Using Spatial Context to Facilitate Inductive Inference for Word Learning

Sophie Domanski (smd@umd.edu)
Department of Hearing and Speech Sciences, 7251 Preinkert Drive
College Park, MD 20742 USA

Yi Ting Huang (ythuang1@umd.edu)
Department of Hearing and Speech Sciences, 7251 Preinkert Drive
College Park, MD 20742 USA

Abstract

The spatial context of everyday speech to children is remarkably consistent. Words tend to repeatedly occur in the same locations, and these words are learned earlier than those which are more scattered in use. Yet little is known about *how* spatial contextualization mediates this relationship. Does more constrained spatial context itself lead to better word learning? Or does it simply correlate with other informative cues in the input? Here, we assess how word learning is influenced by different levels of spatial contextualization in naturalistic scenes. We use different teaching methods (other-directed versus self-guided) as a proxy for distinguishing how the need for inductive inference mediates reliance on spatial context. We found that greater spatial contextualization led to better word learning, but only when inductive inference was needed. Fixation patterns during familiarization further showed that learners paid more attention to objects with more contextualized labels when learning was self-guided. Overall, these findings suggest that learners can leverage spatial context to support word learning in the absence of rich linguistic input.

Keywords: spatial context; word learning; language acquisition; eye tracking

Introduction

Over the course of days and months and years, rhythms and regularities emerge from patterns of caregiver language that can be leveraged by children to support language acquisition (Bruner, 1985; Custode & Tamis-LeMonda, 2020). For example, caregivers tend to repeatedly use words in the same locations, and these words are learned earlier than those which are more scattered in use (Roy et al., 2015). A word like "shoe," which usually appears by the front door, is therefore likely to be learned earlier than a word like "floor," which can occur anywhere in the home. By forming the physical backdrop upon which word learning takes place, spatial context can provide a supplementary source of trackable input to help learners infer connections between words and referents (Samuelson et al., 2011).

Nonetheless, a challenge to interpreting the role of spatial context in word learning is that children's everyday lives contain many correlated cues to meaning. That is, words that consistently occur in the same locations also tend to be used in consistent routines, at similar times of day, and by the same speakers (Custode & Tamis-LeMonda, 2020). This presents a challenge for isolating the role of spatial context alone. While observational studies of child-caregiver interactions provide insight into how word exposures are distributed across natural settings, they cannot control for the presence of overlapping cues in the input (Custode & Tamis-LeMonda, 2020; Soderstrom & Wittebolle, 2013). At the same time, reliance on lagging measures of learning (e.g., age of first production) means that an indeterminate amount of time may pass between when learning actually occurs and when it is displayed (Roy et al., 2015). In the case of experimental studies, spatial context has typically been defined narrowly, with "location" referring to position in front of the child or position on a screen with an otherwise blank background (Samuelson et al., 2011; Benitez & Smith, 2012; Roembke & McMurray, 2016; Dautriche & Chemla, 2014).

To isolate the role of spatial context in word learning, experimental methods are needed that will a) manipulate spatial and linguistic information across *naturalistic scenes* and b) measure learning as it happens to provide insight into the real-time processing of the visual environment. This approach is justified by previous work in developmental psychology showing that spatial information is richly represented in the mind via geometric pathways that allow for the representation of distance, angle, and direction and non-geometric pathways that allow for representation of cues like color, pattern, and landmarks (Hermer & Spelke, 1994; Spelke et al., 2010; Lee, et al., 2006). While this work shows that spatial information is represented early in development, it remains unclear how spatial context may be informative for word learning.

A well-known obstacle to aligning words with their referents is that the visual environment is messy and laden with potential options (Medina et al., 2011). Without knowing the perspective of the speaker, it can be extremely challenging to draw conclusions about what words mean (Quine, 1960). One possible way in which spatial context eases this challenge may be by facilitating inductive inference. By helping to narrow potential word-referent mappings to those repeatedly occurring together in the same environments, spatial context may act as a backdrop upon which likely pairings are highlighted. This possibility is consistent with work showing that more spatially-consistent words are learned more easily, particularly when learners can internally guide their attention to predicted object locations (Benitez & Smith, 2012). It is also supported by work showing that caregivers naturally tend to maintain objects in spatially consistent locations when interacting with their

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.*

children (Roy et al., 2015; Benitez & Smith, 2012; Samuelson et al., 2011).

Furthermore, shared location may help connect words/objects that are thematically linked in a way that supports learning. This is consistent with evidence showing that learning gains may be boosted when named objects that repeatedly occur together also share a semantic relationship (e.g., animals found in a zoo, clothes found in a closet) (Dautriche & Chemla, 2014; Chen & Yu, 2017).

An alternative possibility is that spatial context does not drive learning itself, but simply *correlates* with other cues that are useful for word learning. Indeed, since all interactions must inevitably take place somewhere in space, it is possible for a strong spatial consistency in word-object co-occurrences to arise without necessarily being informative for the learner. This possibility is consistent with the fact that abundant social, linguistic, and temporal cues are regularly layered with spatial context (Custode & Tamis-LeMonda, 2020; Roy et al., 2015; Soderstrom & Wittebolle, 2013). Resolving these possibilities will require linking traditional theories of word learning to theories of information processing that consider how different sources of data can be integrated to support learning across domains. Achieving this resolution will in turn require a more nuanced understanding of how attention and memory mechanisms are implicated in word learning.

Both hypotheses about the role of spatial context are plausible a priori given past research on word learning, and traditional accounts of word learning vary in the degree to which they expect learners to attend to and recall information across disparate learning events. Under a cross-situational learning account, learners can track co-occurrences between words and potential referents across many encounters, using distributional statistics to extract word-referent mappings (Yu & Smith, 2007, Smith & Yu, 2008). Under a propose-but-verify account, such an approach is understood to be computationally untenable, such that learners might instead follow a procedure by which they make a 'best guess' about a word's meaning that is bolstered or weakened by further encounters with that word (Trueswell et al., 2013; Medina et al., 2011). Given these diverging views on what learners can attend to and recall, figuring out where spatial context fits within these accounts will require understanding how different attentional and memory processes are affected by changes to contextualization.

Here, we investigate the role of spatial context in word learning by examining how learning accuracy is affected by how words are taught (i.e., other-directed versus self-guided) and the degree to which learning is spatially constrained (i.e., words are limited to appearing in one location versus many locations, or some degree in between). In the other-directed teaching condition, the need for inductive inference is low as attention is exogenously-driven and word-objects pairs are taught directly. In the self-guided teaching condition, the need for inductive inference is high as attention is endogenously-driven and word-object pairs must be inferred independently by learners. This teaching manipulation is motivated by evidence suggesting that internal and external attentional processes are governed by different neural mechanisms, and that information processing differs by whether attention is directed by top-down versus bottom-up influences (Landau et al., 2007; Chica et al., 2013). We additionally leverage real-time eyetracking to observe patterns of fixations in naturalistic scenes in order to understand how contextualization influences learners' processing of the visual environment during learning.

If increased spatial contextualization facilitates word learning by supporting the learner's ability to infer word-referent mappings, then learning accuracy should be higher for word occurrences concentrated in fewer locations. Furthermore, if learners track spatial context primarily to support inductive inference, then this effect should be more pronounced when teaching is self-guided (and thus the need for inductive inference is high) compared to when teaching is other-directed (and thus the need for inductive inference is low). In either case, accuracy is expected to be higher overall when teaching is other-directed due to the high degree of transparency between words and referents compared to when teaching is self-guided.

Method

Participants

Sixty-four undergraduate students participated in exchange for course credit. Half were assigned to the self-guided condition and the other half to the other-directed condition. Four additional participants were excluded due to eyetracker malfunction during the experiment.

Procedure

Participants were told they would be participating in a study about the interaction between location and word learning. They were asked to look around and listen while they encountered novel labels and objects embedded in four rooms of a virtual home during the first part of the experiment (familiarization phase). They were informed they would later be asked to match objects with their labels in the second part of the experiment (test phase).

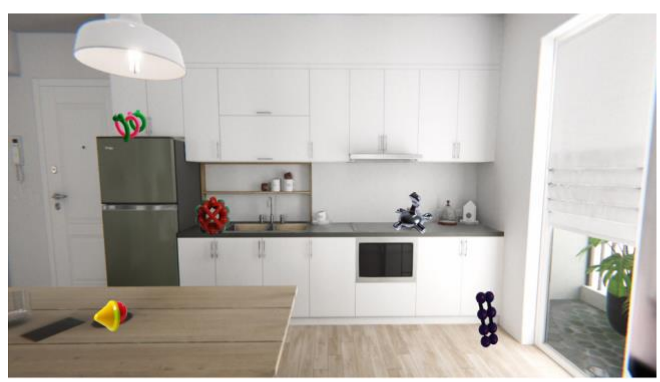

Figure 1: Screenshot of the "kitchen" location, with target items embedded throughout.

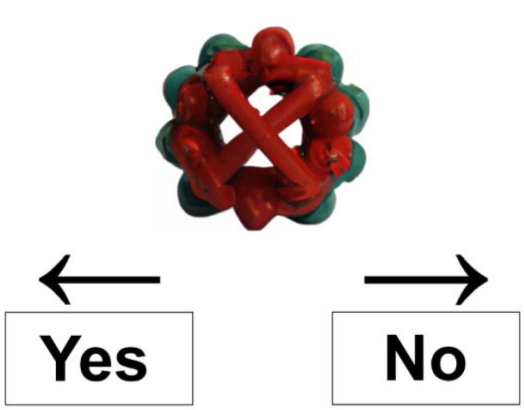

Figure 2: Sample test phase trial

Figure 3: Set of eight stimulus items used in this study.

Familiarization overview In the familiarization phase, participants cycled through viewing four "rooms" of a virtual home corresponding to a kitchen, bathroom, living room, and bedroom (in that order) (Fig. 1). Each of the four rooms was viewed four times total.

In the self-guided condition, participants had to look at target objects to hear their corresponding labels. Thus, labeling events were triggered when participant gaze remained fixated on a target object for at least 50 ms. Importantly, participants were not told that labeling events were gaze-contingent, and no explicit indication was given on screen that the label being spoken matched the object that was looked at. A given object label could be triggered up to 5 times each time it was a target within a room. A room would be exited after each target object label was triggered 5 times or after 105 seconds, whichever came first.

In the other-directed condition, participant behavior did not influence labeling events. Instead, the order, frequency, and time-course of label presentation were controlled to mirror the experience of a randomly matched self-guided participant. This information was extracted from the fixation report automatically generated by the eyetracking software upon completion of the experiment. Importantly, target objects flashed conspicuously while labels were being spoken to indicate a correspondence between them, in contrast with the continuously static images seen by self-guided participants.

While vastly abstracted, differences in word-referent transparency across conditions were intended to simulate two teaching situations often encountered by children: one in which a knowledgeable adult directly teaches a word-referent relationship with explicit and salient cues (relatively rare) versus one in which the child must infer meaning and discover relationships between words and referents independently (more common).

Testing overview The test phase consisted of 32 trials. In this phase, participants encountered individual target objects one at a time on a blank white background while hearing one target label. On each trial they were asked to respond "yes" (it's a match) or "no" (it's not a match) using left and right arrow keys on a keyboard (Fig. 2). Each of the eight stimulus items was subject to an identification response four total times, in which two trials were a match between object and label, and two trials were a mismatch between object and label.

Materials

Novel objects were created using OpenAI's DALL-E text-to-image model (Fig. 3). Objects were selected following a norming procedure in which individuals rated items on familiarity and perceived likelihood to appear in different rooms of a home. Final items were those rated as low in overall familiarity and those that were not strongly associated with any particular place. Novel words/objects varied in contextualization level, or whether they occurred in only one room ("most contextualized"), four rooms ("most dispersed"), or two rooms ("moderately contextualized," "moderately dispersed") (Fig. 4). The moderate levels were distinguished by how concentrated word occurrences were, with word occurrences either equally divided between the two rooms (moderately dispersed level) or occurring more frequently in one room than the other (moderately contextualized level). This distinction permitted investigation of how learners attend to objects differently based on concentration of *labeling events* (i.e., equally divided or skewed toward one room) rather than simple visual saliency, since they occurred equally often (i.e., in exactly two rooms). Words were presented in generic sentence frames ("look, a <nonce>") and were all two-syllable nonce words obeying phonotactic properties of English (*zollix*, *lorbu*, *meepin*, *dopeck*, *porvah*, *froobo*, *garbee*, and *shahkar*).

Results

We assessed the influence of spatial contextualization on word learning in two ways. First, we evaluated how attention to target objects differed by teaching method as evidenced by fixation patterns during the familiarization phase. Second, we evaluated how overall word learning differed by teaching method and how these differences were split by contextualization level during the test phase.

Location | Word	zollix	lorbu	porvah	garbee	meepin	dopeck	froobo	shahkar
Kitchen	1x	1x	0x	2x	1x	0x	4x	0x
Bathroom	1x	1x	2x	0x	3x	0x	0x	0x
Living Room	1x	1x	2x	0x	0x	3x	0x	0x
Bedroom	1x	1x	0x	2x	0x	1x	0x	4x
Contextualization Level	Most Dispersed		Moderately Dispersed		Moderately Contextualized		Most Contextualized	

Figure 4: Contextualization levels for the eight stimulus words-object pairs used in the study. Numbers refer to how many times a novel object was a target in a given location. Labeling events could be triggered up to 5 times per round in which an object was a target, but might be triggered as few as 0 times due to gaze-contingency.

Familiarization phase

Fixations were analyzed during the 1750 ms following the onset of a labeling event to its completion. This analysis compared looks to target objects across the self-guided and other-directed conditions, with looks further split by contextualization level. "Looks to target" were defined as looks to the labeled object while looks to any other part of the screen (or offscreen) were coded as "other" across 50 ms time bins, with the overall fixation for a 50 ms increment based on where the participant fixated for more than 50% of the increment. Figure 5 illustrates that overall looks to target were higher in the other-directed condition compared to the self-guided condition. It also shows that within each condition, looks were further differentiated by contextualization level. A linear mixed-effects model predicting proportion of looks to target confirmed an interaction between teaching condition and contextualization (χ^2 (3) = 20.75, $p < 0.001$).

Follow-up comparisons examined pairwise differences for contextualization levels within each condition. In the other-directed condition, the most dispersed objects were looked at more than moderately dispersed objects (M = 0.05, SE = 0.02, $p < 0.01$) and most contextualized objects (M = 0.05, SE = 0.02; $p < 0.05$), but were not looked at differently than moderately contextualized objects ($p = 0.49$). Furthermore, moderately contextualized objects were looked at more than moderately dispersed objects (M = -0.04, SE = 0.02, $p < 0.05$). There was no difference in looking times between the most contextualized objects and moderately contextualized ($p = 0.11$) or moderately dispersed objects ($p = 0.57$).

In the self-guided condition, the most contextualized objects were looked at significantly more than the moderately dispersed (M = -0.06, SE = 0.02, $p < 0.01$) and most dispersed objects (M = -0.08, SE = 0.02, $p < 0.01$), but were not looked at differently than moderately contextualized objects ($p = 0.16$). Moderately contextualized objects were looked at significantly more than moderately dispersed (M = -0.08, SE = 0.02, $p < 0.001$) and most dispersed objects (M = -0.10, SE = 0.02 $p < 0.001$). No other pairwise differences were significant (all p's > 0.30).

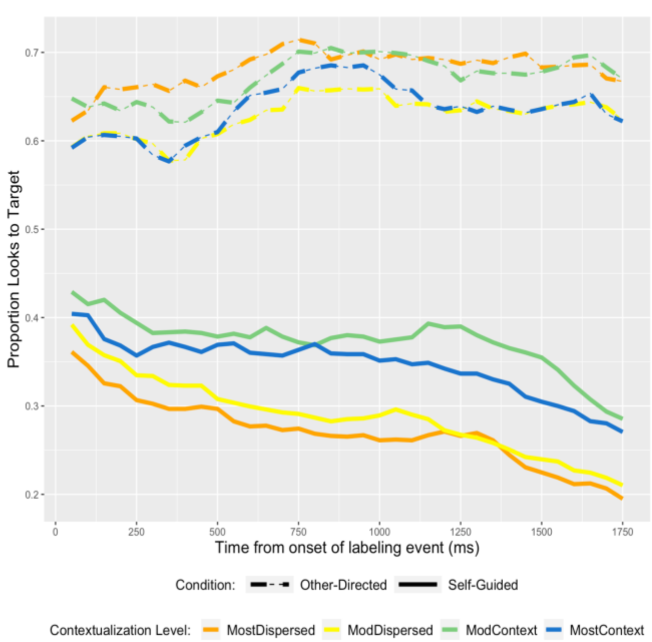

Figure 5: Time course of looks to target for different contextualization levels across each condition.

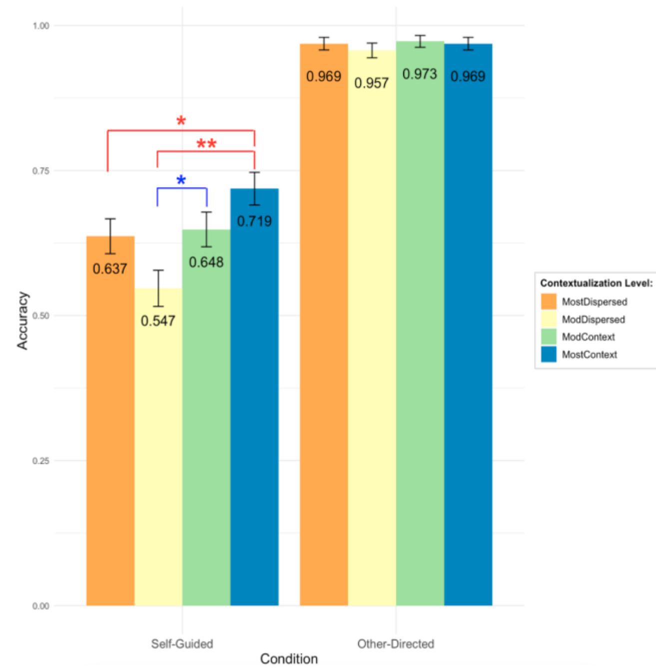

Figure 6: Word learning accuracy by teaching condition and level of spatial contextualization.

Behavioral phase

The test phase was assessed for overall accuracy in identifying words-object pairs in each condition. Accuracy was defined as the number of correct identifications of word-object matches and correct rejections of mismatches over the total number of trials. Figure 6 shows that accuracy was significantly higher in the other-directed condition (M = 96.7% , SD = 0.18) than in the self-guided condition (M = 65.0%, SD = 0.48), consistent with the expectation that words are learned more easily when word-referent mappings are readily available versus when they must be inferred (χ^2 (1) = 49.29, $p < 0.001$). However, learning was notably different across contextualization levels in each condition. In the self-guided condition, accuracy varied by level of contextualization (range: 54.7% to 71.9%; χ^2 (3) = 10.97, $p < 0.05$). In contrast, contextualization level did not impact accuracy in the other-directed condition (range: 95.7% to 97.3%; χ^2 (1) = 1.40, $p = 0.70$) .

Follow-up comparisons were performed to examine the effect of contextualization level within each condition. As stated, there were no significant differences in accuracy by contextualization level in the other-directed condition (all *p*'s > 0.30). In contrast, in the self-guided condition, accuracy was significantly better for the most contextualized level compared to the moderately dispersed (z = -3.24, $p < 0.01$) and most dispersed levels (z = -2.21, $p < 0.05$), though not better than the moderately contextualized level ($p = 0.13$). Accuracy for the moderately contextualized level was in turn significantly better than the moderately dispersed level (z = -2.42, $p < 0.05$) but not the most dispersed level ($p = 0.39$), while accuracy for the moderately dispersed level was not significantly different from the most dispersed level ($p = 0.13$).

Discussion

This study investigated how spatial context influences word learning and examined how this relationship is affected by the need for inductive inference. We discovered that learners attend more strongly to objects with more contextualized label use when learning is self-guided and inductive inference is needed, but not necessarily when word-referent pairings are taught directly. Furthermore, we found that greater spatial contextualization in word use results in better overall word learning, but *only when inductive inference is needed for learning*. Our findings suggest that spatial context is attended to by learners in a way that enables it to be used for linguistic purposes, in line with a growing literature suggesting that constrained spatial context benefits word learning (Roy et al., 2015; Benitez & Smith, 2012; Goldenberg, et al., 2022). This outcome points to a blurred boundary between "linguistic" and "non-linguistic" information, implying that learners can integrate information across domains to support vocabulary acquisition. This result is also in line with work in perception suggesting that when individuals encounter a novel object, they do not consider it in isolation, but process it as part of a scene in the manner of a "figure/ground" relationship (Wagemans et al., 2012). It is thus reasonable that information about spatial context would not be discarded, but might be tracked and used to support learning when necessary.

With regard to traditional views of word learning, the current findings are potentially compatible with both cross-situational and propose-but-verify accounts (Yu & Smith, 2007; Trueswell et al., 2013). One possibility is that spatial context is useful because it acts as a source of statistical regularity that can help narrow possible word-referent mappings to those co-occurring in the same places, consistent with predictions from cross-situational learning. Alternatively, spatial context may make it easier to form an accurate initial hypothesis about meaning for highly contextualized words (which have a smaller pool of potential referents), which in turn makes these mappings more likely to be confirmed and remembered with subsequent encounters, consistent with predictions from propose-but-verify. Further work is needed to disambiguate these possibilities. In particular, assessing word learning after a longer delay or in the same spatial context where learning took place may provide more insight as to what is carried over from the initial learning encounter.

Nevertheless, our findings also depart from traditional accounts in critical ways. Recall that under the propose-but-verify view, only a single hypothesis about a word's meaning is carried over from a given learning encounter (Trueswell et al., 2013). Yet, we found that word learning accuracy differs depending on a word's spatial contextualization, implying that learners attend to and carry over more than only a word's meaning by also retaining information about the spatial context in which it occurs. This challenges the propose-but-verify assumption that only the target word-referent mapping is carried over because additional information about the encounter would be too onerous to track. Further recall that under a cross-situational learning account, learners are expected to continuously track co-occurrences between words and referents (Yu & Smith, 2007). Yet, we found that contextualization impacted learning *only* when inductive inference was necessary. Furthermore, learners attended more strongly to objects with more concentrated labeling events *even when* those objects visually occurred equally often to objects with more dispersed labeling. This challenges the cross-situational learning assumption that tracking statistical regularities is a dumb associative process, instead suggesting that it may be selective depending on the needs of the learner (Yu & Smith, 2012).

Our findings further raise questions about what exactly makes spatial context special. Could other aspects of the input be used in the same way? We know that the brain processes visual information via distinct "what" and "where" (or "how") pathways, corresponding to ventral and dorsal streams, respectively, and that humans can systematically represent spatial features of the environment like colors, patterns, and landmarks (Goodale & Milner, 1992; Hermer & Spelke, 1994; Spelke et al., 2010; Lee et al., 2006). These facts support the notion that spatial context is not a random attribute of the environment but a highly salient feature. They further support the possibility that learners do indeed regularly track this aspect of their experience *independent* of its usefulness for word learning given its critical role in self-orientation and navigation through the world (Chiandetti & Vallortigara, 2008). Given this framework, it becomes more plausible that learners *may* in fact be able to carry over spatial "background" information across learning instances if the mind has existing mechanisms for tracking and storing this data efficiently. This in turn makes it more likely that it could be leveraged for learning across domains (such as for linguistic purposes) when needed.

Future studies might seek to validate the assumption that naturalistic scenes experienced on two-dimensional screens are an adequate proxy for locations experienced in everyday life. This could be done by replicating these findings using virtual reality (VR) to immerse participants in three-dimensional environments. Another critical step is to investigate whether these findings hold true in word learning with children who may have different underlying learning mechanisms and strategies than adults. If children indeed rely on spatial context like adults, then this may have implications for how clinical intervention is carried out in language-disordered populations, by shifting away from a focus on linguistic input only and towards a more holistic focus on the underlying learning mechanisms impacting language use and comprehension. Finally, methodological changes may bolster confidence in the conclusions drawn in this study. While no effect of contextualization was observed when teaching was other-directed, a potential difference may have been obscured by high overall accuracy approaching ceiling. Future variations may involve introducing a delay between familiarization and test phases, or including additional stimulus items to increase learning difficulty to better discern the presence or absence of a potential effect. Overall, the current study provides new evidence that learners can leverage spatial context to support word learning in the absence of rich linguistic input. Further work on this topic may enrich our understanding of how information is integrated across domains by learners.

References

Benitez, V. L., & Smith, L. B. (2012). Predictable locations aid early object name learning. *Cognition*, *125*(3), 339–352.

Bruner J (1985) Child's Talk: Learning to Use Language (W. W. Norton & Company, New York).

Chen, C. H., & Yu, C. (2017). Grounding statistical learning in context: The effects of learning and retrieval contexts on cross-situational word learning. *Psychonomic bulletin & review*, *24*, 920-926.

Chiandetti, C., & Vallortigara, G. (2008). Spatial reorientation in large and small enclosures: Comparative and developmental perspectives. *Cognitive Processing*, *9*(4), 229–238.

Chica, A. B., Bartolomeo, P., & Lupiáñez, J. (2013). Two cognitive and neural systems for endogenous and exogenous spatial attention. *Behavioural brain research*, *237*, 107–123.

Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy*, *25*(6), 809–826.

Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 892–903.

Goldenberg, E. R., Repetti, R. L., & Sandhofer, C. M. (2022). Contextual variation in language input to children: A naturalistic approach. *Developmental Psychology*, *58*(6), 1051.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20–25.

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature*, *370*(6484), 57–59.

Landau, A. N., Esterman, M., Robertson, L. C., Bentin, S., & Prinzmetal, W. (2007). Different effects of voluntary and involuntary attention on EEG activity in the gamma band. The Journal of neuroscience : the official journal of the Society for Neuroscience, 27(44), 11986–11990.

Lee, S. A., Shusterman, A., & Spelke, E. S. (2006). Reorientation and Landmark-Guided Search by Young Children: Evidence for Two Systems. *Psychological Science*, *17*(7), 577-582.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9014–9019.

OpenAI. (2023). *DALL-E* (Version 2) [Text-to-image model]. https://labs.openai.com/

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language*, *87*, 105–127.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), 12663–12668.

Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS one*, *6*(12), e28095.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS one*, *8*(11), e80646.

Spelke, E., Lee, S. A., & Izard, V. (2010). Beyond core knowledge: Natural geometry. *Cognitive science*, *34*(5), 863–884.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: fast mapping meets cross-situational word learning. *Cognitive psychology*, *66*(1), 126–156.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological bulletin*, *138*(6), 1172–1217.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological science*, *18*(5), 414–420.

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word–referent learning: Prior questions. *Psychological Review*, 119, 21.