Similarity in object properties supports cross-situational word learning: Predictions from a dynamic neural model confirmed

Ajaz A. Bhat (ajaz.bhat@ubd.edu.bn)
School of Digital Science, Jalan Tungku Link,
Universiti Brunei Darussalam, BE 1410, Brunei

John Spencer (j.spencer@uea.ac.uk)
School of Psychology, 0.9A Lawrence Stenhouse Building
University of East Anglia, NR4 7TJ, UK

Larissa K. Samuelson (l.samuelson@uea.ac.uk)
School of Psychology, 0.113B Lawrence Stenhouse Building
University of East Anglia, NR4 7TJ, UK

Abstract

Learning names for novel objects has been shown to be impacted by the context in which they appear. Manipulations of context, therefore, provide a key pathway to explore these learning dynamics. Here we use a neural process model that instantiates the details of 'context' to generate novel, counterintuitive predictions about how similarity in object properties influence learning. Specifically, we use a dynamic field model, WOLVES, to simulate and predict learning in a cross-situational word learning task in two conditions: one where the two objects presented on each learning trial are metrically similar in a property ('NEAR') and another condition where the two objects are always dissimilar ('FAR'). WOLVES predicts—counterintuitively—that participants should learn better in the 'NEAR' condition (where objects are potentially confusable) than in 'FAR' condition (where objects are distinctive). We then tested this prediction empirically, finding support for the novel prediction. This study shows the utility of process models which instantiate the details of 'context' during learning and provides support for WOLVES. We know of no other theory of cross-situational word learning that captures these novel findings.

Keywords: cross-situational word learning; dynamic neural model; DFT; metric similarity; attention and memory; learning

Introduction

A central issue in cognitive science is how learning is affected by the context in which it occurs. One area in which this has been demonstrated recently is Cross-Situational Word Learning (CSWL). CSWL refers to tasks in which word object mappings are learned via accumulation of information gained over successive, individually ambiguous, trials. In a classic demonstration of CSWL with children, Smith and Yu (2008) presented infants with 30 trials composed of two objects and two words. On each trial, it was unclear which word mapped to which object, but every time a given object appeared, its associated word was heard. Thus, over trials the correct word-object mappings could be inferred. In a preferential looking test, two objects were again presented, but with only one word. Infants looked more to the object that had most-often occurred with that word for 4 of the 6 tested words, suggesting learning.

Subsequent studies have confirmed CSWL in adults and children under a variety of task manipulations designed to examine the limits of and possible mechanisms supporting learning (see Roembke, Cimonetti & Koch, 2023). An active literature presents an ongoing debate about the mechanisms supporting learning and includes multiple computational models examining the mechanisms and processes involved (see Bhat, Spencer & Samuelson, 2022 for review). Manipulations of various aspects of the learning context, such as the number of other objects presented on a trial, have contributed to the understanding of CSWL.

Suanda and Namy (2012) assigned objects to trials randomly such that some objects occurred together more often than with other objects. They found that the these spurious correlations during training reduced learning. Kachergis, Shiffrin and Yu (2009) demonstrated that increased contextual diversity (e.g., appearing with more items across trials) supported learning of pairs presented less frequently over trials. These studies suggest learners encode aspects of the context such as what other objects were present during individual CSWL trials.

Other studies have examined the influence of similarity between the presented words. Mulak, Vlach and Escudero (2019) found that using pairs of words that differed by one vowel or one consonant reduced learning accuracy overall, compared to other studies using more dissimilar words. Tuninetti, Mulak and Escudero (2020) further showed that the similarity of novel auditory stimuli to participants' native language impaied learning. These studies demonstrate the influence of auditory context on CSWL and additionally that the similarity of stimuli to known words can reduce learning.

The Current Study

The current work sheds light on how multiple cogntive processes come together to support word learning using a neural process model of CSWL. We test a unique *a priori*

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

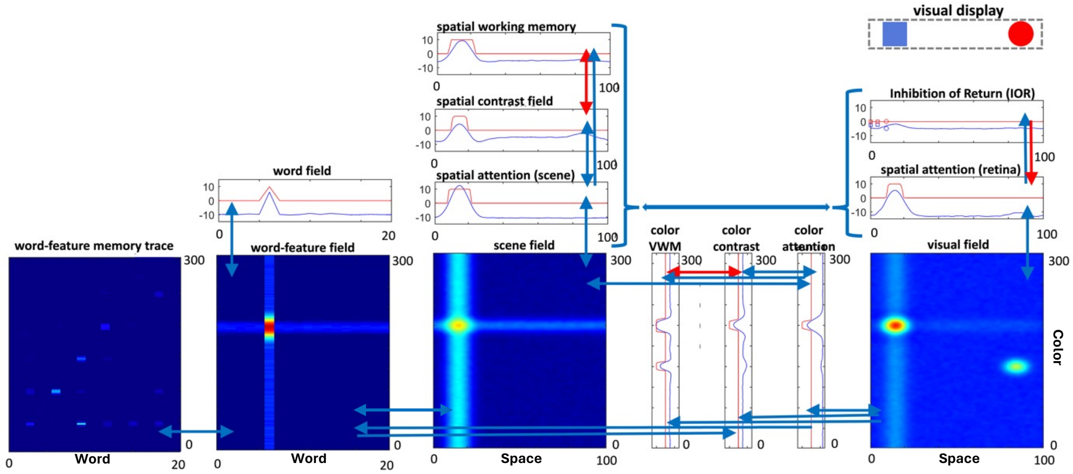

Figure 1. The WOLVES model neural architecture. The one-dimensional (1D) and two-dimensional (2D) dynamic fields (DFs) in the model are responding to the visual display in the top right. Arrows represent uni/bidirectional connectivity (blue: excitatory, red: inhibitory). Only one feature dimension, and some working memory and memory trace fields are not shown for simplicity. 1D fields show activation profile in blue and above-threshold neural activity in red. 2D fields visual and scene fields show activation representing specific colors in specific spatial locations, with higher activation in 'hotter' colors. The 2D word-feature and word-feature memory trace field show binding of specific words and the object color feature.

prediction from a recent model of CSWL to both gain insight on the role of local context in word learning and as a probe of this specific model. In contrast to prior CSWL studies that have manipulated local context in terms of the co-occurrence rates of objects, we examine how the similarity of the objects presented on a trial influences learning, akin to work using minimal auditory pairs. This is possible because the process model we use—Bhat et al.'s (2022) Word-Object Learning via Visual Exploration in Space (WOLVES)—is unique amongst models in its ability to represent similarity relations between objects and their neural interactions. Investigating systematic variations in object properties in CSWL is crucial for better understanding of the cognitive processes underlying language acquisition.

WOLVES is a neural model implemented within Dynamic Field Theory (DFT) (Schöner, Spencer & The DFT Research Group, 2015; for a comprehensive model description, refer to Bhat, et al, 2022). Before describing the simulations, prediction, and empirical test, we provide a concise overview of DFT and a description of WOLVES.

Dynamic Field Theory

Dynamic Field Theory (DFT) posits that cognitive processes emerge from neural processes occurring within dynamic neural fields (DNFs), which simulate the dynamics of neural populations. These fields are organised based on cortical tuning curves into metric dimensions such as color, shape, and space. Neurons within a field interact recurrently with each other according to a local excitation, lateral inhibition function leading to the formation of peaks that serve as stable states of the field (Amari, 1977).

A DNF architecture comprises fields (see Figure 1) which interact along shared dimensions through unidirectional or bidirectional projections (shown as arrows in Figure 1). Fields can also incorporate a form of Hebbian learning at a slower timescale, allowing neural populations to learn and encode statistical information across trials. This 'memory trace' (the leftmost 2D field in Figure 1) creates a local boost for repeatedly active sites , enhancing the likelihood of peak formation in those regions.

WOLVES: The Model

A visual representation of WOLVES is presented in Figure 1. For simplicity this figure only shows fields processing objects' color; objects' shape is processed in a similar manner by adding another row of fields tuned the shape features. Each box, excluding the visual display in the upper right, shows a dynamic field. Starting with the fields on the right, the model captures processing related to visual exploration in space (Perone & Spencer, 2013; Schneegans, Spencer & Schöner, 2016). Visual inputs (e.g., the red circle and blue square in Figure 1) are presented to a visual field (far right in Figure 1), encoding features such as color and shape, along with their respective locations on the visual field. This information is processed along two pathways: a dorsal pathway for spatial information (horizontal 1D fields in Figure 1) and a ventral pathway for feature information (vertical 1D fields in Figure 1). The dorsal pathway represents the location of objects in the world, creating a spatial working memory for crucial locations within the current scene. The ventral pathway represents object features (like color), including a visual working memory for specific features. The scene attention field combines this information into a scene representation, discerning 'what is where' (Treisman & Gelade, 1980).

The word-object learning portion of the model (Samuelson, Smith, Perry & Spencer, 2011) is on the left of Figure 1.

Auditory inputs are presented to the word field, representing the provided word labels. Words and object features are then integrated in the word-feature (binding) field. Over time, memory traces grow in the memory trace layer, supporting the recognition of word-object mappings and guiding attention to familiar items through top-down connectivity.

Figure 1 illustrates neural activation in WOLVES at a moment when the model recalls a word-object mapping and directs attention to the corresponding object in the visual field. Understanding the model involves considering 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 observes visual input, generating activation peaks in the visual field. The color and spatial features are then conveyed to the contrast fields, which detect visual novelty. Peaks in the visual working memory fields can suppress contrast layer peaks, distinguishing between 'known' and novel elements. Peaks in the contrast fields transmit activation to the attention fields. These 'winner-take-all' fields support a single peak—the focus of attention. The attention peak amplifies activation in the visual field, selecting the focused element and enhancing spatial attention. After consolidation in working memory field, the information enters the 2D scene field, binding feature and spatial inputs. The inhibition of return field detects the consolidated item, suppressing spatial attention and releasing the item from the attentioanl focus. The cycle repeats as the model explores the next item.

The word-object learning cycle starts when an auditory word is presented, creating a peak in the word field. This peak influences the word-feature field, where attended visual features are also projected. If a word is presented while an object is attended, the combined inputs build a peak in the word-feature field, binding the word to the attended visual features and leaving a trace in the memory trace layer. These traces grow over subsequent presentations allowing the model to learn word-feature mappings.

The top-down attention cycle starts when word input intersects a strong memory trace for a word-feature mapping. This causes a peak in the word-feature fields, passing a 'top-down' signal to the contrast field. This signal activates the associated feature, directing attention to the object via interactions with the feature attention fields.

These cycles emerge over various timescales as neural activation propagates in the model. On a real timescale of milliseconds and seconds, the model autonomously shifts attention between visible objects and recognizes words. Visual habituation emerges as strong memory traces alter the visual exploration in space cycle, causing the model to swiftly release fixation from 'known' items and spend more time exploring novel items. Learning occurs over a longer timescale, as repeated attention to objects in the presence of words builds stronger memory traces. As learning progresses, it will affect subsequent processing. The word-object learning cycle can help the model build correct word-object associations because robust word-object traces can block the formation of new incorrect associations. Learning also impacts the top-down attention cycle, as strong memory traces direct attention to labelled objects.

Together these processes allow WOLVES to *autonomously* allocate attention to the objects in the visual field on the real timescale of milliseconds and seconds and build associations over trials that grow to allow recognition of words. This real-time behaviour aligns with participants' looking behaviour, enabling the model to be embedded in the exact same experiments as participants using identical visual and auditory inputs as those in CSWL tasks.

CSWL with WOLVES

Since WOLVES autonomously generates looking data, adaption to preferential-looking CSWL tasks such as Smith & Yu (2008) is direct. Each object presented to the model is represented by a specific color-shape feature pair. To present an object to the model we input a gaussian-shaped stimuli pair at the corresponding feature values of the model's visual field. Words were locally represented with distinct Dirac functions randomly located across the word dimension. Following presentation of inputs, activation evolves autonomously during each training and testing trial.

As in typical CSWL studies, two objects are presented to the model on each training trial along with two words. The duration of trials and stimuli onset and offset are made to exactly match the target experiment. For example, in simulations of Smith & Yu, 2008 (Bhat et al., 2022) training trials were 4000ms long with the first word turned on at 500ms and off at 1500ms and the second turned on at 3500ms and off 1000ms later. Learning in WOLVES can be assessed in two ways. Because the model autonomously explores objects during training and test, behavioral measures such as time looking to the target can be taken, just as in studies with child and adult participants. In addition, the model's internal processing can be anlayzed, examining measures such as the strength of memory traces in word-feature fields as a measure of the correct/incorrect associations learned by the model.

Bhat et al. (2022) used WOLVES to capture data from 7 adult and 5 child studies of CSWL, capturing more data than two competitor models with good quantitative fits. WOLVES also generalized better to three "held-out" experiments. The developmental account of CSWL instantiated in WOLVES, the only such account to date, demonstrates how changes in memory processes from infancy to adulthood influence task performance. Further WOLVES sheds light on the processes supporting CSWL, showing how visual exploration and selective attention in CSWL are dependent on and also indicative of learning and how learning is driven by the real-time synchrony of words and gaze and dynamically constrained by memory processes. Here we seek additional insight on how relations between stimuli influence the learning process by simulating conditions of high and low similarity between objects to make an a priori prediction that is then tested empirically.

Simulation and Experiment Design[1]

A unique feature of DNF models and thus WOLVES is the use of metrically organised dimensions to represent object stimuli and experimental features such as the space in which objects are presented. Practically, this means that the representations of a red circle and pink oval presented simultaneously to WOLVES would be closer together on the color and shape dimensions than a red circle and a green triangle. Further, the interaction kernel governing how sites in WOLVES's feature fields interact is such that sites close to each other in color and shape have excitatory interactions while those farther apart have inhibitory interactions.

Combined, these aspects of WOLVES mean that we expect the relative similarity of the two objects presented on a CSWL trial will matter for the representations the model develops over the course of a trial and for learning outcome. Indeed, such effects have previously been documented in DNF models. Johnson, Spencer, Luck & Schöner (2009) used a change detection task to look at the effect of stimulus similarity on working memory. They found that when participants were asked to consolidate similar colors or orientations in working memory, they were better able to discriminate a feature change presented at test, compared to when stimuli were far apart in color or orientation space.

These prior findings lead to the counterintuitive prediction that learning in CSWL may be improved if similar object pairs are presented together during training. To test this prediction, we designed a CSWL stimulus set that allowed precise control of the relative similarity of the stimuli.

Object Stimuli

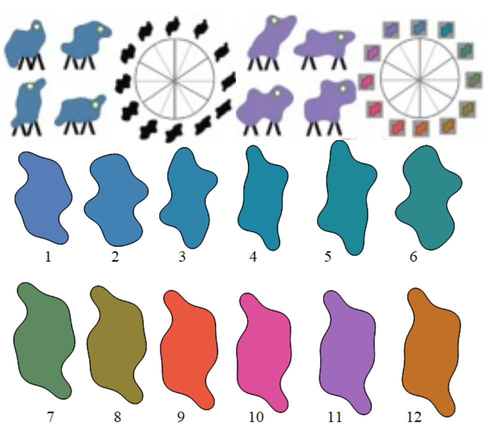

Figure 2. Top: Beasties and 360° color and shape spaces. Bottom: Stimuli objects used. Note that the objects in the top row are similar in color but have dissimilar shapes and vice versa for the objects in the bottom row.

We created a set of novel stimuli by combining metric variations of object shapes and colors. Object shapes were defined by radial frequency components (Zahn & Roskies, 1972, Figure 2 top left), providing an evenly parameterized similarity space without category boundaries. Object colours were sampled equidistantly from a 360° continuous colour space (CIELab, 1976). Twelve Beastie stimuli were created by first selecting two mutually exclusive sets of equidistant feature values. In Set 1 feature values were forty degrees apart: 105, 145, 185, 225, 265, 305, 345. In Set 2 features were 10 degrees apart: 20, 30, 40, 50, 60, 70. A 35-degree gap separated stimuli in the two sets (e.g., from 70 to 105; and from 345 to 20). These values were then combined as shown in Table 1 (column 'Feature value-pair') to create the twelve Beasties (Figure 2, bottom). We added legs and eyes to highlight shape differences and help maintain participant interest. The resulting "Beasties" were generated in Matlab and presented a on uniform grey background.

Object	Name	Feature value-pair
1	Manu	20-345
2	Bosa	30-265
3	Gake	40-185
4	Regli	50-105
5	Colat	60-145
6	Kaki	70-225
7	Virdex	105-20
8	Loche	145-30
9	Tanzer	225-40
10	Boscot	305-50
11	Fiffin	345-60
12	Chatten	185-70

Table 1. Word-referent pairings and corresponding object feature value pairings for stimuli used in the study.

Word Stimuli

Twelve disyllabic nonwords consistent with the phonological probabilities of the English Language were recorded in isolation by a female British speaker (Table 1). Each sound was 1000 ms long in duration.

Design

Both the empirical study and simulations used an independent measures design, with different groups (of people/simulations) in two learning conditions, NEAR and FAR. Based on Smith & Yu (2008), each condition involved two phases, learning and test. The learning phase consisted of 30 trials each presenting two objects for 4000 ms. During this time, two words were played through the speaker for participants or turned on for the model with the timing matching that of Smith and Yu (2008). A test phase of 12

[1] The design and qualitative predictions were pre-registered on OSF: https://osf.io/d9zeq/

trials followed, wherein participants/simulations were exposed to two objects and a single word on each trial, for a duration of 8000 ms each. One object, the target, had always been presented with the tested word during learning. The other object served as a distractor.

The same stimul were used for both conditions, but paired differently (Table 2). In the NEAR condition, paired objects were just 10 degrees apart on one of the dimentions. In the FAR condition, objects were were 35 degrees apart on both dimensions. See Figure 3 for example trials.

The 10 object pairings were repeated three times to make up 30 training trials. Trial order, right/left position of objects, and order of first/second word presentation was randomized separately in each block. Note that the design also leads to equal contextual uncertainty (number of co-occurring objects) for each object in both conditions.

The same test trials were used for the NEAR and FAR conditions, one for each object/word. The selection of distractors and their side/location in the test trials was randomized as was the order of the test trial presentations.

NEAR condition (similar on one dimension)	**FAR condition** (different in both dimensions)
1 vs 2	1 vs 7
2 vs 3	2 vs 8
3 vs 4	3 vs 9
4 vs 5	4 vs 10
5 vs 6	5 vs 11
7 vs 8	2 vs 11
8 vs 9	3 vs 12
9 vs 10	4 vs 7
10 vs 11	5 vs 8
11 vs 12	6 vs 9

Table 2: Object pairings in training trials

Simulation Study

We situated WOLVES in the exact experiment and examined its preferential looking to the target at test in each condition. Based on prior findings of Johnson et al. (2009), we expected the local excitation/lateral inhibition function in the feature fields would cause stimuli in the NEAR condition, which are highly similar on one dimension, to interact in an excitatory fashion. This should affect working-memory formation, change looking dynamics, and thereby modify learning. Critically, because the same objects are used in both conditions; it is just the NEAR/FAR pairing that creates metric differences in the moment.

A visual inspection of the looking/learning dynamics in WOLVES revealed that in the (FAR) condition, as working memory traces of objects bult over training, the model became more efficient at consolidating objects and switching attention. This resulted in quicker release of fixations and, consequently, habituation—meaning less time spent looking

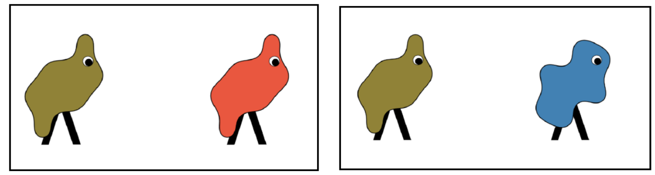

Figure 3. Object 8 v. Object 9 in the NEAR condition (left, similar shape, different colour); Object 8 v. Object 2 in the FAR condition (right, dissimilar in both shape and colour).

at the objects over time. This habituation reduced looking and contributed to a decline in the opportunity to learn correct word-object associations (see also, Smith & Yu, 2013). In contrast, in the NEAR condition, the overlap of similar stimuli peaks created a broad WM trace, allowing a WM peak to persist even after the model shifted attention to the other object. This lingering scene-WM activity hindered the swift formation of a new scene representation. This extended the time needed to consolidate new objects, leading to prolonged fixation and longer looks at the NEAR objects.

Simulation Methods

Simulations were conducted in MATLAB 2016b via the COSIVINA framework, a modeling package for designing DF models (Schneegans, 2012; Schöner et al., 2015). Two machines both using Intel i5 processors were used to run all the simulations: a PC with 36 parallel processing cores and a High- Performance Cluster with 28 parallel processing cores. Time was scaled such that each simulation step equaled eight real-time milliseconds. Simulation results were aggregated over 300 runs (i.e., 300 individuals). Model parameters matched those used by Bhat et al. (2022) to capture adult data.

Results

The left panel of Figure 4 presents WOLVES' preferential looking to the target for the two conditions out of the total trial length (scaled to 8s). As can be seen, the model's looking to the target on test trials was higher when trained using stimuli pairs that were highly similar (NEAR condition;

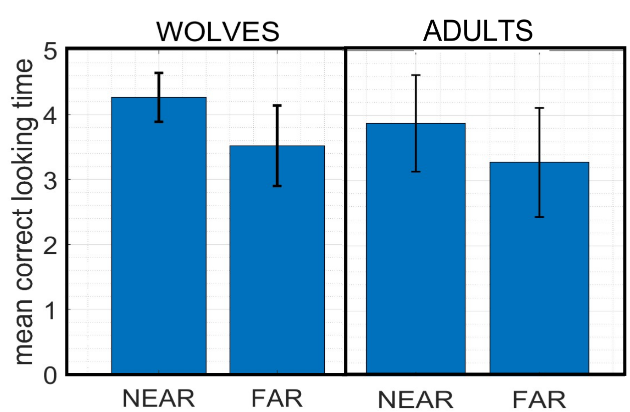

Figure 4. Total looking to target in WOLVES model (left panel) and adult participants (right panel) at test for the two conditions with standard deviation bars.

M=4.26s, SD=0.38s) than when trained with stimuli pairs that were more dissimilar (FAR condition; M=3.52s, SD=0.62s).

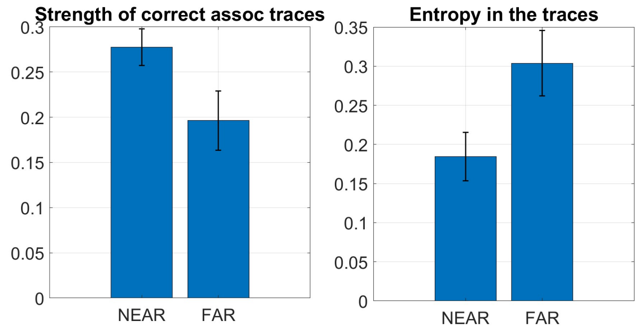

Figure 5. Strength of the memory traces for ‘correct’ associations in the word-feature field (left) and entropy in memory traces associations in the word-feature field (right).

To further assess learning, we measured the strength of correct associations learned in both the conditions. As shown in Figure 5 (left), the memory traces of correct associations are stronger in the NEAR condition than in the FAR condition, confirming that the model learned word-object mappings better in the NEAR condition. We also looked at the overall entropy in the memory traces of the association matrix. Higher entropy is an estimate of more incorrect and random associations. Figure 5 (right) plots the entropy in word-object association memory traces in the two conditions. Entropy values are higher in FAR condition suggesting formation of a larger proportion of random/incorrect associations between words and features.

Thus, the model predicts better learning when similar objects are presented together (NEAR condition) than when dissimilar objects are presented together (FAR condition).

Empirical Study

The simulation data from WOLVES make the counterintuitive prediction that when presented with highly similar stimuli in a two-item CSWL task, adults will learn better than when stimuli are less similar. To test this prediction, we ran the same experimental design in two groups of adult participants.

Methods

Thirty adult participants (23 Females) with a mean age of 20 (sd = 5.1) were randomly assigned to the NEAR and FAR conditions (15 each). Color-blindness was used as an exclusion criterion. An eyelink 1000 eyetracker was used to measure looking behavior during training and test.

Results

Dwell times to the target and distractor on each testing trial were compiled from the raw eyetracking data and averaged across participants within each condition. As can be seen in the right panel of Figure 6, participants proportion looking to the target out of total looking on a trial was higher in the NEAR (M = 3.88s, SD = 0.74s) compared to the FAR (M = 3.28s, SD = 0.84s condition; $t(28) = 2.1$, $p = 0.045$. The mean distractor looking times were also significantly different for the NEAR (M = 2.74s, SD = 0.91s) and FAR (M = 3.38s, SD = 0.65s) conditions; $t(28) = -2.21$, $p = 0.035$. These findings support WOLVES’ prediction of better learning in the NEAR condition than the FAR condition. Target looking in the simultations closely fits empirical means with RMSE=0.32 and MAPE=8.62.

Discussion

This work sheds light on the representations and processes supporting CSWL by examining a neural model’s predictions about the effects of stimulus similarity on learning. Bhat et al.’s (2022) WOLVES model has previously been used to capture a range of CSWL data. Here we capatalized on a unique feature of WOLVES—metrically organized object feature dimensions—to examine the influence of object similarity in learning. Prior work with similar DF models suggested that the local excitation/laterial inhibition function that governs how object representations interact would, counterintuitively, lead to better learning with highly similar stimuli. Indeed, model simulations using a metrically-controlled set of stimuli led to the prediction that adults would learn better when exposed to similar pairs on each learning trial, compared to training that used the same stimuli but presented dissimilar pairs on each trial. An empirical study confirmed this prediction.

This finding is particularly interesting in light of work suggesting that highly similar auditory stimuli make CSWL harder. Malek et al. (2019) found that word pairs that only differed in one vowel or one consonant were harder to learn in CSWL. Future work that directly compares the impact of auditory versus object stimulus similarity on learning will be critical to unpack the bases for the apparent differences between the current work and that of Malek et al. However, one current limitation of WOVES, is that words are represented locally and without metric features.

WOLVES is currently the only model to capture developmental changes in CSWL. Thus it will be important to examine wither WOLVES predictions hold in studies of infants and children. In addition, examination of the interplay of similarity and object presentation order (c.f., Carvalho & Goldstone, 2015) would be revealing of the interaction of the multiple timescales on which representations build in WOLVES. Finally, future analyses will examine changes in looking, including patterns of habituation and novelty detection, over the course of learning. These future directions notwithstanding, the current study represents a strong test of WOLVES that complements recent work generalizng the model to new tasks (Bhat et al., 2023).

Acknowledgments

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 National Institutes of Health (NIH). We wish to thank Ruejan Mamorno and Martyna HaVan for their help with data collection. Simulations presented in this article were carried out on the High-Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia.

References

Amari, S. I. (1977). Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields. *Biological Cybernetics*, *27*(2), 77–87. https://doi.org/10.1007/BF00337259

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

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2022). Word-Object Learning via Visual Exploration in Space (WOLVES): A Neural Process Account of Cross-Situational Word Learning. *Psychological Review*, *129*(4). https://doi.org/https://doi.org/10.1037/rev0000313

Carvalho, P. F., & Goldstone, R. L. (2015). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in psychology*, *6*, 130375.

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science*, *20*(5), 568–577. https://doi.org/10.1111/j.1467-9280.2009.02329.x

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009). Frequency and contextual diversity effects in cross-situational word learning. *Proceedings of Cognitive Science*.

Mulak, K. E., Vlach, H. A., & Escudero, P. (2019). Cross-Situational Learning of Phonologically Overlapping Words Across Degrees of Ambiguity. *Cognitive Science*, *43*(5). https://doi.org/10.1111/cogs.12731

Perone, S., & Spencer, J. P. (2013). Autonomy in action: linking the act of looking to memory formation in infancy via dynamic neural fields. *Cognitive Science*, *37*(1), 1–60. https://doi.org/10.1111/cogs.12010

Roembke, T. C., Simonetti, M. E., Koch, I., & Philipp, A. M. (2023). What have we learned from 15 years of research on cross-situational word learning? A focused review. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1175272

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., Spencer, J. P., & Schöner, G. (2016). Integrating "what" and "where": Visual working memory for objects in a scene. In G. Schöner, J. P. Spencer, & & the D. F. T. R. Group (Eds.), *Dynamic Thinking-A Primer on Dynamic Field Theory.* (pp. 197–226). Oxford University Press.

Schöner, G., Spencer, J. P., & Group, T. D. F. T. R. (2015). *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford University Press.

Smith, L. B., & 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

Smith, L. B., & Yu, C. (2013). Visual Attention Is Not Enough: Individual Differences in Statistical Word-Referent Learning in Infants. *Language Learning and Development*, *9*(1), 25–49. https://doi.org/10.1080/15475441.2012.707104

Suanda, S. H., & Namy, L. L. (2012). Detailed behavioral analysis as a window into cross-situational word learning. *Cognitive Science*, *36*(3), 545-559.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1). https://doi.org/10.1016/0010-0285(80)90005-5

Tuninetti, A., Mulak, K. E., & Escudero, P. (2020). Cross-Situational Word Learning in Two Foreign Languages: Effects of Native Language and Perceptual Difficulty. *Frontiers in Communication*, *5*. https://doi.org/10.3389/fcomm.2020.602471

Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, *C–21*(3), 269–281. https://doi.org/10.1109/TC.1972.5008949