

# Formal theories clarify the complex: Generalizing a neural process account of the interaction of visual exploration and word learning in infancy

Ajaz A. Bhat<sup>1</sup> | Larissa K. Samuelson<sup>2</sup> | John P. Spencer<sup>2</sup>

<sup>1</sup>School of Digital Science, Universiti Brunei Darussalam, Gadong, Brunei Darussalam

<sup>2</sup>School of Psychology, University of East Anglia, Norwich, UK

## Correspondence

Larissa K. Samuelson, School of Psychology, University of East Anglia, 0.09 Lawrence Stenhouse Building, Norwich NR4 7TJ, UK.  
Email: [l.samuelson@uea.ac.uk](mailto:l.samuelson@uea.ac.uk)

## Funding information

Eunice Kennedy Shriver National Institute of Child Health and Human Development, Grant/Award Number: HD045713

## Abstract

The interaction of visual exploration and auditory processing is central to early cognitive development, supporting object discrimination, categorization, and word learning. Research has shown visual–auditory interactions to be complex, created from multiple processes and changing over multiple timescales. To better understand these interactions, we generalize a formal neural process model of early word learning to two studies examining how words impact 9- to 22-month-olds' attention to novelty. These simulations clarify the origin and nature of previously demonstrated effects of labels on visual exploration and the basis of mutual exclusivity effects in word learning. We use our findings to discuss key questions for this special section: what makes a good theory and how should formal theories interface with empirical paradigms and findings?

Visual exploration is a central process in early cognitive development. As infants gaze around their environment, they gather new information about the world and strengthen existing representations. These representations form the basis for learning to discriminate between new and old stimuli, building categories, learning which sounds go with seen events, and—when labels are added to the mix—learning words. A long history of experimentation and theory in cognitive development has provided a rich picture of infants' looking behaviors with and without labels present and how those behaviors change over development. Here, we examine the developing influence of words on visual exploration in the second year, a period characterized by early, but often rapid, vocabulary development.

The prior literature has shown that the interaction of visual exploration and words is shaped by numerous aspects of the visual and auditory input and evolves over

multiple timescales from moment-to-moment changes in gaze to trial-to-trial changes as representations form, to developmental changes across age. Given this complexity, we ground our examination of these visual–auditory interactions in a formal neural process theory that instantiates the multiple processes involved in early visual exploration and word learning and how these processes change over time. We highlight how formal process theories can be usefully generalized across tasks, shedding light on the multiple processes that underlie infants' and toddlers' performance. This serves two key goals. First, we use model simulations to explore the origin of effects from two key experiments situated in a complex literature that uses multiple methods and concepts. These simulations highlight how process models can *clarify the complex*. Second, we use the simulations to address questions central to the special section: what makes a good model or theory of development and how should

**Abbreviations:** DF, dynamic field; DFT, dynamic field theory; MAPE, mean absolute percentage error; ME, mutual exclusivity; RMSE, root mean squared error; SWM, spatial working memory; VWM, visual working memory; WOLVES, Word Object Learning via Visual Exploration in Space.

The title for Special Section is **Formalizing Theories of Child Development**, edited by Willem E. Frankenhuys, Denny Borsboom, Daniel Nettle, and Glenn I Roisman.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Child Development* published by Wiley Periodicals LLC on behalf of Society for Research in Child Development.



formal theories interface with empirical paradigms and findings? We envision a future where models and experiments are tightly integrated with a rich dialog back and forth.

## Visual exploration and the impact of labels in early development

In laboratory tasks of visual processing, infant and toddler-looking preferences are influenced by a range of factors including the timing of stimulus presentation, the complexity of the stimuli, and the age of the child (Aslin, 2007; Jadva et al., 2010; Richards, 2010; Rose et al., 1982). Infants initially build a visual representation of an object, preferring to look at a familiar stimulus. Once that representation becomes stronger, infants show habituation and the emergence of a novelty preference (Hunter & Ames, 1988; Perone & Spencer, 2012, 2013). Importantly, if neither of the presented stimuli are more familiar and if they are of equal complexity, initial looking can be expected to be driven by idiosyncratic factors across infants and, thus, looking preferences at the group level should be at chance. If looking is analyzed at the level of individual infants, however, the familiarity-to-novelty pattern is seen (Roder et al., 2000). The formation of robust memories is also suggested to underlie performance in speed of processing tasks (Rose et al., 2002, 2003).

Studies examining the influence of auditory input on visual processing span a range of research topics including multimodal processing, categorization, and word learning (see Gogate & Maganti, 2016 for review). This diversity means the experimental paradigms used can vary in ways that make direct comparison or succinct conclusions difficult. For example, one line of investigation has examined whether processing of auditory stimuli “overshadows” processing of visual stimuli (e.g., Sloutsky & Robinson, 2008; Zupan & Sussman, 2009). In these studies, pairs of auditory and visual stimuli are presented in conjunction for some familiarization period. Then, either the auditory or visual stimulus is changed, and an increase in looking is used to gauge detection of the change (e.g., Robinson & Sloutsky, 2007a, 2007b; Sloutsky & Robinson, 2008). In contrast, studies of infant and toddler categorization include multiple exemplars of a category of visual stimuli and may include pre-familiarization of some stimuli (e.g., Plunkett et al., 2008). These studies may use non-linguistic sounds, novel made-up words, words unknown to the young participants, or words expected to be familiar to the participants. Moreover, these words or sounds may be presented in non-labeling (e.g., “look at that”) or labeling contexts marking them as nouns, verbs, or adjectives (e.g., “look at the blicket/its blicking/its blickish,” respectively). Finally, studies of early word learning may include a single object and word presented a handful of

times (e.g., Woodward et al., 1994), multiple word–object pairs that are presented over a series of individually ambiguous naming trials as in cross-situational word learning tasks (e.g., Smith & Yu, 2008), or a mix of novel and well-known words and objects as in studies of mutual exclusivity (ME), referent selection, or disambiguation (e.g., Bion et al., 2013; Horst & Samuelson, 2008; Kucker et al., 2018; Mather & Plunkett, 2009).

Review of this expansive literature is complicated due to the diverse methods; however, evidence suggests that by the end of the first year, infants' experiences have led them to process visual stimuli in the context of words differently from stimuli presented in the context of non-linguistic sounds (Perszyk & Waxman, 2018). And, from late in the first year onward, overshadowing of visual processing by auditory stimuli appears to decline. In categorization tasks, words appear to support the formation of categories in infants and toddlers (Perszyk & Waxman, 2018), resulting in increased looking to *novel*, non-category stimuli (see Ferguson & Waxman, 2017, for a review). In contrast, in cross-situational word learning studies, 12- to 14-month-olds demonstrate learning by looking more at *familiar* targets at test (Smith & Yu, 2008).

Although this literature shows that infants and toddlers can learn categories and words, differences in methods cloud conclusions about the influence of words on visual processing. One set of studies, however, provides a direct assessment of the influence of words on visual processing: Mather et al. (2011) compared infants' visual processing of the same novel stimuli in conditions with and without a word. In two experiments, they showed that the addition of a word slowed visual processing in 9- to 21-month-old infants, both across blocks of the experiment and within individual trials. Mather and Plunkett (2012) later manipulated object familiarity to examine how word knowledge and object novelty/familiarity interact in the formation of new word object mappings. They showed that in 22-month-old infants, the novelty of visual stimuli, rather than knowledge of familiar object names, supports referent selection. These studies help clarify the influence of words on visual processing, highlighting the operation of basic processes such as encoding, habituation, novelty detection, attention shifting, and learning. They also point to the multiple timescales upon which changes in these processes occur—from millisecond-to-millisecond shifts in attention to trial-to-trial learning as infants build word–object mappings, and to developmental changes with age and vocabulary growth.

## What is the current state of theory in this literature?

There are multiple computational models of word learning and other models of visual exploration (see

**Supporting Information**). Few models, however, explicitly integrate these processes. Moreover, few models have tried to tackle the diversity of empirical methods used in the literature. This is not surprising: it is often hard enough to develop a theory of one task/domain; generalizing across tasks presents major hurdles because the details of the stimuli across paradigms can vary, the number of word–object pairings can vary, the behaviors measured can vary, and so on. Of course, it would be particularly useful to have theories that can generalize across tasks because these could inform experiments, shedding light on which factors are key, how multiple factors/processes impact behavior, and integrating findings across studies.

One way for theories to generalize across tasks is if they are formalized as *process models*. What does this mean? Process models take time seriously, that is, they specify or formalize how processes unfold over time or across iterations. Part of this involves specifying how “inputs” change over time—which objects appear when, which auditory events occur and when, and so on. But this also requires formalizing the processes that operate on these inputs through time, processes of attention or memory formation, or learning. Finally, process models are grounded in behavior, that is, they must specify how the evolution of the model's operations through time give rise to key aspects of what children do in the task (looking, pushing buttons, responding verbally, etc.). Process models are particularly good at generalization because they often capture the ways in which empirical paradigms differ. Consider the set of experiments described above by Mather and colleagues. Across experiments, they varied whether visual events were presented in isolation or with words, they varied the number of word–object pairings, and they varied the novelty/familiarity of the presented items. Because process models must capture the details of the inputs and how these vary over trials, such manipulations should be readily captured by a process model of visual exploration and word learning.

In the remainder of this paper, we illustrate this strength of process models using a specific example: a dynamic field (DF) model of early visual exploration and word learning called WOLVES (Word Object Learning via Visual Exploration in Space; Bhat et al., 2022). This model was recently used to capture data from cross-situational word learning tasks, that is, word learning in the context of multiple, individually ambiguous naming episodes. Critically, because WOLVES is a process model, it has a way to capture the details of the auditory–visual inputs in the task, it specifies the attention and memory processes that operate on these inputs through time, and it specifies how these processes give rise to changes in looking behavior through time. Thus, in principle, using this model to explain the findings from, for instance, studies by Mather and colleagues should largely be a

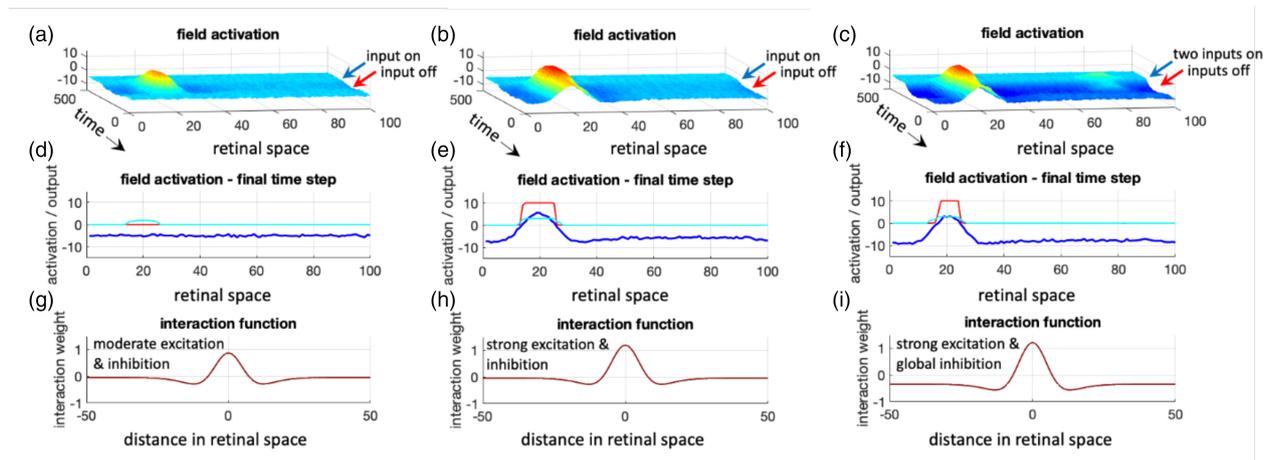
programming task: we just need to embed WOLVES in these new tasks and see what the model does. We can then ask whether the model is a good match to the child data, what matters in the model (and by analogy in children), and see if WOLVES is a useful tool, not just for fitting behavioral data, but for shedding light on what might matter in these experiments. Through this exercise, we address key questions of the special section including how process theories are formalized, implemented, and evaluated, as well as how process theories can be useful tools for understanding empirical findings.

## OVERVIEW OF WOLVES AND DF THEORY

WOLVES is instantiated in the framework of dynamic field theory (DFT; Schönner et al., 2016; see Bhat et al., 2022, for a complete model description). Thus, we begin this section with a brief overview of DFT. This is followed by an introduction to the WOLVES model.

DFT proposes that cognition arises from activation within DFs. DFs are collections or “populations” of neurons, which are wired up in a particular way that mimics how neural populations in the brain are wired up. In particular, neurons are wired up based on the perceptual, cognitive, or motor dimensions they represent. 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 be “tuned” to inputs at “left” spatial positions, while neurons on the right side of the field would be tuned to inputs at “right” spatial positions. With this setup, “neighboring” 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). Consequently, a DF can “encode” visual inputs to the network. For instance, one could present a visual stimulus 20° to the left of 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 neighbors, further driving up activation. At the same time, the excited neurons would inhibit neurons “far away” in the field, preventing excitation from growing out of control. The result will be a “peak” of activation—a local above-zero “bump” of activity centered at 20° in the field—which faithfully represents the presence of the visual stimulus at this location (see Figure 1a).

Interestingly, by changing the wiring in a DF, 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



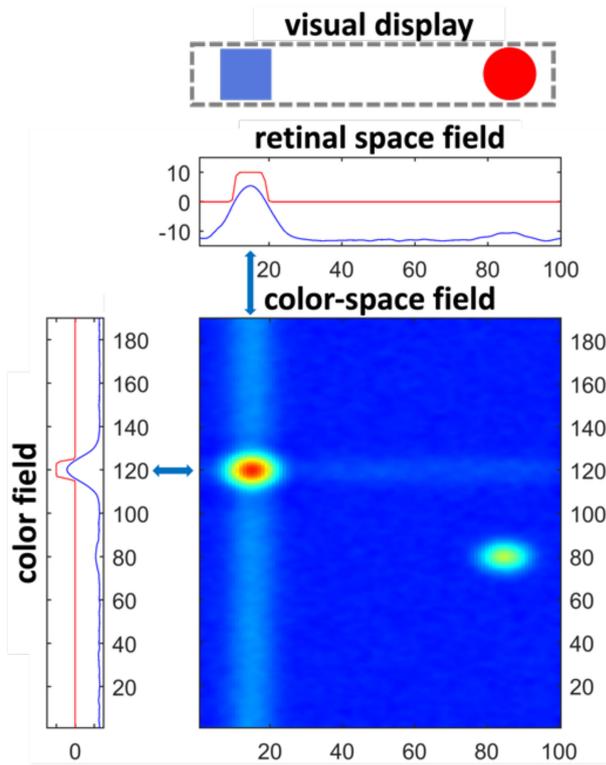
**FIGURE 1** The graphs in the top row (a–c) show how activation ( $z$ -axis) evolves through time ( $y$ -axis) in the dynamic field across locations in retinal space ( $x$ -axis) as inputs are turned on and off. The blue arrow in each panel shows when input is turned on; the red arrow shows when input is turned off. The panels in the middle row (d–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–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” dynamic field (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).

(we call such fields “input-driven,” see Figure 1a). 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*, forming only one peak at the 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 is “tuned” to particular types of information. For instance, we might have one DF that is sensitive to retinal spatial position and another that is sensitive to color (i.e., hue). In this case, how would we pass activation from the spatial field to the color field as these dimensions do not have any a priori relationship (i.e., blue objects can be on the left or right; the same for red objects)? In this case, the spatial and color fields need to be joined up into a two-dimensional DF where we have neurons “tuned” to all possible combinations of space and color (see the “color-space” field in Figure 2). In this case, the presentation of a blue object on the left would lead to a peak in the “blue-left” position of a color-space field while the presentation of a red object on the right would lead to a peak in the “red-right” position of the color-space field (Figure 2). Interestingly, we can also pass activation to and from this two-dimensional

field to separate retinal space and color fields as shown in Figure 2 (see bidirectional 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 color field), and the “blue object is on the left” (in the color-space field). Note that we do not 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. To date, these include spatial dimensions (as in Figure 2) as well as more abstract binding dimensions such as words or labels. Because WOLVES uses these binding dimensions, we return to this issue below.

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? DFT uses a variant of Hebbian learning to capture such effects. In particular, “memory traces” can form in DFs when strong 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 particular region of the field. For instance, if the model was 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 in future trials (because the memory trace boosts local excitation in this



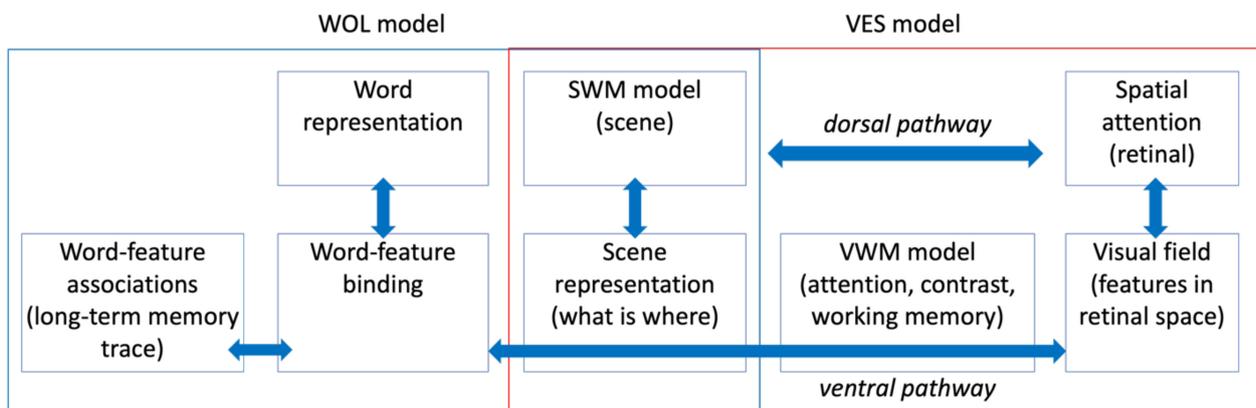
**FIGURE 2** An example of a dynamic field model with a two-dimensional color-space field that shares activation with a retinal space field (top) and a color field (left). Input from the visual display is passed directly into the two-dimensional field, mapping the location of the input along both the 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.

region of the field). This results in a “pre-shaping” effect, facilitating recognition of familiar inputs.

**Overview of WOLVES**

WOLVES integrates prior DF models. One prior model captured the processes of linking words and objects together (Samuelson et al., 2011) and others captured the processes of visual attention and memory (Perone & Spencer, 2013; Schneegans et al., 2015). By bringing these prior models together in one neural architecture, WOLVES was able to explain data from studies of cross-situational word learning, capturing data from 7 studies with adults and 5 studies with children (Bhat et al., 2022). The WOLVES model has been implemented in Matlab using the COSIVINA framework, a free object-oriented framework to construct DF architectures in Matlab, simulate the evolution of activation in these architectures, and create graphical user interfaces to view the activation time course and adjust model parameters online (Schneegans, 2012). The model code and the simulation files are available on OSF (<https://doi.org/10.17605/OSF.IO/NG2VF>), including instructions for how to run the model. The full model specification including model equations can be found in Bhat et al. (2022). We do not attempt to reproduce all these details here in favor of focusing on the general concepts of the model as well as what the modeling exercise below can tell us about the role of formal models in developmental science.

A schematic of WOLVES is shown in Figure 3. This box-and-arrow view highlights what each part of the neural architecture does. The schematic is laid out like a brain with higher-level cortical fields to the left (i.e., at the front of the brain) and lower-level cortical fields to the right (i.e., at the back of the brain). Thus, when



**FIGURE 3** A schematic of WOLVES which integrates two previous models: the word–object learning (WOL; blue box) model and the visual exploration in space (VES; red box) model. Note that the VES model is also an integration of earlier models of visual processing, including models of the neural dynamics in early visual fields, spatial attention, visual working memory (VWM), and spatial working memory (SWM).



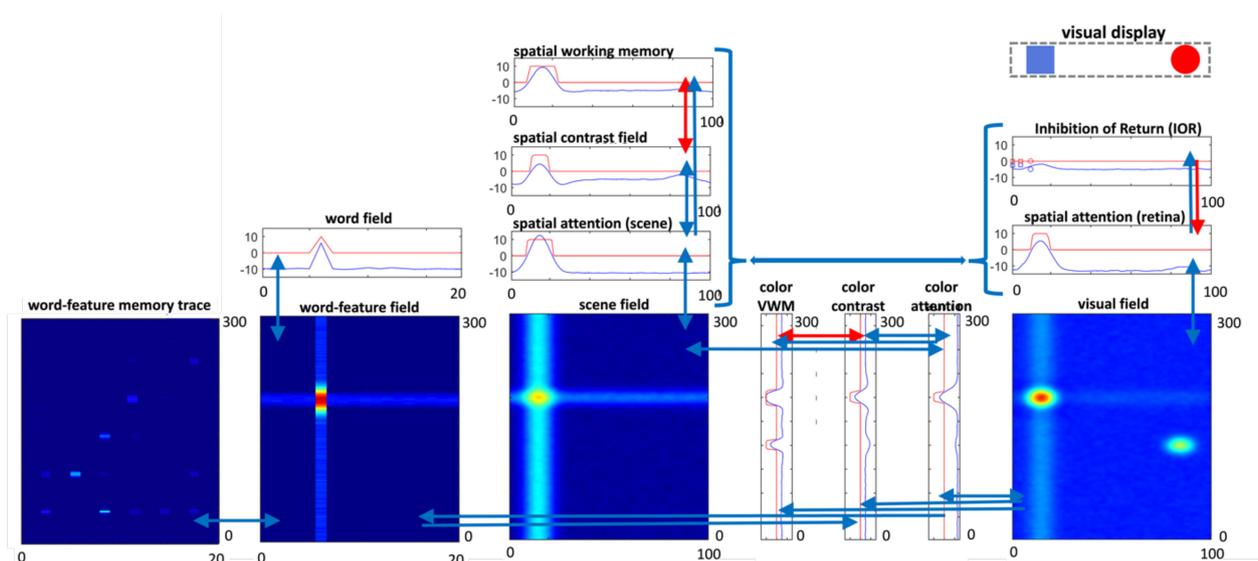
viewing a brain from the side facing left, information flows through the model in a “back to front” (or right to left) fashion. Starting in the red box to the right, one can see parts of the model that capture visual exploration in space (Perone & Spencer, 2013; Schneegans et al., 2015). Initially, visual inputs are represented in a visual field to the far right, which encodes features like color and shape and where these features are on the retina. This information is then passed along two pathways: a dorsal pathway for spatial information and a ventral pathway for feature information. Along the dorsal pathway, the model represents where objects are in the world, including a spatial working memory (SWM) for important locations within the current scene. Along the ventral pathway, the model represents object features, including a visual working memory (VWM) for which features are present. This information is re-combined into a scene representation (i.e., what is where) using focal attention (see also Treisman & Gelade, 1980).

The left side of the schematic in the blue box shows the word–object learning portion of the model (Samuelson et al., 2011). Auditory inputs to the model are fed into the word representation field which identifies the labels provided. Words and object features are then integrated into the word–feature binding field. Over time, memory traces of word–feature bindings build up in the long-term memory trace layer. These can support recognition of word–object mappings and direct looks to familiar items when a word is presented via top-down connectivity (i.e., passing activation

from the word–feature field to “lower-level” fields in the model; see long, horizontal blue arrow toward the bottom of the figure).

## Cycles of action in WOLVES

A picture of the full WOLVES neural architecture is shown in Figure 4. This shows neural activation at the moment when the model has remembered a word–object mapping and directed attention to the object in the visual field. Each box (except the visual display in the upper right) is a DF. To explain how the model works, it is useful to think about three autonomous cycles of action in WOLVES—the visual exploration in space cycle, the word–object learning cycle, and the top-down attention cycle. The visual exploration in space cycle starts when the model looks at a visual display; for instance, the blue square on the left and the red circle on the right (see “visual display”). This pattern is input to the visual field, creating an activation peak on the left at the intersection of blue/left and an activation peak on the right at the intersection of red/right. The blue and red features are then input into the contrast field (color contrast field). This field detects visual novelty (defined as any feature that “contrasts with” or fails to match the contents of working memory). At the start of a visual exploration of space cycle, all inputs are novel. At other times, peaks in VWM (color VWM) can suppress peaks in the contrast layer, differentiating what is “known” (i.e., the



**FIGURE 4** The WOLVES model neural architecture, showing one-dimensional and two-dimensional dynamic fields in the model responding to the visual display in the top right. Arrows represent uni-/bidirectional connectivity (blue: excitatory, red: inhibitory). Some working memories and memory traces are not shown for simplicity. One-dimensional fields show activation profile in blue and above-threshold neural activity in red as in prior figures. Horizontally oriented 1D fields show activation across space ( $x$ -axis); vertically oriented 1D fields show activation over the color (hue) dimension ( $y$ -axis); and the word field shows activation distributed over a “word” dimension with each site in the field representing a different auditory word. Two-dimensional visual and scene fields show activation patterns along space ( $x$ -axis) and color ( $y$ -axis), with higher activation in “hotter” colors. 2D word–feature field and word–feature memory trace show activation patterns along the word dimension ( $x$ -axis) and color ( $y$ -axis).

peak in VWM) from what is novel (i.e., the peak in color contrast).

Peaks in the contrast field pass activation to the attention field (color attention). The attention field is a “winner-take-all” or selective field and can support only one peak at a time—the focus of attention. The peak in the attention field then amplifies activation in the visual field, selecting this peak as the focus of attention and boosting spatial attention. In [Figure 4](#), WOLVES is currently attending to the blue square on the left with color attention highlighting the “blue” feature and spatial attention highlighting the “left” location. Once a visual item is attended, it is consolidated in working memory. This happens along the ventral pathway in VWM (color VWM) and along the dorsal pathway in SWM. As these peaks are consolidated in the working memory fields, activation is also passed into the 2D scene field. This field binds the feature and spatial information together, but now at the level of the spatial scene (i.e., a non-retinal position, typically anchored to the body or an external frame of reference). Once the item is consolidated at the level of the scene (i.e., a peak forms in the scene field), it is detected by the inhibition of return field. Consequently, a peak forms in the inhibition of return field at the location of the currently attended item. This inhibition of return peak, in turn, suppresses spatial attention, releasing the current item from the focus of attention. The model is then free to attend to another novel item detected by the color contrast field and the visual exploration in space cycle repeats.

The word–object learning cycle is engaged when an auditory word is presented, activating a word in the word field. The peak in the word field sends input into the word-feature field. If the model is currently attending to a visual object, the features of this object (i.e., blue) are simultaneously projected into the word-feature field. When these inputs combine, they can build a peak of activation in the word-feature field, “binding” the word to the visual features (see hotspot of activation in the word-feature field). This peak leaves a trace in the memory trace layer. Such traces build slowly over time and decay even more slowly. With repeated presentations of the word and the visual object, the model learns the mapping. Note that the model in [Figure 4](#) has been presented with multiple words and multiple objects, so there are many mappings in the word-feature memory trace (see light blue dots). Some of these are correct; some are incorrect. This reflects the ambiguity of the cross-situational word learning task on which WOLVES was initially benchmarked.

The final cycle central to how WOLVES works is the top-down attention cycle. When the model has a strong memory trace for a word–object mapping, presenting that word causes a peak to form in the word-feature field at the location of the strong memory trace. This peak then passes a “top-down” signal from the word-feature field to the color contrast field, causing

strong activation of the associated feature in the contrast field and directing attention to the object via interactions with the color attention field. This is how the model directs attention to a visual object when hearing a known word.

### Autonomous visual exploration over multiple timescales in WOLVES

The visual exploration in space, word–object learning, and top-down attention cycles emerge over multiple timescales as neural activation propagates in the model as it performs the task. On the real timescale of milliseconds and seconds, the model autonomously looks back and forth between the visible objects in the visual field (i.e., it attends to the left item, then attends to the right item, and so on). It also recognizes words and binds visual features and words together. At this level, the model's real-time behavior can be mapped to participants looking behavior, enabling us to embed the model in the same visual and word learning scenarios as participants simply by instantiating the same presentations of visual and auditory inputs and asking if the model “looks” like children and adults do.

Over a learning timescale, we can ask if WOLVES learns as participants learn. Here, it is important to emphasize that learning is not isolated to the word-feature memory trace in [Figure 4](#). In fact, every field in the neural architecture (except the visual field, the attention fields, and the inhibition of return field) has a memory trace. This has consequences for each cycle. For the visual exploration in space cycle, learning is felt most strongly in the working memory fields. Here, strong memory traces can support the fast consolidation of peaks in working memory as well as robust maintenance of these peaks. This can alter the visual exploration in space cycle, causing the model to quickly release fixation for “known” items and, consequently, spend more time looking at novel items later in learning. This is the basic process underlying visual habituation in WOLVES (see [Perone & Spencer, 2013](#)).

Learning also impacts the word–object learning cycle. Here, learning cascades on itself, helping the model to figure out the statistics of correct word–object co-occurrences. Critically, as the model forms robust word–object memories, these can “block” the formation of new erroneous associations. This mechanism is particularly evident in adult studies of cross-situational word learning (for discussion, [Bhat et al., 2022](#)). Finally, learning impacts the top-down attention cycle as strong memory traces in the word-feature fields can lead to a strong “top-down” signal to the contrast field which drives looking to the labeled object.

We have also used WOLVES to understand how visual exploration and word learning change on a developmental timescale. In particular, [Bhat et al. \(2022\)](#) proposed



that memory traces decay less quickly over development. Consequently, models of older infants and children show faster habituation to repeated visual stimuli because memory traces build more robustly (i.e., they decay less quickly). Likewise, if words are repeatedly presented with objects, individual word–object mappings will build more robustly / decay more slowly. This allows individual word–object associations to have more of an impact on visual exploration as it is less likely the representation of the mapping will have decayed fully before the word–object pairing is encountered again. Thus, models of older infants and toddlers will show faster learning of new word–object mappings and more top-down attention to familiar stimuli.

In summary, WOLVES is well positioned to advance our understanding of how core processes of attention, working memory, long-term memory, and word–object associative learning create patterns of visual exploration, novelty detection, habituation, and word learning biases over the timescales of a trial, an experiment, and development. In what follows, we apply WOLVES to studies by Mather et al. (2011) and Mather and Plunkett (2012) examining how words impact visual exploration and how novelty influences word learning. In so doing, we highlight how process models like WOLVES can be generalized beyond their initial testing context.

This exercise also highlights several issues central to the special issue regarding how models and data “talk” to one another. For instance, we note that the data from studies by Mather et al. (2011) and Mather and Plunkett (2012) are variable as is often the case with infant data from looking experiments. Thus, we highlight what is statistically significant in these studies below, but we acknowledge that these findings have not yet been replicated. We also note that our goal was not to ask if the model could precisely quantify all details of the empirical data. This would require optimizing the many parameters in WOLVES (an issue we return to in the General Discussion). Instead, we varied a small set of parameters and asked how the model captures the patterns from these studies. Although this limits the modeling exercise to a more qualitative, exploratory comparison (rather than a detailed statistical analysis of the model data), it emphasizes how process models can have a useful dialog back and forth with data even as understanding of empirical phenomena is still developing. Process models can inspire new ways of thinking about different aspects of the data even when their fit to data is not (yet) optimal, and data can provide useful constraints for models, even when replication has not yet happened.

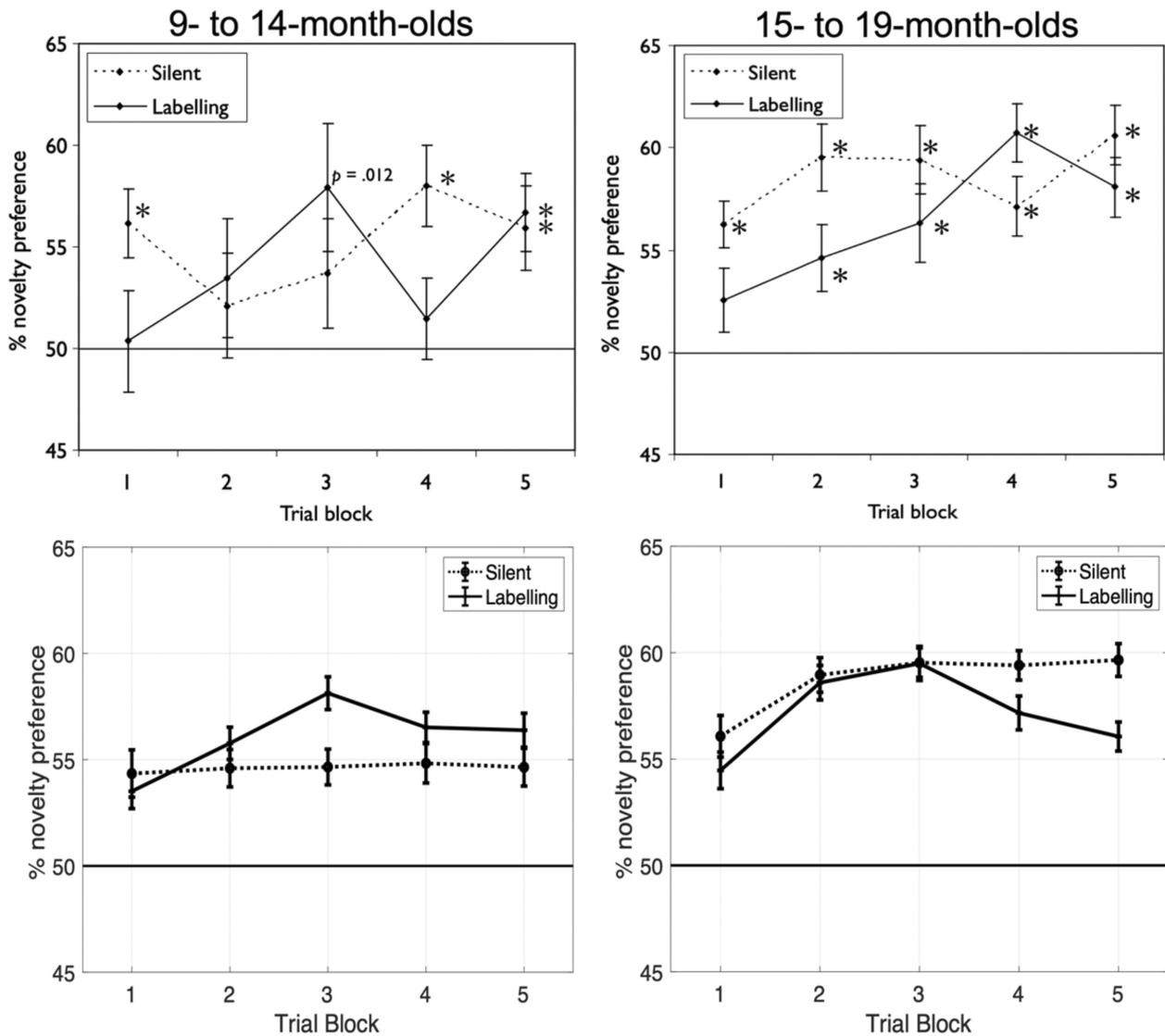
## **SIMULATION STUDY 1: MATHER ET AL. (2011), EXPERIMENT 1**

The goal of this study was to examine whether auditory disruption of visual processing, seen in prior studies

with younger infants, occurred during the developmental period of early vocabulary growth. Prior studies had found that infants between 8 and 12 months of age processed visual input better when they were presented in silence than when presented with accompanying sounds (Sloutsky & Robinson, 2008). These findings were taken to suggest that when visual and auditory input are presented together, infants preferentially process the auditory input, overshadowing the visual input. However, other work suggested that labels enhanced processing of visual stimuli, facilitating infants' abilities to categorize stimuli, and that this might particularly be the case after 12 months of age as vocabulary development progresses (see Perszyk & Waxman, 2018).

To examine these possibilities directly, Mather et al. (2011) used a visual paired comparison procedure to observe 9- to 19-month-old infants' processing of novel objects with or without accompanying label input. On each trial, infants were presented with two novel cartoon-style creatures in either a silent or labeling condition. One of the two creatures repeated on each of the 30 trials and the other was a different new creature on each trial. Mather et al. (2011) hypothesized that attentional resources are shared across the auditory and visual modalities and therefore that there would be a disruptive effect of labels on visual processing of novel objects, especially during early stimulus exposure. They used suppression of novelty preference in the context of words as the indicator of this disruption. The data were separated by age for analysis into younger (9- to 14-month-olds) and older (15- to 19-month-old) groups. The dependent variable was the novelty preference, calculated as proportion looking at the novel stimulus divided by total looking on a trial. Data were grouped into five blocks of six trials (note that block 1 had five trials because the first trial, when both stimuli were novel, was dropped from analyses). The results are presented in the top two panels of Figure 5.

Mather et al. (2011) reported the following effects in their paper as verified with one-tailed *t*-tests following a significant three-way interaction in an ANOVA comparing performance across trial blocks, age groups, and conditions: (a) younger infants showed a stronger novelty preference in the silent condition than in the labeling condition during the first trial block ( $t(34) = -1.94$ ,  $p = .031$ ,  $d = .64$ ) and (b) the fourth trial block ( $t(34) = -2.32$ ,  $p < .015$ ,  $d = .76$ ); (c) older infants showed a stronger novelty preference in the silent condition than in the labeling condition during the second trial block ( $t(34) = -1.82$ ,  $p < .038$ ,  $d = .55$ ). In addition, Mather et al. compared performance in each condition to chance levels using one-tailed *t*-tests with alpha set at .01 to correct for multiple comparisons. They found: (d) in the silent condition, younger infants showed a novelty preference during the first trial block and in the fourth and fifth trial blocks; (e) in the labeling condition, younger infants showed a marginal novelty preference in the third block and a significant novelty



**FIGURE 5** Top two panels show empirical data from Mather et al. (2011) Experiment 1. Infants were grouped into a younger, 9- to 14-month-old, and older, 15- to 19-month-old, groups. Data are shown as function of condition and trial block. Error bars are  $\pm 1$  SE; \*comparison to chance  $p < .01$ , one-tailed. The bottom panels show simulation data for corresponding age groups.

preference in the fifth block; (f) in the silent condition, older infants showed a significant novelty preference in all blocks; and (g) in the labeling condition, older infants showed a significant novelty preference in all blocks except block 1.

**Simulation methods**

WOLVES is a process model; thus, at the most basic level, simulating the Mather et al. (2011) study is a programming task—we have to write code to embed the model in the same task infants experienced. This involves exposing the model to the same sequence of auditory and visual events infants experienced on each trial, recording the same behavioral measures (i.e., looking left and right), and then iterating through all the trials in the experiment with the same trial

randomization. WOLVES has two feature pathways along the ventral pathway shown in Figure 4—one for color and one for shape. Gaussian inputs were used to represent the color and shape features of the visual input to the visual field; Dirac functions were used to capture auditory inputs to the word field. We assumed all the objects and words were distinct and randomly spaced across the shape, color, and word fields. Each simulation time step was scaled to equal eight real-time milliseconds as in our prior work. Thus, we turned inputs on and off to mimic all timing details in the experiment. We recorded looking in the model at each time step based on whether the model was attending to the left, right, or away (i.e., whether there was a peak in spatial attention on the left, right, or no peak/a peak in some other location).

The next question was how many “infants” to run. One approach is to run the same number of models as

infants in the experiment; however, as in experiments, if you run too few infants, estimates of performance will be noisy. In experiments, we might conduct a power analysis to determine the sample size. In a model, we can do something similar—run batches of simulations with different numbers of infants and evaluate when the model's performance converges on roughly the same values from batch to batch. Bhat et al. (2022) conducted such an analysis and found that 300 simulations were needed to converge on stable performance. The reason is quite simple: to model infants' performance, we need a noisy model, and to sample the behavior of that noisy model robustly, we need a lot of iterations. We used the same simulation approach here and conducted 300 runs of the model for each condition/age group.

Running 300 simulations per condition/age group takes a lot of time. Thus, we used parallel computing to speed things up. With DF models, we cannot parallelize the computations for each model/infant because what the model does on each time step influences the next time step; however, we can run each individual simulation on a separate processing “core.” In an ideal world, we would have 300 cores available; thus, running one “batch” of infants would take roughly the same amount of time as running one infant model. Instead, we ran the simulations on two machines: an Intel i5 processor-equipped PC with 36 parallel processing cores and a high-performance cluster with 96 parallel processing cores (RSCS-UEA, 2015). All simulations were conducted in Matlab 2016b using the COSIVINA framework (Schneegans, 2012; Schöner et al., 2016). To quantitatively evaluate the model's performance, we computed two standard metrics: the root mean squared error (RMSE) and mean absolute percentage error (MAPE) between the simulated and empirical data. Note that we opted to evaluate the fit of the model qualitatively in the present report, rather than statistically. Statistical comparisons with a model  $N$  of 300 would not be terribly meaningful as even small differences would be significant. If we had more detailed information about the variance in the data, we could make a more detailed quantitative comparison. As this was not available—and as the data have not yet been replicated—we think a more qualitative comparison is the best approach.

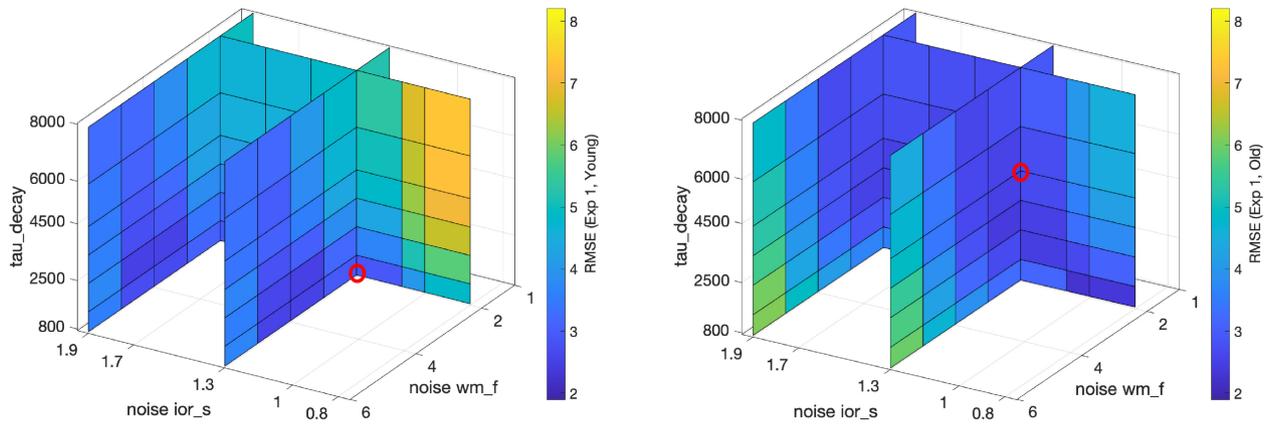
Even with a more qualitative comparison, we must evaluate whether the model “worked,” that is, does the model show a low RMSE/low MAPE score when compared to the young and old infants in the Mather et al. study and did it reproduce the right qualitative patterns. To evaluate this, we must first “tune” the developmental parameters of the model to fit the data from infants because the specific ages examined in Mather et al. did not precisely match the ages examined in Bhat et al. (2022). In that prior work, Bhat et al. (2022) captured developmental changes by changing the *tau\_decay* parameter, with lower values of *tau\_decay* (faster forgetting) for younger infants. Concretely, then, we had to examine a

range of *tau\_decay* values, picking the values that minimized RMSE/MAPE across the two age groups such that *tau\_decay* of “younger” models would be lower than *tau\_decay* of “older” models.

After doing this exercise, we noted that the model did not show a strong enough change in the novelty bias over learning. In some ways, this is not surprising: WOLVES was previously optimized to capture data from cross-situational word learning studies and the Mather et al. task is rather different. But how should we modify WOLVES to better capture the data? WOLVES is a complex neural architecture, with many parameters, so which parameters should we change? When we looked at the model's performance, we noticed two key things. First, the model oscillated its looking left and right more regularly than infants did (e.g., the model did not show enough variation in looking, with some short looks and some long looks). Second, the model formed robust working memories for the objects quite quickly; consequently, there was not much change in looking dynamics over learning. This insight came from our prior work on visual habituation in a “speed of processing” task analogous to the “silent” conditions of Mather et al. (see Perone & Spencer, 2013). That work showed how the slow buildup of memory traces over learning gradually supported peaks in working memory fields leading to habituation.

There are probably multiple ways to alter these two problematic behaviors in WOLVES. We opted to explore increasing the amplitude of noise in the model. This was appealing in that noise is something hypothesized to change in development (with stronger noise early in development; see, e.g., Spencer et al., 2022). Thus, “tuning” noise is a bit like “tuning” *tau\_decay*: the baseline parameters of WOLVES can remain fixed but we can vary developmental parameters to optimize the fit of the model to the age groups of interest. We modified noise in two ways. First, we increased the noise amplitude in the inhibition of return field (*ior\_s*) to introduce variability in the release from fixation (to create some short looks and some long looks). Second, we increased the noise amplitude in the feature working memory field (*wm\_f*). This created variability in the stability of working memories for recently seen visual stimuli creating more habituation-related change in the model's looking over trials.

To fit WOLVES to the empirical data, we conducted a  $7 \times 7 \times 7$  grid search, varying three parameters (*tau\_decay*, *ior\_s* noise strength, and *wm\_f* noise strength) to find parameter values that resulted in minimal RMSE for each age group (see Figure 6). The range of values for exploration of each of these parameters was chosen around the baseline value used by Bhat et al. (2022). Specifically, we varied noise in the inhibition of return field between 0.8 and 1.9 (baseline value = 1.0), the noise in the working memory feature fields between 1.0 and 6.0 (baseline value = 1.0), and memory decay parameter *tau\_decay* between 800 and 8000 (baseline values for 14- to 68-month-old models varied from 800 to 5000



**FIGURE 6** Plots of the volumetric data of root mean squared error (RMSE) values between the model and empirical data against three parameters (noise in the inhibition of return field [*ior\_s*], noise in the working memory feature fields [*wm\_f*], and *tau\_decay*). Left panel plots model RMSE from the data from 9- to 14-month-olds (‘Young’) and the right panel shows RMSE against 15- to 19-month-olds (‘Old’). Points highlighted in red circles indicate the best-fitting parameters for the two age groups (see text for details).

**TABLE 1** Parameter values and model fit metrics for the best-fitting grid search parameters for the “younger” model (9–14 months) and three best-fitting model parameters for candidate “older” models (15–19 months) when fit to data from Experiment 1 of Mather et al. (2011).

	<i>ior_s</i>	<i>wm_f</i>	<i>tau_decay</i>	RMSE	MAPE
9–14 months	1.3	2	800	2.48	3.91
15–19 months param set 1	1.3	1	800	2.13	3.21
15–19 months param set 2	1	1	800	2.22	3.24
15–19 months param set 3	1.3	2	4500	2.33	3.29

Abbreviations: MAPE, mean absolute percentage error; RMSE, root mean squared error.

**TABLE 2** Parameter values and model fit metrics for parameter sets 1 and 3 for candidate “older” models when fit to the 16-month-old data from Experiment 2 of Mather et al. (2011).

	<i>ior_s</i>	<i>wm_f</i>	<i>tau_decay</i>	RMSE	MAPE
16 months parameter set 1	1.3	1	800	4.55	6.97
16 months parameter set 3	1.3	2	4500	3.51	5.29

Abbreviations: MAPE, mean absolute percentage error; RMSE, root mean squared error.

with the “adult” baseline value set at 15,000). For each one of the  $7 \times 7 \times 7$  parametric combinations, the model was simulated for 300 runs, resulting in 205,800 simulations for the two task conditions in experiment 1. Figure 6 plots the RMSE values when the model is fit to data from “younger” children (left panel) and “older” children (right panel). The plots show that the RMSE values vary smoothly and systematically as the three parameters are varied, indicating that the model’s behavior changes systematically across parameter space.

Data points corresponding to the best-fitting parameters for the “younger” model are circled in red in the left panel of Figure 6 (for details, see Table 1). For the “older” model, we examined the three best-fitting parameter sets in more detail because they all had comparable errors yet suggested different developmental hypotheses. As can be seen in Table 1, parameter sets 1 and 2 suggest that noise is reduced over development. In the first parameter set, noise is reduced in the working memory fields.

In the second, noise is reduced in the inhibition-of-return field and in the working memory fields. We note that Simmering (2016) proposed a DF model of VWM where noise reduction in working memory was a key aspect of the developmental changes in parameters. The third parameter set suggests that *tau\_decay* changes over development. This is consistent with the developmental proposal set forth in Bhat et al. (2022).

Given that these best-fitting parameter sets for the “older” model suggest two different developmental hypotheses—noise reduction and change in *tau\_decay*—we explored parameter sets 1 and 3 further. Note that parameter set 2 was not explored further as the other noise reduction parameter set (set 1) fit the data slightly better.

In the next step, we examined the fit of parameter sets 1 and 3 to the 16-month-old data from experiment 2 of Mather et al. (2011). Full details of these simulations can be found in Supporting Information. As can be seen in Table 2, parameter set 3 fared considerably better in this

generalization experiment. Across experiments 1 and 2 from Mather et al. (2011), parameter set 1 had a mean MAPE of 5.09 while parameter set 3 had a mean MAPE of 4.29—a considerable improvement. Thus, we report results for parameter set 3 as our primary findings. Note that parameter set 3 is circled in red in the right panel of Figure 6.

## Results

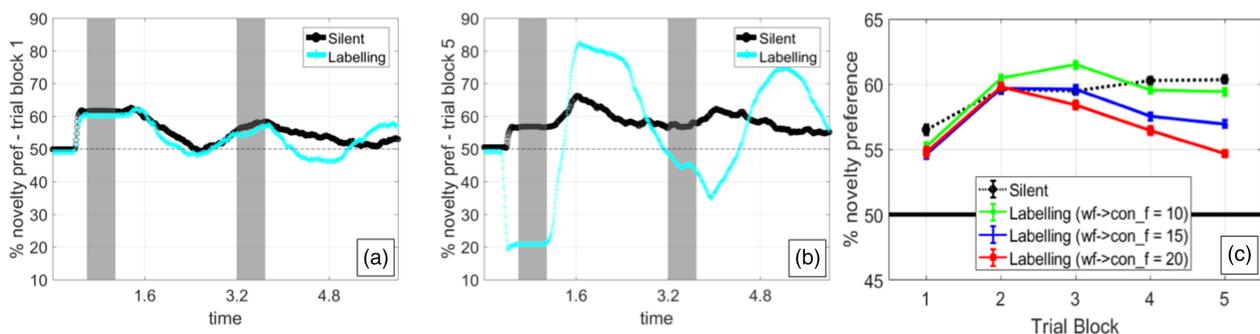
Simulation results from the best-fitting parameters for the “younger” and “older” models are shown in the bottom panels of Figure 5. To facilitate more meaningful comparisons between the model and the data, we computed standard errors by running 10 random permutations on the data, sampling sets of 40 simulations (the  $N$  from Mather et al., 2011), and computing a mean standard error over these permutations.

Overall, WOLVES shows a solid fit to the empirical data; the RMSE and MAPE for all data points are 2.40 and 3.61, respectively. In terms of comparisons between the silent and labeling conditions, recall that younger infants showed a stronger novelty preference in the silent condition than in the labeling condition during the first trial block (effect “a” in the summary above). This is the case in the model as well, although the effect is quantitatively small. The younger model does not capture the condition difference for the younger infants in the fourth trial block (effect “b”). The older model shows a stronger novelty preference in the silent condition than in the labeling condition during the second trial block, reproducing effect “c”; again, however, the effect in the model is quantitatively small. In terms of above-chance performance, the model is robustly above chance for all trial blocks. Recall that for the younger infants, blocks 2 and 3 in the silent condition were not above chance, and blocks 1, 2, and 4 in the labeling condition were not above chance. The younger model is relatively flat in the silent condition, so it fails to capture these variations (effect “d”), but it

does reproduce the lower performance in blocks 1, 2, and 4 of the labeling condition (effect “e”). In addition, the older model shows a strong novelty preference over trials (effects “f” and “g”), particularly in the silent condition with some decline in block 5 for the labeling condition. It also shows the smallest novelty preference in block 1 of the labeling condition; recall that this was the only data point that was not robustly above chance for the older infants (effect “g”).

To more closely examine how the word presentation impacts novelty preference over the course of the experiment, we examined the model's looking dynamics within trials. Figure 7a,b show the time course of percentage looking to novel stimuli across the 6000 ms of a trial for the models simulating the older age group. As can be seen in Figure 7a, in both the silent (black curve) and labeling (cyan) conditions, the model looks to the novel stimulus at the start of the trial with a second look to the novel item roughly 3 s into the trial. During the first block of the trials in both conditions (Figure 7a), the memory traces associated with the contrast fields, scene attention fields, and word-feature fields (see Figure 4) are just forming; thus, these memory layers do not have much effect on where the model looks initially. Why, then, does the model tend to look at the novel item first? This is caused by sustained activation in the working memory fields as the model tends to maintain peaks for the features of the repeated stimulus.

As the experiment proceeds, the memory traces and model dynamics drive looking differently for the silent and labeling conditions. In the silent condition, looking is like earlier in the task, but with a stronger novelty preference overall, particularly early in the trial. This reflects habituation to the repeated item driven by stronger working memory peaks for the repeated features supported by the associated memory traces. In the labeling condition, however, the memory traces for the word-object associations become stronger and counteract habituation, directing attention to the repeated object at the start of the trial. This can be seen in panel b as a sharp decrease in looking at the novel object in the labeling



**FIGURE 7** Panel a shows percentage preferential looking by the model to the novel object during the 6 s time course of a trial for both conditions, aggregated over trials 2–6 of the first block. The vertical gray rectangles indicate timing and duration of label presentation. Panel b shows a similar time course aggregated over the last block of six trials. Panel c shows the percentage novelty preference when the parameter controlling top-down attention is manipulated from weak (10) to strong (20).

condition (first big dip, cyan curve). Interestingly, the word–object associations are strong enough to cue attention even before the word is presented. This reflects how subtle factors can bias looking at the start of the trial when the model is not currently looking at either item.

During the remainder of the trial, every time the word is presented, it activates top-down attention again. This drives the second dip in the cyan curve in [Figure 7b](#), just after the second word presentation. At this moment, the model has just released attention from the novel item; thus, top-down attention drives the model to look to the repeated object. This is followed by a final cycle of looking at the novel item as fixation is released from the repeated item.

The top-down attention effect on novelty preference is further illustrated in panel c of [Figure 7](#). Here, we manipulated the top-down attention parameter (i.e., the strength of the projection from the word-feature field to the feature contrast field, see long horizontal blue arrow in [Figure 4](#)). This parameter controls the extent to which acquired word–object associations modulate looking behavior in the system with lower values meaning word–object associations have less impact on looking. As [Figure 7c](#) shows, in the first block of the experiment, differences in this parameter do not influence novelty preference much. However, as word–object associations grow, higher values of top-down attention allow those associations to direct attention more to the repeated object, and the model shows less novelty bias. Nevertheless, the strong influence of habituation is also evident in the figure—the novelty preference shown by the model is still above chance, even with a top-down attention parameter 5 units stronger than the value used in our baseline simulations (15). Thus, while words clearly influence visual processing, they do not completely override the tendency to attend more to novelty.

## Discussion

Data from the empirical study by Mather et al. (2011) suggest that the repeated presentation of a label decreased the strength of infants' novelty preference. Thus, in a direct comparison of conditions that only differed in the inclusion of labels, labels “disrupted” visual processing and slowed the development of a novelty preference (Mather et al., 2011). In the silent condition, models and both the younger and the older infants showed a novelty preference early in the first block of trials. In the labeling condition, in contrast, the novelty preference in the first block was lower in both infants and models, although this effect was quantitatively small in the model. This was also the case in block 2 for the labeling condition for older infants and models. The model failed to capture the increase in novelty preference in the silent condition in block 4 for the younger infants and, more generally, failed to show much modulation across trial blocks in

this condition. The younger model showed more modulation in the labeling condition, capturing, for instance, the increase in novelty preference in block 3. For the older model in the labeling condition, the model showed the lowest novelty preference in block 1, consistent with the data. Overall, then, the model's performance was mixed: it captured some of the statistically significant effects from the data but missed other significant patterns.

Nevertheless, the model can inform our understanding of how words impact visual processing. Mather et al. (2011) reviewed two possible causes for the impact of labeling. One possibility is that because attentional resources are shared across visual and auditory modalities, labeling introduces additional load that suppresses processing of visual novelty. Alternatively, they suggested that separable visual and word learning processes competed in directing infants' gaze, and this competition slows visual processing. Our simulations with WOLVES support and augment the later explanation. Labeling results in the formation of word–object associations over the first few trial blocks. These associations have a top-down effect on looking behavior and can drive looking to the corresponding object. At the same time, the visual exploration of the repeated object leads to habituation to that object, directing attention toward novelty. Thus, word learning “competes” with visual habituation, competition which is manifest in the contrast fields as these fields receive both top-down input from the word-feature fields and inhibitory input from working memory (see [Figure 4](#)). Note that WOLVES suggests that looking less at novel objects in the task is not necessarily indicative of slower visual processing but a combined effect of preference for novelty, created by habituation, and preference for familiarity created by learning the statistical relationships between words and objects.

This conclusion is further supported by simulations of Mather et al.'s (2011) second experiment (reported in [Supporting Information](#)). In that study, Mather et al. increased the number of repeating objects and labels, arguing that it would increase the processing demands and result in a reduction in the novelty bias. A reduction in the novelty bias was indeed found in the empirical study and captured in WOLVES. Importantly, however, this reduction occurs in both the labeling *and* silent conditions. WOLVES provides additional insight on this finding as it shows how the stated experimental manipulation—increasing the number of objects and words presented—results in reducing the frequency of presentations of each object which results in less habituation to the repeating objects and, thus, less novelty bias in both conditions. This demonstrates, then, how the application of process models that capture the details of tasks and experiments over time can help unpack how changes to an experiment produce changes to participant behavior.

Note that we restricted the grid search to three parameters (two noise parameters and *tau\_decay*) to explore whether WOLVES could effectively generalize to

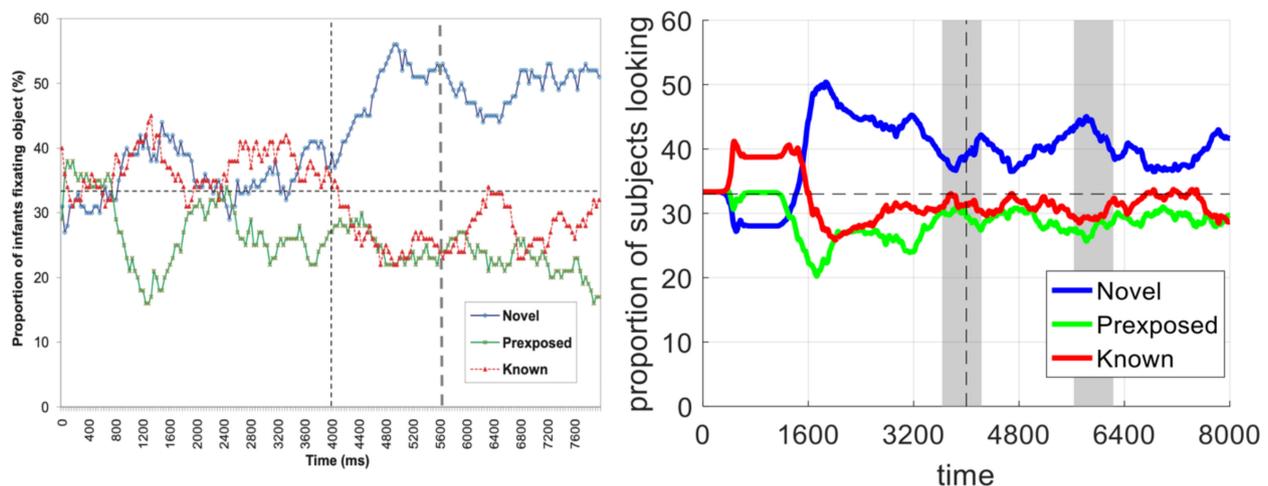
a new paradigm. Simulation results were mixed; despite this, the model was still a useful tool in clarifying the factors that might underlie infants' performance in this task. Thus, even though data from Mather et al. (2011) are variable and have not yet been replicated, and even though our quantitative fit was not extremely precise, we can still see a useful dialog back and forth between theory and data. It is important to acknowledge that there are likely other parameter sets that fit the data from the present experiment with greater precision. Given that there are no global optimization methods available for DFT yet, parameter tuning is a time-consuming task. This is clearly an important area for future work (see General Discussion for additional details).

## SIMULATION STUDY 2: MATHER AND PLUNKETT (2012), EXPERIMENT 1

In this experiment, Mather and Plunkett (2012) examined the role of novelty in early word learning with the aim of elucidating the processes underpinning the ME bias. ME is the bias children demonstrate when they associate novel words with new objects (Markman, 1989, 1990). With young infants, ME is frequently explored via looking-based tasks wherein infants are presented with images of one or more familiar objects they would already know the name of along with a novel “name-unknown” object. Across trials, known and novel names are presented and the stimulus that infants look to more is taken to be the one they have mapped to the word (see, e.g., Bion et al., 2013; Halberda, 2003). The typical finding is that infants will look to the novel object when novel names are presented, thus demonstrating a bias to associate novel words with objects that do not yet have a name (see Mather, 2013, for review).

As reviewed by Mather and Plunkett (2012), much research has examined the basis for this bias, but many of these studies use “name-unknown” objects that are also completely novel to the infant, confounding novelty and name knowledge. Mather and Plunkett untangled these factors by presenting three objects on each testing trial—a familiar object, a pre-exposed name-unknown novel object, and a completely novel name-unknown object. The experiment began by familiarizing infants with two novel objects each presented individually six times for 4000 ms. No names were presented during these familiarization trials, such that on subsequent novel label trials, these objects were pre-familiarized but name unknown. On each of the six novel label trials, three objects—one of the pre-familiarized name-unknown objects, a completely novel name-unknown object, and a familiar, name-known object were presented together for 8000 ms. During these trials, one of two novel names was presented with onsets of 3633 and 5633 ms thereby dividing the trial into a 4000 ms pre-naming baseline period (0–4000 ms) and a post-naming period (4000–8000 ms).

The data are presented in the left panel of Figure 8, aggregated across trials. For analysis, infants' total fixation times to each object across the 8000 ms test trial were measured and proportion of infants fixating at each object in each 40 ms time bin was calculated in the pre- and post-naming phases. Mather and Plunkett (2012) reported the following main effects in their paper: (a) infants displayed a significant preference for the novel object over the pre-exposed object during pre-naming phase ( $t(24)=3.05$ ,  $p<.01$ ,  $d=.61$ ); (b) infants also displayed a significant preference for the novel object over the pre-exposed object during post-naming phase ( $t(24)=4.72$ ,  $p<.001$ ,  $d=.94$ ); (c) prior to naming, infants had a significant preference for the known object over the pre-exposed object, ( $t(24)=2.73$ ,  $p<.015$ ,  $d=.55$ ); (d) with naming, there was no difference in attention to



**FIGURE 8** Time series of proportion of infants (left) and models (right) attending to the known, novel, and pre-exposed objects over the course of the 8000 ms novel label trials. The vertical dashed lines in the left panel and the right side of the gray bars in the left panel indicate naming offset. Horizontal dashed lines indicate chance.

the known object and the pre-exposed object ( $p > .5$ ); (d) prior to naming, there was no significant preference for the known and novel objects; (e) with naming, attention to the novel object over the name-known object increased significantly ( $t(24) = 4.75$ ,  $p < .001$ ,  $d = .95$ ); and (f) the increase in preference for the novel object pre-naming versus post-naming was significant ( $t(24) = 4.15$ ,  $p < .001$ ,  $d = .83$ ; for related findings, see Horst et al., 2011; Kucker et al., 2020). Mather and Plunkett (2012) argued these data provided compelling evidence that novelty is the mechanism behind ME as infants looked most to the novel object and least to the pre-exposed object.

## Simulation methods

We simulated the task in WOLVES with the same trial structure and timing details reported in the empirical work. We used the same simulation methods and parameters used for Mather et al. (2011), Experiment 2 (see [Supporting Information](#)), including the *tau\_decay* values that correspond to the 22-month-olds tested. Since this task involves presentation of *known* objects to participants, we started models with two known word–object associations. The strength of these memories was determined via an iterative search to minimize RMSE that ended in a final strength value = 0.3. Likewise, the model was provided with pre-familiarization trials with two novel objects that were not named before the novel name trials. Thus, on novel label trials, the model, like infants, was presented with one name-unknown pre-familiarized object, one name-known familiar object, and one completely novel object along with a novel name. Note that, inspired by the grid search from the first simulation experiment, we explored whether a noise reduction parameter set might improve the fit of the model to this experiment. This was not the case: a model with working memory feature field noise equal to 1.0 instead of 2.0 (with inhibition of return noise = 1.3 and *tau\_decay* = 8000) did not fit the data better (RMSE = 6.13, MAPE = 9.65).

## Results

The simulation results are presented in the right panel of [Figure 8](#). WOLVES captured the infant data reasonably well with overall RMSE and MAPE values of 4.83 and 7.66, respectively. The model looks most to the novel object and least to the pre-exposed object, consistent with infants' behavior. In the pre-naming phase of the test trials, WOLVES looks more to the name-known and novel objects compared to the pre-familiarized object like infants. However, the model's looking pattern is more systematic—unlike infants, the model consistently looks at the name-known object first and then releases fixation to look at the novel (name-unknown) object. This reflects top-down attention caused by the relatively

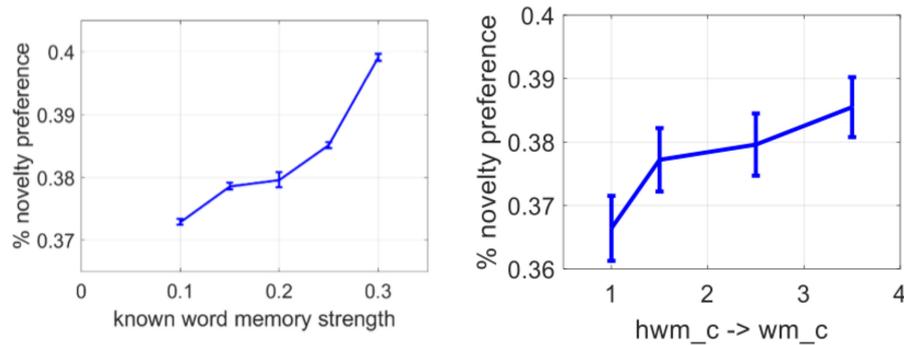
strong word–object associations for the name-known object like the bias to look toward the repeated item late in learning in the labeling condition of Simulation Study 1 (see [Figure 7b](#)). From the list of effects in Mather and Plunkett (2012), as reported above, WOLVES reproduced effects “a,” “b,” “c,” and “e.” Effect “d” showed a non-significant difference between the novel item and the name-known item during the pre-naming period. The model, by contrast, shows more looking at the novel item during this period. In addition, the model failed to show effect “f”—there was no clear increase in looking at the novel item in the naming period.

We suspect it would be possible to create more variability in the model's initial looking preference to more closely mimic infants' performance which may also help capture effects “d” and “f.” We opted not to pursue such parameter tuning as this simulation study was focused on how process models can be usefully generalized to new paradigms, rather than on producing an optimal fit to all empirical details. Even without detailed parameter fitting, WOLVES shows shifts in attention to the pre-exposed name-unknown object in the pre-naming phase similar to the pattern shown by infants. In the post-naming phase, the model showed a clear preference to look at the novel object more than either the name-known or pre-exposed name-unknown objects, with some increase in looking at the novel item time locked to the naming instances. Thus, consistent with infant data, the model showed a preference for the *most novel* of the two name-unknown objects after hearing the novel label.

To establish whether the model's preference for the most novel object correlated with ME behavior, that is, assigning the novel label to the most novel object and not to other two objects, we examined WOLVES' newly learned word–object associations. We compared the average number of times the model associated the novel label with the novel object against chance (i.e., 1/3). WOLVES associated the novel label with the novel object significantly above chance ( $M = .54$ ,  $SE = .05$ ,  $p < 0.01$ ) compared to the pre-exposed ( $M = .32$ ,  $SE = .04$ ) and the name-known objects ( $M = .14$ ,  $SE = .03$ ).

Mather and Plunkett argued based on the infant data that ME is driven by novelty: when the novel name was presented, infants looked to the most novel object of the two name-unknown objects available. However, the prior simulations with WOLVES have shown that novelty preferences in the context of words are influenced by the competing action of habituation to more familiar stimuli and top-down attention driven by word–object mappings. Thus, we further examined the basis for infants' novelty preference in the context of this study via additional simulations manipulating the strength of these competing processes.

To analyze the role of top-down word-driven attention from the known words on preferences for the most novel item following label presentation, we ran a set of simulations wherein we systematically varied the strength of the



**FIGURE 9** Left: Novelty preference as strength of known-object memory trace was varied. Right: Effect of habituation parameter on novelty preference in the post-naming phase.

memory trace of the known word (i.e., the word–object association) and examined how this modulated novelty preferences in the task following the word presentation (after 4000 ms). The left panel of Figure 9 shows the results in terms of percentage preference in looking at the novel object compared to the other objects in the latter half of trials, following word presentation. As can be seen, with increasing strength of known-word representations, novelty preference in the model increases. This suggests that prior knowledge boosts the novelty preference following word presentation. This is consistent with other work by Bion et al. (2013), which shows that with increasing age and vocabulary, the ability to disambiguate novel objects improves.

To examine the effect of habituation on infants' attention to the most novel object following label presentation, we ran a set of simulations manipulating the strength of the working memory traces for the scene. This parameter influences how quickly the model habituates. The right panel of Figure 9 shows that the novelty preference consistently increases with stronger working memory traces (i.e., stronger habituation). Stronger habituation means less looking at known and pre-exposed objects that are presented to the model many times during familiarization. Thus, habituation also contributes to the increase in looking at the novel object.

## Discussion

This experiment and the corresponding simulations delve deeper into the origins of ME and tease apart the relative contributions of visual novelty, familiarity, and naming. Both the empirical findings and WOLVES simulations concur on the pivotal role of novelty in the ME bias. The model and children show a robust relative preference for the novel object in the post-naming period of the trials. Since looking drives learning in the model, a ME bias appears wherein the model maps the novel label to the more-looked-at novel object. This mapping is manifested in the word–object memory traces in the model, highlighting the synchrony in visual behavior

and internal knowledge representation. Critically, the model provides a mechanistic view of the ME bias, explaining how interactions between attentional processes and word association processes can drive looking and learning behaviors in this task.

While WOLVES is consistent with Mather and Plunkett's viewpoint that novelty is a primary driver behind ME, the model also suggests that novelty is impacted by multiple factors including the strength of object knowledge and the level of habituation in the system. A stronger memory of known objects improves attentional processing, allowing faster selection and consolidation of known objects. This then allows more time to explore the novel objects. Furthermore, as the strength of the known-object association increases in the experiment, so does the probability that the model will look first at the known object. This builds a working memory for the known object, driving attention away from this object during the latter part of the trial when novel labels are presented. This further increases the chances of ME.

Looking away from the pre-exposed object is largely influenced by the level of habituation in the system. The more the pre-exposed objects are familiarized with the model (by increasing the habituation rate in the system or by increasing pre-exposure time), the shorter the look durations of the pre-exposed objects become during the test trials. Shorter looks to pre-exposed objects result in higher relative preferential looking to the novel object at test. Considering these factors together, we suggest that novelty is not an isolated, necessary, and sufficient condition for ME as Mather and Plunkett (2012) propose; rather, *relative* novelty conspires with other task-specific processes to create the ME bias (see also Mather & Plunkett, 2012, Experiment 2).

## GENERAL DISCUSSION

The goal of this special section is to showcase what mathematical and simulation-based modeling can accomplish for central questions in the study of child development. The present study used a model of early word

learning—WOLVES—to provide a formal, process-based understanding of the different cognitive processes underlying novelty preferences and word learning in early development. We focused on a complex literature that has examined how auditory and visual events are processed during a period in early development when vocabulary knowledge is rapidly expanding. We suggested that process models might be a useful tool to *clarify the complex* in this literature because they take time seriously, that is, they specify the details of input, processing on input, and how that processing leads to behavior. We then used WOLVES as a tool to probe how the different cognitive processes underlying novelty preferences and word learning might impact findings from the studies by Mather et al. In what follows, we reflect on this simulation exercise and what it reveals about the role process models can play in developmental science.

Our first take-home conclusion is that process models like WOLVES have a real strength—generalization over tasks/conditions. We argued that this is because such models take time seriously. Consequently, adapting WOLVES to a new paradigm involved situating the model in the context of the auditory and visual events in the task and measuring the relevant behaviors (e.g., looking left and right). This was largely a programming exercise. We then asked a key question: how do we evaluate the effectiveness of this generalization exercise, that is, how do we know if the model works? Here, we discussed the challenges of parameter fitting and running quantitative simulations, highlighting that the nature of parameter fitting depends on the goals of the modeler (for discussion, see Pitt & Myung, 2002). In our case, our goal was not to fit every detail of the empirical findings. This reflects, in part, the nature of the data. While differences in visual processing in the context of words are well established in the literature, the data from the simulated studies are somewhat variable. Thus, our goal was not to mimic the data in precise detail; rather, we wanted to use the simulations more qualitatively to explore what processes might push infants' performance around. Thus, we kept our parameter explorations to a minimum. Ultimately, we achieved reasonable fits to the data from the first simulation experiment by changing three parameters—two noise strength parameters and *tau\_decay* (which implemented the main hypothesis of less forgetting over development from Bhat et al., 2022). We identified “best-fitting” parameter values using a grid search and subsequently refined these parameters through a probe of generalization to a second experiment (see [Supporting Information](#)). We then showed how these parameter sets captured findings from Mather and Plunkett (2012).

A key question is did this exercise, in fact, help *clarify the complex*? One important way WOLVES provides clarity is by instantiating how processes of attention, working memory, long-term memory, and word-object associative learning create patterns of visual exploration,

habituation, and novelty biases, and by specifying how these evolve over the timescales of a trial, an experiment, and development. The exercise of generalizing WOLVES to studies by Mather et al. clarified some aspects of the results, pointing toward future work that would add additional clarity regarding the origin of some empirical effects. WOLVES also failed to reproduce some empirical findings clarifying ways in which the modeling efforts need enhancement as well.

The experiments we simulated addressed the relationship between familiarity and novelty preferences in different ways. This is a fundamental issue in the developmental literature, with data showing that infants' preference for novel versus familiar objects is impacted by a range of factors such as infants' age, stimulus complexity, and task durations (see, e.g., Hunter & Ames, 1988; Rose et al., 2003). WOLVES sheds light on the possible origin of novelty and familiarity biases within the specific tasks simulated here. In Simulation Study 1, WOLVES showed that an initial novelty bias which was slightly stronger in the silent condition reflects habituation due to the continued presence of a repeated item from trial to trial. In the labeling condition, this process competes with the formation of word-object associations. This competition created a U-shaped pattern over trial blocks in the labeling condition as the word-object associations became stronger, particularly for the older models with slower memory decay. There was mixed statistical support for a U-shaped effect in the data; we suggest that future empirical work might investigate this in more detail. We note that WOLVES was not directly able to explore the explanation offered by Mather et al. (2011) that auditory input slows down visual processing as there is no explicit auditory process in the model. It would be useful for future modeling work to formalize how this explanation might work. Nevertheless, WOLVES suggests that competition between visual habituation and word learning processes may be key in this task. We also showed that more word-object pairs reduce both habituation and word learning (see [Supporting Information](#)). Here, we saw how the task structure matters, highlighting a key point: it is not the processes that changed across studies, but the processing due to the change in task structure.

A key contribution in Simulation Study 2 was the demonstration that novelty may be better conceptualized as *relative* novelty. Simulations showed that the bias toward the novel item was influenced by the strength of the known word-object associations as well as habituation to the pre-familiarized item. This was clarified via additional simulations where we manipulated key parameters, revealing systematic relationships among processes in the model which all contribute to the ME bias. This is a good example of how models can be useful, allowing us to directly manipulate factors that push behavior around.

We note that the simulations reported here could have been conducted differently to serve other goals.

For instance, the central goal of Bhat et al. (2022) was to develop a comprehensive theory of cross-situational word learning that explained key findings from the literature, out-performed competitor models, and provided the first account of how cross-situational word learning changes over development. Bhat et al. (2022) reviewed 19 models of cross-situational word learning, highlighting that most word learning models are one-shot—they do not unpack what happens in real time on each trial. Thus, the pattern of looking from second-to-second through time cannot be captured. Bhat et al. (2022) then compared WOLVES with two competitor models, simulating 132 data values from 12 experiments with children and adults. This included generalization to three experiments without any fitting of parameters in advance (similar to “hold-one-out” validation). Quantitative measures of fit (e.g., Akaike information criterion) were then compared across all experiments. Critically, WOLVES captured the data better than the competitor models in the generalization experiments, explained data that the other models failed to capture, and offered the only developmental account. In the context of the present report, we use this example to highlight that there is not one single way to run model simulations. Rather, the approach to modeling should vary with the goals.

Indeed, reflecting on the different simulation approaches across studies, we see important contributions from both more quantitative approaches (e.g., Bhat et al., 2022) and more qualitative approaches (i.e., the present report). For instance, the qualitative model usage reported here offered potential new insights into the data, even though the data were somewhat noisy and the model failed to capture some significant findings. We perceive that models are more often used to capture data from literatures that are more “settled,” where studies have been replicated multiple times (often including pre-registration). While this is certainly a fundamental use of models, we envision a future where models and data can have more of a dialog back and forth; where models can be readily applied to multiple data sets and the model results are used to inform the next experiments even if model interpretations are somewhat preliminary, and even when the model does not achieve an ideal quantitative fit.

For instance, Bhat et al. (2022) made several novel predictions within the cross-situational word learning domain that are currently being tested. What about building on the work presented here? Simulations of experiments from Mather et al. (2011) show that the recurrence of the same novel word–object pair across trials helps infants acquire the repeated mapping, modulating novelty biases relative to the silent condition. One way to mitigate the impact of word–object associations would be to modify the task by providing a novel label on each trial. If labels are not repeated, the system will form only weak word–object associations, eliminating top-down influences on the visual dynamics.

We probed this prediction in WOLVES by simulating the same task structure as Simulation Study 1 (Experiment 1 from Mather et al., 2011) but included a different novel label on each trial. In this new label novel condition, WOLVES showed nearly identical behavior to the silent condition from the original study (see Figure S1). Because the labels change on every trial, however, there is no consistent word–object association as in the labeling condition of the original study. This means there is no top-down word-driven attention in the model. Thus, looking in the model is dominated by visual memory and habituation factors only as is the case of the silent condition. This highlights how WOLVES—and process models more generally—can be a useful platform for exploring novel experimental situations and generating a priori novel predictions.

Although several aspects of our simulation study were promising, the study also revealed limitations. As noted previously, one major limitation is parameter tuning with complex models like WOLVES. We approached this by using a three-parameter grid search; however, the fit of the model to data was clearly not optimal, with the model failing to capture several statistically significant effects in the data. We are currently developing a new tool called “Dynamic Field Flow” that uses machine learning tools to optimally fit parameters of DF models. This requires that one specify constraints for the activation time course that must be satisfied to optimally fit the model to data. We have shown that this works well for reaction time data but have yet to generalize this approach to fitting accuracy data or looking time courses. Another key limitation of the study was the variability of the data we modeled; as noted, replication of the findings from Mather et al. (2011) and Mather and Plunkett (2012) is necessary to identify which patterns are truly robust in the data. On this front, it would be ideal to have more information on the within-subjects variance in the data to move simulation work toward modeling individuals instead of groups. This would enable us to model not just group means but also the variance in the data, and to test the fit of the model statistically. Finally, we did not examine competing models of the phenomena of interest. Such work is critical moving forward to probe which explanations best explain the processes driving the central effects reported here.

In conclusion, simulations from the present study show how formal process models can clarify complexity by making theoretical concepts and processes explicit and concrete. Looking forward, we resonate with goals of the special section which also highlights how models can contribute to the grounding of empirical findings within an open science framework. In line with this, we again point to our OSF site where all materials from the present report are available. Readers might also be interested in materials on our DFT website, a central repository for the theoretical framework, with case studies, tutorial videos, and other tools including

model code and instructions ([www.dynamicfieldtheory.org](http://www.dynamicfieldtheory.org)). We hope these efforts can contribute to a richer theory–experiment dialog in developmental psychology and related fields.

### FUNDING INFORMATION

This research was supported by HD045713 awarded to Larissa K. Samuelson. The content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH.

### DATA AVAILABILITY STATEMENT

All materials necessary to reproduce the model and simulations are available at <https://doi.org/10.17605/OSF.IO/NG2VF>. The authors wish to thank Emily Mather for supplying data from her studies that we simulated for inclusion in the online repository associated with this paper. The simulations presented in this paper were carried out on the high-performance computing cluster supported by the Research and Specialist Computing Support Service at the University of East Anglia.

### ORCID

Ajaz A. Bhat  <https://orcid.org/0000-0002-6992-8224>

Larissa K. Samuelson  <https://orcid.org/0000-0002-9141-3286>

### REFERENCES

- Aslin, R. N. (2007). What's in a look? *Developmental Science*, 10(1), 48–53. <https://doi.org/10.1111/J.1467-7687.2007.00563.X>
- Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2022). Word-Object Learning via Visual Exploration in Space (WOLVES): A neural process model of cross-situational word learning. *Psychological Review*, 129, 640–695. <https://doi.org/10.1037/REV0000313>
- Bion, R. A. H. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, 126(1), 39–53. <https://doi.org/10.1016/j.cognition.2012.08.008>
- Ferguson, B., & Waxman, S. (2017). Linking language and categorization in infancy. *Journal of Child Language*, 44(3), 527–552. <https://doi.org/10.1017/S0305000916000568>
- Gogate, L., & Maganti, M. (2016). The dynamics of infant attention: Implications for crossmodal perception and word-mapping research. *Child Development*, 87(2), 345–364. <https://doi.org/10.1111/CDEV.12509>
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87, 23–34. [https://doi.org/10.1016/S0010-0277\(02\)00186-5](https://doi.org/10.1016/S0010-0277(02)00186-5)
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128–157. <https://doi.org/10.1080/15250000701795598>
- Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. A. (2011). What's new?: Children prefer novelty in referent selection. *Cognition*, 118(2), 234–244. <https://doi.org/10.1016/j.cognition.2010.10.015>
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, 5, 69–95.
- Jadva, V., Hines, M., & Golombok, S. (2010). Infants' preferences for toys, colors, and shapes: Sex differences and similarities. *Archives of Sexual Behavior*, 39(6), 1261–1273. <https://doi.org/10.1007/S10508-010-9618-Z/TABLES/5>
- Kucker, S. C., McMurray, B., & Samuelson, L. K. (2018). Too much of a good thing: How novelty biases and vocabulary influence known and novel referent selection in 18-month-old children and associative learning models. *Cognitive Science*, 42, 463–493. <https://doi.org/10.1111/cogs.12610>
- Kucker, S. C., McMurray, B., & Samuelson, L. K. (2020). Sometimes it is better to know less: How known words influence referent selection and retention in 18- to 24-month-old children. *Journal of Experimental Child Psychology*, 189, 104705. <https://doi.org/10.1016/j.jecp.2019.104705>
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. MIT Press.
- Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14(1), 57–77. [https://doi.org/10.1207/s15516709cog1401\\_4](https://doi.org/10.1207/s15516709cog1401_4)
- 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 Computational Biology*, 6(9), e1000919. <https://doi.org/10.1371/journal.pcbi.1000919>
- Mather, E. (2013). Bootstrapping the early lexicon: How do children use old knowledge to create new meanings? *Frontiers in Psychology*, 4, 1–3. <https://doi.org/10.3389/fpsyg.2013.00096>
- Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, 14(1), 60–76. <https://doi.org/10.1080/15250000802569702>
- Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science*, 36(7), 1157–1177. <https://doi.org/10.1111/j.1551-6709.2012.01239.x>
- Mather, E., Schafer, G., & Houston-Price, C. (2011). The impact of novel labels on visual processing during infancy. *British Journal of Developmental Psychology*, 29, 783–805. <https://doi.org/10.1348/2044-835x.002008>
- Perone, S., & Spencer, J. P. (2012). Autonomy in action: Linking the act of looking to memory formation in infancy via dynamic neural fields. *Cognitive Science*, 37, 1–60. <https://doi.org/10.1111/cogs.12010>
- Perone, S., & Spencer, J. P. (2013). Autonomous visual exploration creates developmental change in familiarity and novelty seeking behaviors. *Frontiers in Psychology*, 4, 648. <https://doi.org/10.3389/fpsyg.2013.00648>
- Perszyk, D. R., & Waxman, S. R. (2018). Linking language and cognition in infancy. *Annual Review of Psychology*, 69, 231–250. <https://doi.org/10.1146/ANNUREV-PSYCH-122216-011701>
- Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, 6(10), 421–425. [https://doi.org/10.1016/S1364-6613\(02\)01964-2](https://doi.org/10.1016/S1364-6613(02)01964-2)
- Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106(2), 665–681. <https://doi.org/10.1016/j.cognition.2007.04.003>
- Richards, J. E. (2010). The development of attention to simple and complex visual stimuli in infants: Behavioral and psychophysiological measures. *Developmental Review*, 30(2), 203–219. <https://doi.org/10.1016/J.DR.2010.03.005>
- Robinson, C. W., & Sloutsky, V. M. (2007a). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy*, 11(3), 233–253. <https://doi.org/10.1111/j.1532-7078.2007.tb00225.x>
- Robinson, C. W., & Sloutsky, V. M. (2007b). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science*, 10(6), 734–740. <https://doi.org/10.1111/J.1467-7687.2007.00627.X>
- Roder, B. J., Bushneil, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy*, 1(4), 491–507. [https://doi.org/10.1207/S15327078IN0104\\_9](https://doi.org/10.1207/S15327078IN0104_9)
- Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2003). Infant visual recognition memory: Independent contributions of speed and attention. *Developmental Psychology*, 39(3), 563–571. <https://doi.org/10.1037/0012-1649.39.3.563>

- Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology, 18*(5), 704–713. <https://doi.org/10.1037/0012-1649.18.5.704>
- Rose, S. A., Jankowski, J. J., & Feldman, J. F. (2002). Speed of processing and face recognition at 7 and 12 months. *Infancy, 3*(4), 435–455. [https://doi.org/10.1207/S15327078IN0304\\_02](https://doi.org/10.1207/S15327078IN0304_02)
- RSCS-UEA. (2015). *High performance compute cluster*. <https://www.uea.ac.uk/groups-and-centres/research-and-specialist-computing/high-performance-computing>
- Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PLoS One, 6*(12), e28095. <https://doi.org/10.1371/journal.pone.0028095>
- Schneegans, S. (2012). *COSIVINA—Compose, simulate, and visualize neurodynamic architectures overview*. <https://dynamicfieldtheory.org/cosivina/>
- Schneegans, S., Lins, J., & Spencer, J. P. (2015). Integration and selection in multidimensional dynamic fields. In G. Schöner, J. P. Spencer, & The DFT Research Group (Eds.), *Dynamic thinking* (pp. 121–150). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199300563.003.0005>
- Schöner, G., Spencer, J. P., & The DFT Research Group. (2016). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press.
- 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), 7–127.
- Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. *Cognitive Science, 32*(2), 342–365. <https://doi.org/10.1080/03640210701863495>
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*(3), 1558–1568. <https://doi.org/10.1016/j.cognition.2007.06.010>
- Spencer, J. P. (2020). The development of working memory. *Current Directions in Psychological Science, 29*, 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*, 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.
- Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology, 30*(4), 553–566. <https://doi.org/10.1037//0012-1649.30.4.553>
- Zupan, B., & Sussman, J. E. (2009). Auditory preferences of young children with and without hearing loss for meaningful auditory–visual compound stimuli. *Journal of Communication Disorders, 42*(6), 381–396. <https://doi.org/10.1016/J.JCOMDIS.2009.04.002>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** 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, 00*, 1–20. <https://doi.org/10.1111/cdev.14023>