The balance between excitation within PF and WM and the inhibition coming from Inhib allow the excitatory layers to serve different cognitive functions (i.e., encoding versus maintenance).

As items are maintained in WM during the delay, the shared connectivity with Inhib produces inhibitory troughs in PF at the remembered values. These troughs support a comparison function between the items held in WM and new items presented to PF. In the change detection task, this comparison serves as the basis of the required same or different response, which is implemented through the addition of a simple responses system coupled to the three-layer architecture. This system consists of two nodes, one indicating a same response receives projections from WM, and one indicating a different response receives projections from PF. When the test array is presented, this signals to the model that a response is required; we implement this by gating activation to project from PF and WM to the response system and raising the resting level of the nodes. Activation to the response nodes builds in parallel, each with its own self-excitation and competing through mutual inhibition such that only one can “win” and generate the response for that trial. This response system allows the three-layer architecture to perform the change detection task.

Fig. 1 illustrates the model’s performance with time slices through the three layers at critical points during the trial. Fig. 1A shows the end of the memory array, once the items have been encoded; note that the three peaks in WM area are inhibiting peaks corresponding to the same colors in PF (via Inhib). Throughout the delay (Fig. 1B), the peaks in WM sustain, leading to continued inhibition of the corresponding areas in PF. Functionally, this allows the model to recognize an item as familiar (falling within a trough) or novel (falling outside of a trough) at test. When the test array is presented in the task, if the items are all the same, these inputs are projected into the troughs in PF and activation does not pierce threshold (Fig. 1C). If one item has changed, however, this input falls into a relatively uninhibited region of PF and forms a peak (circle in Fig. 1D), detecting the new item. The result of this comparison process then drives a decision. In Fig. 1C, the peaks in WM activate the same node (Fig. 1E), while in Fig. 1D, the peak in PF activates the different node (Fig. 1F). The projections to these nodes are tuned such that a single peak in PF (any new item) will generate a different decision, regardless of the number of items being maintained in WM.

Simmering (2008) simulated adults’ change detection performance with the DFT, showing that capacity in the model was limited not only by the number of peaks that can be maintained simultaneously in WM, but also the comparison and decision processes. The number of peaks is limited by the neural interaction function, that is, the balance between local excitation and lateral inhibition in WM. As more items are encoded in WM, the overall amount of both excitation and inhibition increases. At some point, a critical threshold is reached where the excitation associated with an additional item is not sufficient to overcome the additional inhibition produced, and another peak cannot sustain through the delay. Even during trials in which all items are maintained as peaks in WM, the model does not perform perfectly; the processes of comparison and decision generation in the model produce systematic errors (Simmering, 2008; see Johnson et al., in preparation, for further discussion).

Simmering (2008) next turned to the question of developmental change in capacity. Previous work in spatial cognitive development established the SPH to account for a range of behavioral phenomena (Simmering & Schutte, in press). According to this hypothesis, neural interactions in the model become stronger over development, purportedly through Hebbian learning and/or synaptogenesis. The emergent consequence of strengthening local excitation and lateral inhibition is that neural representations (i.e., peaks) increase in strength, precision, and stability over development. Although these changes in connectivity influence performance in real time (i.e., on a trial-by-trial basis), the full consequences emerge over repeated trials. Specifically, the precision of peaks refers to how closely they are centered on the presented value, which varies across trials: A peak may be accurately located on one trial for the “child” model, but not across repetitions, leading to lower overall precision than for the “adult” model. This highlights the importance of stability at different timescales; not only is an individual peak more susceptible to perturbation on an individual trial, but the peak position and strength also vary considerably across trials early in development. Simmering (2008) tested whether these consequences of the SPH could also account for developmental increases in capacity.
Quantitative simulations showed that the DFT could produce the same pattern of performance as 3–5-year-old children and adults through changing a subset of the model’s parameters according to the SPH (Simmering, 2008, submitted for publication). The child parameters were tuned to have weaker local excitation (within PF and WM) and lateral inhibition (from Inhib to PF and WM), simulating less-established connectivity within and between brain areas. Additionally, direct inputs to PF and WM were weaker and broader to approximate less precise projection from early visual areas. Noise within each layer was also stronger in early development, as would result from poorer myelination. These changes within the fields produce peaks that are weaker, broader, and less stable early in development (see Simmering, 2008, submitted for publication, for complete details). Direct neural evidence for these types of developmental change is incomplete; only recently has technology allowed the type of examination to ask these questions. General evidence for the developmental trajectory of changes in prefrontal cortex (e.g., synaptic density, Huttenlocher, 1979; myelination, Sampaio &
Truwit, 2001) is qualitatively consistent with the SPH. More recently, an architecture similar to the DFT was shown to predict developmental changes in BOLD signals based on increasing excitatory and inhibitory connections (Edin, Macoveanu, Olesen, Tegnér, & Klingberg, 2007).

To quantitatively fit the behavioral data from children in the change detection task, Simmering (2008) also modified the parameters in the decision system to reflect the types of errors children made. Similar to changes made in the three-layer architecture, the self-excitation and mutual inhibition of the decision nodes were weaker for the child parameters. In addition, the projection from WM to the same node was weaker for the child parameters, reflecting children’s tendency to respond “different” more often in the change detection task. Essentially, this made the child model more sensitive to novelty or more input-driven (relative to the adult model). The consequences of these changes are illustrated in Fig. 2, which shows change detection performance with parameters modified to capture 3-year-olds’ performance. These trials demonstrate how the SPH leads to increased errors in the model.

The panels in Fig. 2 follow the same progression as in Fig. 1, with an additional delay panel (Fig. 2C) showing how maintenance of the WM peaks differed across trials. As Fig. 2A shows, the first notable difference in performance with the child parameters is the nature of peaks in WM: with the weaker interactions, the child model encodes items much more tenuously. Peaks are weak and unstable. Indeed, by the end of the delay on the no-change trial (Fig. 2B), activation of one peak in WM has dropped below threshold, leading to the loss of the corresponding inhibition (see circled region of Inhib and WM). By comparison, this item is held in memory during the delay on the change trial (Fig. 2C). Note that the difference between these two trials is due to fluctuations in noise, as the input has not yet differentiated the trial type. Thus, the loss of a peak is equally likely to occur on change versus no-change trials.

The weak peaks in WM formed shallow, broad troughs in PF at the corresponding values on both trials (Fig. 2B–E). When a no-change test array is presented (Fig. 2D), the inputs reinforce the two peaks that remain in WM, but the “forgotten” item is identified as new (circle in PF), producing an inaccurate different decision (Fig. 2F). This is a common source of “false alarm” errors in the model (Johnson et al., in preparation). By contrast, on the change trial, the model has maintained all three items in WM (Fig. 2C). At comparison (Fig. 2E), one item has changed, resulting in no input at −90° and a new input at 120°. Although this new input causes activation in PF to increase, weaker interactions with these parameters are not strong enough to push activation above threshold (circle in PF). Thus, no activation is projected to the change node, and the model produces an inaccurate same decision (Fig. 2G).

The sample trials in Fig. 2 illustrate how the child parameters in the DFT produce errors; correct responses arise as with the adult parameters (Fig. 1). Overall, the model makes more errors with the child parameters than with the adult parameters, capturing children’s poorer performance and, by extension, lower capacity estimates (Simmering, 2008). Combined with simulations of adults’ performance, Simmering’s model provides the first computational account of the source of VWM capacity limits in adults as well as a mechanism underlying developmental increases in capacity. A critical part of theory and model development, however, is generating and testing new predictions, which we do here.

2. Novel predictions derived from the model

As Fig. 2 shows scaling parameters according to the SPH lead to differences in how the model is able to perform the encoding, maintenance, comparison, and decision processes involved in the change detection task. Here we explored these characteristics of children’s performance further by developing a new task to measure the nature of children’s individual memory representations more directly. We designed a single-item change detection task that tests very small changes in color to assess the stability and precision of children’s memory. Specifically, we predicted that young children’s performance will reveal less precision and stability for colors early in development.

This prediction arises from the processes illustrated in Fig. 2. In particular, weaker “child” interactions have two consequences for detecting small changes of a single item. First, the weaker encoding and maintenance of the item means that peaks will be less precise, that is, more likely to be centered on a different value than the actual target on each trial (for similar developmental changes in spatial
Fig. 2. Time slices through the three layers of the model at critical points in two trials, using parameters tuned to capture 3-year-olds’ performance: (A) encoding; maintenance during (B) no-change and (C) change trials; comparison for (D) no-change test array and (E) change test array. Also shown is activation of the decision nodes following the (F) no-change and (G) change test arrays. Note that the model made errors on both trials. Arrows indicate progression through the trial(s). Dashed lines in each panel indicated the activation threshold (i.e., 0).
memory, see Simmering & Spencer, 2008). A second, related consequence is lower stability in the color value across trials. Across repeated trials with the same target value, the child parameters will show more variability in the location of the peak than the adult model. Thus, measuring the model’s performance in detecting changes in that color value should show less accuracy and more variability across repeated trials.

We tested this prediction by using the parameters Simmering (2008) developed to capture performance by 3-year-olds, 5-year-olds, and adults, testing the model in a single-item change detection task with small separations between items.\(^2\) Specifically, we presented a target color at 0° followed by a second stimulus that either matched exactly (0°) or changed incrementally across different values for the test stimulus.\(^3\) We simulated 40 trials at each separation for each parameter setting (i.e., “age”) and tabulated the percent of trials at each separation on which the model responded same versus different.

Fig. 3 shows our simulation results, demonstrating the developmental effects of changes in the precision and stability of peaks in the model. With the adult parameters (solid line), performance begins with 100% same responses for the matching items and small stimulus separations. As the items become less similar (i.e., increasing steps), the number of same responses drops dramatically. For the child parameters (dashed and dotted dashed lines), however, this is not the case. First, there are fewer same responses when the items are identical or very similar, and the number of same responses drops off more gradually. For the adult model, the transition from mostly same to mostly different responses happens within only a few steps. For the child model, however, this transition takes much longer, with the 3-year-old parameters only settling near 0% same responses 6 steps further than the adult parameters.

To quantify performance of the model over development, we fit a probit function to each line in Fig. 3 and estimated the distance (in steps) between items necessary for the model to produce different responses 75% of the time (following Simmering, Spencer, & Schöner, 2006). These estimates were approximately 7.6 steps for the adult parameters, 8.9 steps for the 5-year-old parameters, and 10.3 for the 3-year-old parameters. Moreover, we looked at variability in the estimates across repeated runs of the model as an indication of changes in stability. For the 3-year-old parameters, the standard deviation across 10 runs of the model was 0.76 steps, whereas it was 0.37 steps for the 5-year-old parameters, and 0.23 for the adult parameters.

---

\(^2\) We made one change to simulate our single-item task relative to Simmering’s (2008) parameters. Pilot data with young children suggested more of a bias to respond “same” in our task, leading us to decrease the amount of change in the projection to the same node over development. In the original parameters, this projection was scaled to 54% for 3-year-olds and 80% for 5-year-olds (relative to adults); here we decreased the developmental changes by half, resulting in a 77% scaling for 3-year-olds and 90% for 5-year-olds.

\(^3\) In these simulations, we tested increments of 4° but refer to them as “steps” because the color space used to generate the stimuli in Simmering (2008) was different than the colors we test here. This difference could possibly result in a differential scaling of nodes in the model to degrees in color space, precluding precise quantitative predictions here.
These estimates suggest that performance should improve through early childhood and into adulthood, both in terms of the ability to discriminate colors in memory, and the reliability of participants’ discrimination across repetitions of the same stimulus pairs. Here we test the prediction that performance in a single-item change detection task will improve as memory representations become more precise (requiring smaller differences between colors to respond “different”) and more stable (with more consistent responding over repetitions of the same stimulus separations) over development.

An important question to consider is whether this model, or any model, is necessary to generate this prediction. After all, predicting developmental improvements in memory tasks is certainly not unique to the DFT and SPH. Our model goes one step further than many theories in predicting generalization from one type of measure (capacity) to a second (precision). Moreover, implementation of this theory in a formal model allows for exploring hypotheses within the model before testing them experimentally (as illustrated later). Models can provide a level of specificity beyond many verbal theories because, in order to test a model, assumptions regarding how various processes contribute to behavior must be made explicit. This is not to say that all models are necessarily more specific and complete than all verbal theories (Simmering, Triesch, Deák, & Spencer, 2010); however, formal models often force theorists to a deeper level of explanation than their own verbal theories may incorporate.

To test our predictions of improved precision and stability in VWM over development, we developed a color memory task from Simmering and Spencer’s (2008) position discrimination task. They employed a stair-casing procedure in which stimuli were presented with a large separation; when children correctly responded that the positions were different, the next trial would test a smaller separation, continuing in this fashion until the child responded “same”. This method was repeated across four runs to each of two targets to estimate children’s position discrimination thresholds. We adapted this task to measure children’s memory for colors by using a similar stair-casing procedure to test small changes in color (much smaller than the changes probed in typical change detection tasks), with increasing the difference between colors over trials. This task design can assess the stability of memory representations through the standard deviation across repeated within-subject trials.

3. Experiment 1

In this experiment, we used the method of ascending limits to test participants’ discrimination thresholds: on each trial, participants viewed two colored mittens presented sequentially; the “distance” between the mittens in color space (approximated in degrees) increased across trials. We predicted that the difference between colors necessary to elicit reliable “different” responses would decrease over development, reflecting an increase in precision of color memory. Additionally, we tested participants on multiple runs to each of two targets, to assess variability in their responses across trials. We predicted that cross-run variability would decrease over development as memory representations become more stable.

3.1. Method

3.1.1. Participants

Participants included 42 4-year-olds ($M = 4.55$, $SD = 0.21$, 26 female), 33 5-year-olds ($M = 5.40$, $SD = 0.20$, 13 female), 33 6-year-olds ($M = 6.25, SD = 0.21$, 22 female), and 23 adults ($M = 21.16, SD = 1.55$, 18 female). All had normal or corrected-to-normal visual acuity and reported no colorblindness. An additional 9 children participated but were excluded from analyses due to experimenter error (two 6-year-olds), because they chose to end early (six 4-year-olds), or because they did not understand

---

4 The sample from Simmering (2008) included 3-, 4-, and 5-year-olds; however, pilot testing showed our task to be too challenging for 3-year-olds to perform reliably, leading us to sample an older age range here. For the purposes of testing the model, development is continuous throughout early childhood, leading to the same predicted pattern between 4 and 6 years as between 3 and 5 years.
the task (one 6-year-old). We also excluded an additional 16 children (eight 4-year-olds, four 5-year-olds, and four 6-year-olds) because their responses suggested our stair-casing procedure could not accurately measure their memory for the items (see below for further discussion). Five additional 4-year-olds were excluded from our analyses as outliers; we identified outliers by examining histograms of each age group’s performance (after removing the children listed above). These histograms showed thresholds greater than 56° (i.e., 7 steps) to be outside the normal distribution, leading us to exclude children with one or both thresholds beyond this criterion. Because these exclusions work against our hypotheses (as we excluded young children with high thresholds), we regarded this to be a conservative approach.

Children were recruited from the community surrounding the university and tested in the laboratory (27 4-year-olds, 19 5-year-olds, all 6-year-olds) or through area preschools and tested in a designated area within the preschool (15 4-year-olds, 14 5-year-olds). Adults were university students and/or staff who were recruited through undergraduate courses or word of mouth and were tested in a laboratory on campus.

3.1.2. Procedure

Children completed the task on a 15.4 in. widescreen Dell Latitude E6500 laptop computer; adult participants were tested on either this laptop (n = 8) or a 21 in. LCD monitor on a Dell Optiplex 760 desktop computer (n = 15; note that stimuli were scaled to appear the same size on both displays). Participants sat approximately 2′ from the display during the task. Stimulus presentation was controlled by Matlab using the Psychophysics Toolbox extension (version 3, Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). On each trial, the computer displayed a gray background (RGB = 200, 200, 200) for 500 ms, followed by a left mitten for 1 s, a blank gray screen for a 500 ms delay, a right mitten for 1 s, and finally a gray screen with the prompt “Enter your response (same/different)” that remained visible until a response was entered. At the prompt, participants identified whether the mitten colors were “same” or “different”; children responded verbally, often using terms “match” and “no match”; adults entered their responses on a numeric keypad. Mittens appeared as 2.5 in. wide × 3 in. tall rectangles with a white background and black outline; the interior colors of the mitten stimuli were determined according to an ascending stair-casing procedure (described below). Mitten stimuli were centered vertically on the screen and presented on the left half of the screen for trials to Target 1 and the right half of the screen for trials to Target 2. One each trial, the first mitten appeared in the left position and the second appeared on the right position (within the respective half of the monitor).

Trials were presented in eight runs alternating between two target colors. The two colors that served as targets were 70° (teal, RGB = 49, 138, 138) and 290° (pink, RGB = 233, 72, 140); these colors were selected to be far from category boundaries based on a separate group of adults’ (n = 12) categorization of the 360° color space used by Johnson and colleagues (2009; see online supplementary materials for images of colors and corresponding RGB values). We developed a stair-casing procedure (following Abrimov et al., 1984; Simmering & Spencer, 2008) to reduce the number of trials required by children while providing repeated measures (across runs) of responses to the same stimuli.

Fig. 4 illustrates the stair-casing procedure. Which color served as Target 1 versus 2 was counterbalanced across participants. Each trial began with the presentation of the Target color as the left mitten. Trial 1 of all runs presented the target color as the right mitten as well (i.e., matching colors; Fig. 4). Here we step through a sample progression of the procedure across runs alternating between the two targets. If the participant responded “same” on Trial 1 of Run 1, Trial 2 presented a one-step different color (i.e., 78° in Fig. 4) for the right mitten; we chose a step size of 8° from pilot data. Trials continued in this manner, increasing the separation between the colors by 8° steps on each trial, until the participant responded “different” to terminate the run.

5 We recognize that testing in different settings and on different equipment can be problematic for probing color perception (due to differences in ambient light, background luminance, etc.). However, the goal of our task is not to establish precise measurements of perceptual discrimination; rather, we are interested in general patterns of recognition over development.
Fig. 4. Example sequence of trials using the ascending stair-casing procedure from Experiment 1 (S = “same”/D = “different” response). Note that degree labels were not present during the task, and the order of target presentation was counter-balanced across participants.

Following the termination of Run 1 to Target 1, Run 1 for Target 2 proceeded in the same manner, beginning with matching mittens and increasing by 8° steps with each “same” response (Fig. 4). After Run 1 for Target 2, the next run returned to Target 1 (Run 2 to this target), again beginning with identical colors. For Trial 2 in this run, the right mitten differed from the left by one step less than the color identified as “different” in the previous run (i.e., in Fig. 4, Trial 2 of Run 2 for Target 1 presented 86°, or 94–8°). Trials and runs continued in this manner, alternating between Targets, until four runs to each target were completed.

As stated above, inspection of the data revealed that this stair-casing procedure was not appropriate to measure thresholds for a subset of the children (the 16 described as excluded). In particular, these were cases in which the child responded “same” for many trials in Run 1, producing a large separation on the final (i.e., “different”) trial. These children then responded “different” on Trial 2 of all subsequent runs to that target, suggesting their true thresholds may be lower than the separations tested in our stair-casing procedure. For example, if Run 1 terminated with a separation of 64°, then Trial 2 of Run 2 (to that target) would test a 56° separation. When the child responded “different”, Run 2 would terminate and Trial 2 of Run 3 would test a 48° separation. When the child responded “different” again, Trial 2 of Run 4 tested a 40° separation. If the child again responded “different” the experiment would terminate without testing smaller separations for that target. It is possible the child in this example situation would have responded “different” at smaller separations, had they been tested on additional runs to that target. We felt that estimates derived from such a pattern could be artificially inflating children’s thresholds, possibly providing false support for our hypothesis; therefore, we chose to exclude them from analysis. We return to this issue in our design of Experiment 2.
Performance was analyzed by taking the separation at which the participant responded “different” to end each run and computing the mean and standard deviation across runs as the participant’s threshold (i.e., difference in color space) and variability (i.e., differences in responses across runs to the same target) scores, respectively. For example, in the trials shown in Fig. 4, the separations were 24°, 32°, 24°, and 24° for the 70° target, and 16°, 24°, 24°, and 24° for the 290° target; this would produce thresholds of 26° and 22° (means across runs), respectively, and variability scores (standard deviation across runs) of 4° in both cases. Preliminary analyses of gender and testing conditions (preschool versus lab for 4- and 5-year-olds; laptop versus desktop for adults) showed no significant effects. We therefore excluded these factors from further analysis.

Mean thresholds are shown separately for each target across age groups in Fig. 5A. Thresholds were generally higher (i.e., worse performance) for children, decreasing over development, and were higher for the 70° versus 290° target. We analyzed mean thresholds in a two-way analysis of variance (ANOVA) with Target (70°, 290°) as a within-subjects factor and Age (4, 5, 6, adults) as a between-subjects factor. This analysis revealed significant main effects of Target, \( F_{1,127} = 21.11, p < .001 \), and Age, \( F_{3,127} = 21.24, \)

---

6 Alternative approaches would be to use the highest separation at which participants responded “same”, or averaging between the “same” and “different” separations on each run. Because we used a fixed step size of 8°, we cannot say where in that 8° separation between colors each participant’s threshold falls. However, using the same step size across participants allows more straightforward comparison across individuals and age groups by preventing differences in step sizes from being conflated with performance. In our analyses we use the smallest separation that generated a “different” response as a proxy for true thresholds and do not interpret these values as absolute for each participant or age group.
p < .001, as well as a significant Target × Age interaction, $F_{3,127} = 3.46, p < .05$. The Target main effect was driven by higher thresholds to the 70° target ($M = 22.82°$) than the 290° target ($M = 19.40°$). Follow-up Tukey HSD tests ($p < .05$) on the Age main effect showed that 4-year-olds ($M = 24.41°$) did not differ from 5-year-olds ($M = 24.08°$), but all other differences were significant (6-year-olds $M = 19.59°$, adults $M = 13.00°$).

We explored the source of the Target × Age interaction by conducting separate one-way ANOVAs with Age as a factor for each Target. Both analyses revealed significant improvements over development: for 70°, $F_{3,127} = 17.96, p < .001$; for 290°, $F_{3,127} = 10.91, p < .001$. Follow-up Tukey HSD tests ($p < .05$) for the 70° target showed that all age groups differed significantly except 4- and 5-year-olds, mirroring the overall Age main effect. For the 290° target, adults had significantly lower thresholds than all age groups of children, who did not differ from one another. We also conducted separate one-way ANOVAs with Target as a factor for each age group. Performance only differed across targets for the younger children: 4-year-olds, $F_{1,41} = 13.54, p < .001$; 5-year-olds, $F_{1,32} = 8.92, p < .01$. In both cases, thresholds were lower for the 290° target than for the 70° target.

**Fig. 5B** presents cross-run variability scores, separated by target and age group. Children showed higher variability, especially for the 70° target, than adults, who had higher variability for the 290° target. We analyzed these scores in a two-way ANOVA with Target (70°, 290°) as a within-subjects factor and Age (4, 5, 6, adults) as a between-subjects factor. This analysis revealed significant main effects of Target, $F_{1,127} = 11.69, p < .001$, and Age, $F_{3,127} = 4.31, p < .01$. The Target effect was driven by lower cross-run variability to the 290° target ($M = 4.51°$) than to 70° target ($M = 5.87°$). Follow-up Tukey HSD tests ($p < .05$) on the Age main effect showed that 4-year-olds ($M = 6.21°$) had significantly higher variability across runs than adults ($M = 4.01°$); no other differences between groups were significant (5-year-olds $M = 5.13°$; 6-year-olds $M = 4.79°$).

### 3.3. Discussion

This study tested predictions derived from Simmering's (2008) proposal that developmental increases in capacity result from strengthening neural interactions. These changes in connectivity improve the stability, strength, and precision of VWM representations over development in the DFT. We developed a single-item change detection task to test for predicted improvements in both the precision and stability of VWM representations in 4-, 5-, and 6-year-olds and adults. Results generally supported these predictions. First, mean thresholds decreased (i.e., precision improved) between 4 and 6 years and again into adulthood (although the improvement from 4 and 6 years was significant for only one color). Second, variability across runs decreased (i.e., stability improved) from childhood to adulthood.

One unexpected result was the difference in performance and developmental change across targets. The model made no a priori predictions of different performance across target colors. It is unclear why thresholds did not decrease significantly between 4 and 6 years for the 290° target, or why 4- and 5-year-olds showed significantly better performance to this target than to 70°. Anecdotal reports of children's performance suggested that 290° may have been a more familiar color; experimenters reported that children were more likely to name “pink”, and that only a few children produced the term “blue” or “teal” (how most adults labeled the 70° target). Although this is one possible difference across colors, it does not provide a clear explanation of the difference in performance. Other studies have shown that labeling colors decreases adults' ability to discriminate among colors within the category (Winawer et al., 2007), therefore suggesting children would perform worse on colors that they labeled.

Another possibility is that the boundary of the “pink” color category was slightly above 290° for most or all children (perhaps near 310°, as this is where the average thresholds were), leading to reliable discrimination at the category boundary. For the 70° target, however, categories may have varied more across participants, or the boundary may have been much further from the target color. In this case, children's production of the "pink" name could be an indication of this as a familiar and stable category representation, whereas they did not have a strong category for the teal color. We tried to reduce the effects of category boundaries by emphasizing the point that small differences in color should be identified as “different”, even though they would be labeled with the same name. We
illustrated this by demonstrating the task to children using two shades of brown infant socks, followed by flashcards showing different shades of blue mittens.

A final concern is the appropriateness of the stair-casing procedure for young children. As noted, 5 children were excluded as outliers due to high thresholds, and another 16 were excluded because the design may not have accurately measured their memory. The majority of the children who were excluded were 4-year-olds, suggesting that the stair-casing procedure was least suited to their performance. Thus, to ensure that the developmental difference we found for the 70° target was not an artifact of our testing procedure or specific only to that color, we designed a second experiment with a new stair-casing procedure and tested two additional target colors.

4. Experiment 2

In this experiment we hoped to gain a more accurate picture of developmental changes in color memory by modifying our stair-casing procedure to be more appropriate for young children (Fig. 6). We began as in Experiment 1, with ascending runs in which target colors become more dissimilar with repeated “same” responses. However, when returning to the target for a second run, we presented a descending run, beginning with a larger separation than had been identified as “different” on the previous run to that target and increasing similarity with each “different” response. Our hope was that this method would be more robust to a child’s occasional long runs, as it allows the following run to descend as far as needed to reach a “same” response. We also tested two different target colors in this version to determine whether performance to the 70° or 290° target in Experiment 1 was more indicative of developmental change in memory precision.

4.1. Method

4.1.1. Participants

Participants were 25 4-year-olds ($M=4.49$, $SD=0.26$, 15 females), 24 5-year-olds ($M=5.54$, $SD=0.31$, 14 females), 25 6-year-olds ($M=6.46$, $SD=0.25$, 15 females), and 24 adults ($M=20.35$, $SD=0.31$, 14 females).
SD = 3.27, 15 females). All had normal or corrected-to-normal visual acuity and reported no color-blindness. An additional 32 participants were excluded from analyses due to experimenter error (one 4-year-old, two 5-year-olds, four adults), choosing to end early (eight 4-year-olds, one 5-year-old), because they did not understand the task (10 4-year-olds, five 5-year-olds) or due to a vision abnormality (one 4-year-old). We identified 17 additional children (seven 4-year-olds, seven 5-year-olds, and three 6-year-olds) as outliers by plotting performance histograms of each age group. As in Experiment 1, these histograms showed mean thresholds greater than 56° (i.e., 7 steps), outside the normal distributions, leading us to exclude children with one or both thresholds beyond this criterion. Children and adults were drawn from the same population as in Experiment 1.

4.1.2. Procedure

The apparatus was identical to that used in Experiment 1. Children completed the task on the same laptop computer and adults on the same desktop computer as in Experiment 1.

The procedure was similar to Experiment 1 with four modifications. First, participants completed a total of 12 runs rather than 8. Second, runs alternated between two new target colors, 136° (green-brown, RGB = 136, 132, 60) and 356° (purple-blue, RGB = 131, 115, 189). We chose these colors to be far from adults’ category boundaries (see Experiment 1), as well as being named inconsistently by adults (i.e., equal labeling of 136° as “green” versus “brown” and 356° as “blue” versus “purple”). Third, we added a “repeat” response option, allowing the experimenter to repeat an identical trial when the participant indicated s/he was not sure or was not attending to the screen while one or both mittens were presented. Fourth, the stair-casing procedure was modified to begin with ascending runs to each target, then switch to descending on the following run to this target. Fig. 6 illustrates this procedure; as in Experiment 1, which color was presented as Target 1 versus 2 was counter-balanced across participants.

The initial ascending run to each target followed the procedure from Experiment 1, terminating when the participant responded “different”. Following Run 1 for Target 2, the next run returned to Target 1 and began with one step larger than the separation that ended the previous run to that target; in Fig. 6, for example, Run 1 terminated at 160° for Target 1, so Run 2 to Target 1 began with 168°. If the participant responded “different” (as most did), the next trial presented a separation one step smaller than the previous trial, and continued in this manner until the participant responded “same”. If the participant responded “same” on the first trial of Run 2, this run continued in an ascending manner. The goal was to provide a more sensitive estimate of thresholds and allow recovery from unintended responses on previous runs.

At the termination of a descending run, the subsequent run to that target began with a separation one step smaller than the separation that terminated the previous run; in Fig. 6, for example, Run 2 to Target 1 terminated at 160°, leading Run 3 to Target 1 to begin with 152°. If a run terminated with identical stimuli (see Run 2 to Target 2 in Fig. 6), the next ascending run began with this pair rather than stepping in the opposite direction from the Target (see Run 3 to Target 2). Runs continued in this manner, alternating between targets, until six runs to each target were completed (for simplicity Fig. 6 shows only four runs to each target).

4.2. Results

Following Experiment 1, we analyzed performance by taking the smallest separation at which the participant responded “different” in each run and computing the mean and standard deviation across runs as the participant’s threshold and variability scores, respectively, separately for ascending and descending runs. In Fig. 6, for example, Target 1 includes 160° and 176° for ascending runs and 168° and 176° for descending runs, leading to averages of 168° and 172°, respectively. These data yield mean thresholds of 32° for ascending and 36° for descending, with cross-run variance scores (i.e., standard deviation across runs) of 11.31 and 5.66, respectively. Preliminary analyses revealed no significant effect of gender leading us to exclude this factor from further analysis.

Fig. 7A and B shows mean thresholds separately for each target across age groups, with results for ascending versus descending runs in different panels. Thresholds were higher for ascending runs (Fig. 7A) than descending runs (Fig. 7B), suggesting our modification to include both directions helps
avoid overestimating thresholds. The overall pattern across targets and development, however, is the same across directions: performance improved with development, with 4-year-olds showing a notable difference across targets and more comparable performance across targets in other age groups.

We analyzed mean thresholds in a three-way ANOVA with Target (136°, 356°) and Direction (ascending, descending) as within-subjects factors and Age (4, 5, 6, adults) as a between-subjects factor. This analysis revealed significant main effects of Direction, $F_{1.96} = 37.17, p < .001$, and Age, $F_{3.96} = 13.67, p < .001$, as well as significant Direction × Age, $F_{3.96} = 5.84, p < .01$, Target × Age, $F_{3.96} = 6.61, p < .001$, and Direction × Target, $F_{1.96} = 3.98, p < .05$, interactions. The Direction main effect was driven by overall higher thresholds for ascending ($M = 20.27°$) than descending runs ($M = 15.72°$). Follow-up Tukey HSD tests ($p < .05$) for the Age main effect showed that all groups of children had higher thresholds than adults (4-year-olds $M = 23.38°$; 5-year-olds $M = 18.83°$; 6-year-olds $M = 18.28°$; adults $M = 11.77°$); thresholds were also significantly higher for 4-year-olds than 6-year-olds, supporting our prediction of increased precision with age during early childhood.

We explored the source of the Direction × Age interaction by conducting separate one-way ANOVAs with Age as a factor for each Direction. Both analyses revealed main effects of age: ascending, $F_{3.94} = 16.22, p < .001$; descending, $F_{3.94} = 5.38, p < .01$. Tukey HSD follow-up tests ($p < .05$) showed that, for ascending runs, children did not differ (4-year-olds $M = 25.90°$, 5-year-olds $M = 21.64°$, 6-year-olds $M = 21.95°$) but all had significantly higher thresholds than adults ($M = 11.26°$). For descending runs, only 4-year-olds ($M = 19.55°$) had significantly higher thresholds than adults ($M = 12.31°$); 5-year-olds ($M = 15.57°$) and 6-year-olds ($M = 15.31°$) did not differ from other age groups. We also conducted separate one-way ANOVAs with Direction as a factor for each age group. These revealed significant main
effects of direction for each of the age groups of children, but not adults: 4-year-olds, $F_{1,24} = 10.83$, $p < .01$, 5-year-olds, $F_{1,23} = 20.08$, $p < .001$, 6-year-olds, $F_{1,24} = 17.48$, $p < .001$. In all cases, children had higher mean thresholds on ascending than descending runs (see means above).

We next analyzed the Target $\times$ Age interaction by conducting separate one-way ANOVAs with Age as a factor for each Target. Both analyses revealed significant improvements over development: $136^\circ$, $F_{3,94} = 5.24$, $p < .01$; $356^\circ$, $F_{3,94} = 13.90$, $p < .001$. Follow-up Tukey HSD tests ($p < .05$) for the $136^\circ$ targeted showed that 4-year-olds ($M = 19.98^\circ$) and 5-year-olds ($M = 19.00^\circ$) had higher thresholds than adults ($M = 12.94^\circ$); 6-year-olds ($M = 17.17^\circ$) did not differ from any other age group. For the $356^\circ$ target, all groups differed significantly except for 5- and 6-year-olds (4-year-olds $M = 26.79^\circ$, 5-year-olds $M = 18.67^\circ$, 6-year-olds $M = 19.37^\circ$, adults $M = 10.60^\circ$). We also conducted separate one-way ANOVAs with Target as a factor for each age group. These revealed significant main effects of target for 4-year-olds, $F_{1,24} = 6.89$, $p < .05$, and adults, $F_{1,23} = 7.45$, $p < .05$. For 4-year-olds, thresholds were higher to the $356^\circ$ target than the $136^\circ$ target; for adults, the pattern was opposite (see means above).

Lastly, we examined the Direction $\times$ Target interaction by conducting separate one-way ANOVAs with Direction as a factor for each Target. Both analyses revealed significant differences across Directions: for $136^\circ$, $F_{1,97} = 22.58$, $p < .001$; for $356^\circ$, $F_{1,97} = 23.65$, $p < .001$. In both cases, thresholds were higher on ascending ($136^\circ$ $M = 19.14$, $356^\circ$ $M = 21.39$) than descending runs ($136^\circ$ $M = 15.39$, $356^\circ$ $M = 16.04$). We also conducted separate one-way ANOVAs with Target as a factor for each Direction. Only ascending runs showed a significant difference across targets, $F_{1,97} = 5.22$, $p < .05$, with lower thresholds to $136^\circ$ than $356^\circ$ (see means above).

Fig. 7C and D presents cross-run variability scores, separated by target and age group, for ascending (C) and descending (D) runs. Variability across runs did not differ notably across run directions and did not change dramatically over development. Four-year-olds did show a slight difference across targets in parallel to the difference in thresholds shown in Fig. 7A and B. We analyzed these scores in a three-way ANOVA with Target ($136^\circ$, $356^\circ$) and Direction (ascending, descending) as within-subjects factors and Age (4, 5, 6, adults) as a between-subjects factor. This analysis revealed significant main effects of Direction, $F_{1,96} = 11.12$, $p < .01$, and Age, $F_{3,96} = 11.60$, $p < .001$, as well as an Age $\times$ Target interaction, $F_{3,96} = 2.72$, $p < .05$. As Fig. 7C and D shows the Direction main effect was driven by higher variability in ascending ($M = 7.72^\circ$) than descending runs ($M = 5.59^\circ$). Follow-up Tukey HSD tests ($p < .05$) on the Age main effect showed that all children (4-year-olds $M = 10.14^\circ$, 5-year-olds $M = 6.91^\circ$, and 6-year-olds $M = 7.87^\circ$), had significantly higher variability across runs than adults ($M = 2.78^\circ$); differences between groups of children were not significant.

We explored the source of the Target $\times$ Age interaction by conducting separate one-way ANOVAs with Age as a factor for each Target. Both analyses revealed significant improvements with development: $136^\circ$, $F_{3,94} = 7.30$, $p < .001$; $356^\circ$, $F_{3,94} = 8.97$, $p < .001$. Follow-up Tukey HSD tests ($p < .05$) for the $136^\circ$ target showed that all groups of children (4-year-olds $M = 8.75^\circ$, 5-year-olds $M = 7.15^\circ$, 6-year-olds $M = 7.08^\circ$) had higher variability than adults ($M = 3.27^\circ$). For the $356^\circ$ target, 4-year-olds ($M = 11.53^\circ$) had significantly higher variability than 5-year-olds ($M = 6.67^\circ$) and adults ($M = 2.29^\circ$); additionally, 6-year-olds’ ($M = 8.67^\circ$) variability was higher than adults’. We also conducted separate one-way ANOVAs with Target as a factor for each age group. These revealed no significant main effects of target ($p > .08$).

4.3. Discussion

The goal of this experiment was to overcome some limitations of our original stair-casing procedure while testing two different target colors to see which pattern from Experiment 1 would generalize. By testing children with a modified stair-casing procedure including both ascending and descending runs, we were able to better identify children’s true memories for colors. With this modified procedure testing different target colors, we replicated effects from Experiment 1: Thresholds decreased over development, but now for both targets (compared to only $70^\circ$ in Experiment 1); variability also decreased during childhood for the $356^\circ$ target and from childhood to adulthood for the $136^\circ$ target. Additionally, we found that children’s thresholds were reliably higher when estimated in ascending versus descending runs, suggesting that including both types of runs is preferable to using only ascending runs.
A significant difference across targets occurred for 4-year-olds and adults. Interestingly, these differences were in opposite directions, with better performance on the 136° target for 4-year-olds versus the 356° target for adults. This suggests different developmental trajectories for different colors throughout the color space and that the results from Experiment 1 were not necessarily anomalous. However, this result does not give a clear indication of why performance differs across targets early in development, an issue we turn to next.

5. General discussion

Across two experiments, we tested the precision and stability of children’s and adults’ memory for colors in a single-item change detection task. The developmental mechanism proposed by Simmering (2008) to capture increasing VWM capacity predicts that memory becomes more precise and stable as capacity increases. Our results supported this prediction, showing significant improvements in both thresholds and cross-run variability over development, although developmental trajectories varied by target color. The strongest support came from changes in thresholds to the 70° and 356° targets between 4 and 6 years of age, a developmental period during which VWM capacity increases from roughly 2 to 3 items.

Our behavioral results provide support for the cognitive and developmental mechanisms embodied within the DFT and SPH, but they also raise a critical challenge for the model: Why do different target colors show different developmental trajectories? The model represents colors within a homogeneous space (i.e., selectively-tuned neural fields), but this is not necessarily the case in the human visual system; indeed, models of color perception have suggested an inhomogeneous perceptual space for color. We attempted to avoid this experimentally by choosing our target colors from 180 colors equally-distributed through the CIELAB 1976 color space, which approximates perceptually-uniform color changes between different colors (Johnson, Spencer, Luck et al., 2009), and by testing different target colors within participants. Doing so revealed unexpected developmental inconsistencies across targets, in that younger children showed large differences in performance for different target colors, while older children and adults showed only small differences.

Could this difference across targets—or even our key developmental finding of improved discrimination—be driven by perceptual differences? Previous research has shown stable color perception over development for colors that varied in hue (holding brightness and saturation constant; Petzold & Sharpe, 1998). Petzold and Sharpe examined developmental change by comparing across relatively broad age ranges, comparing preschoolers (3–6 years) to preadolescents (9–11 years) and young adults (22–30 years), without assessing differences within each age group. However, there were no significant differences between preschoolers and preadolescents in their two tests of discrimination. This suggests little room for developmental change in perception between 3 and 6 years, but to our knowledge, this question has not been tested directly.

It is also possible that perception of the target colors was equally sensitive at any given point in development, but that differences in performance across targets arose through memory processes. As not, children’s familiarity with colors and/or color categories may affect their discrimination of colors in different categories. Although the familiarity explanation makes intuitive sense in Experiment 1, with pink versus teal, it does not seem as likely in Experiment 2, where both colors were less familiar. Finding the explanation for the color differences will require experiments testing these possibilities through separate assessments of participants’ perception, familiarity, and categorization of target colors. We can explore some of these questions through model simulation, as an initial evaluation of the potential explanations. This is a key advantage of process-based computational models like the DFT over less-specified theories: Potential explanations posed to account for our empirical findings can be tested in the model before conducting further experiments.

We tested these possibilities in the model in relatively simple ways, by increasing the resting levels of PF and WM. Increasing the resting level of PF corresponds to enhanced perceptual processing of the stimuli. For simplicity, we raised the resting level of the entire field and repeated the simulations using the 3-year-old and adult parameters; to implement differences across targets, however, the resting level would vary across color space. Our simulation results are shown in Fig. 8A. Increasing the resting level in PF indeed improved discrimination performance for children by 2–3 steps,
from approximately 10.3 steps in the original simulations (resting level of −7) to 8.4 steps (resting level of −6.5) and 7.6 steps (resting level of −6). However, the effect on the adult parameters was negligible, keeping threshold estimates between 7.5 and 7.6 across resting levels. Empirically, this would predict that children who show better perceptual discrimination of colors should also show better memory for those colors, but that the difference would not be evident in adults’ performance. Moreover, these simulations show a reduction in developmental change (i.e., similar performance between children and adults) as the resting level of PF increases. This type of perceptual difference across colors is a potential source for the behavioral differences we found across targets early in development.

Increasing the resting level of WM corresponds to more familiarity with the color. Implementations of long-term memory within the DFT have taken the form of reciprocally-coupled excitatory long-term memory fields, in which repeated exposure to a stimulus effectively boosts the resting level for the corresponding neurons in WM (Simmering, Schutte, & Spencer, 2008). We approximated this process here by raising the resting level of WM from −4.5 to −4 or −3.5 for both the 3-year-old and adult parameters. The results, shown in Fig. 8B, suggest the opposite effect of our intuition for children’s performance: Rather than familiarity improving discrimination with the 3-year-old parameters, the increased resting level led to a decrease in performance, requiring larger separations to reliably respond “different” (from 10.3 to 10.8 to 16.3 steps as the resting level increased). For the adult parameters, however, we again see very little change (from 7.6 to 7.8 to 8.1 steps). These simulations suggest an interesting prediction—that children, but not adults, should show worse discrimination performance on more familiar colors. This needs to be tested empirically but may relate to the category/labeling effect shown by Winawer et al. (2007).

The simulations described in this article illustrate how computational models can provide the specificity needed to generate novel, testable predictions and to test potential explanations before conducting further behavioral experiments. The type of process-based model we use is especially...
well-suited to these goals, as it performs a task on a trial-by-trial basis and incorporates multiple cognitive processes to generate the same type of behavior measured in laboratory tasks. The present empirical results provide a first step in this direction, but much work remains to provide a more complete test of the explanations embodied in the DFT and SPH. Although these results are consistent with the predictions derived from the model, they do not provide direct evidence for the link between precision and capacity. The change in these processes could be coincidental, driven by different developmental mechanisms. To further test the link between precision and capacity will require experiments comparing performance across tasks within the same groups of participants. Correlated performance across two tasks over development will provide stronger evidence for a common underlying mechanism.

Even if future studies provide evidence for this link, the SPH may not account for the full developmental trajectory of changes in VWM capacity. Empirical evidence has shown developmental improvements in a host of cognitive processes supporting working memory tasks, including speed of processing, executive functioning, resistance to interference, maintenance over delays, and more (Towse & Hitch, 2007). How can a simple mechanism like the SPH address this variety of changes? Although the DFT has not been designed to address these questions specifically, it can be tested in different tasks to examine how broadly the SPH may generalize. For example, Perone and colleagues used a related architecture to capture infants’ looking behavior in a task designed to measure VWM capacity during the first year (Perone, Simmering, & Spencer, 2011). Their simulations showed that the “young” infant model could perform similarly to the “old” infant model by removing the delays between items, suggesting that one consequence of strengthening neural interactions is improved maintenance of items over delay. Similarly, Schutte and Spencer (2009) demonstrated that “young” parameters (capturing 3-year-olds’ performance in spatial recall) showed greater spatial “drift” over memory delays, as well as increased variability across repeated trials to the same target, when compared to “old” parameters (capturing 5-year-olds’ performance). Thus, implementing a theoretical explanation like the SPH in a real-time process model allows exploration of multiple consequences that may arise from a single underlying change.

Beyond these implementations of the SPH, developmental changes in other dynamic neural field architectures have been used to account for behavioral development in other tasks. Buss and Spencer (2008) developed a multi-layered architecture of neural fields to perform the dimensional change card sort task, in which young children (typically 3-year-olds) have difficulty shifting rules used to sort cards. In their model, Buss and Spencer approximated executive control through the process of boosting the resting levels in fields used to represent the different dimensions of the cards (e.g., shape versus color) that generate the sorting rules. Through simple, quantitative changes in the magnitude of the boosts over development, they have successfully captured performance across a variety of conditions of this task.

Samuelson, Schutte, and Horst (2009) applied a dynamic neural field architecture to word learning tasks and showed that changes in the strength of inputs corresponding to object characteristics (e.g., shape versus material) captured developmental changes in children’s novel noun generalization in both yes/no and forced-choice tasks. As these examples demonstrate, the SPH is only one potential developmental change that can be implemented in a model like the DFT.

This brings us to a final question: What can our behavioral and simulation results tell us about the nature of developmental change? We contend that one of the most important implications of the present results is that we cannot answer questions about developmental mechanisms until we specify the ways in which cognitive processes operate in service of the behaviors we measure. We have highlighted how this can be achieved using computational models and how providing rich structure of the real-time processes of cognition and behavior can simplify the “work” that must be done by developmental change. Across a variety of behaviors and domains, neural field models have captured complex patterns of developmental change through relatively simple changes in connectivity in the model. The complexity of developmental change builds on the complexity of real-time behavior, and understanding the former requires a richer understanding of the latter. Computational models provide a powerful tool for understanding processes at both of these levels. The full explanatory scope of these relatively simple developmental mechanisms remains to be seen as models are applied to additional domains and tasks.
Acknowledgments

Thanks to the families who participated in this research, the preschools that allowed us to visit for data collection, and the undergraduate research assistants who assisted in data collection and processing. Thanks also to Rob Olson and Kenneth Jones for experiment programming. Data collection and preparation of this manuscript were funded by internal grant 100365 from the Wisconsin Alumni Research Foundation and National Institutes of Health (R03-HD067481) awarded to VRS. A subset of the data included here was presented at the 69th biennial meeting of the Society for Research in Child Development in Montreal, Quebec.

References


