Title
Timing and Variability Support Children’s Word Learning

Author
Price, Gwendolyn Flesher

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Timing and Variability Support Children’s Word Learning

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychology

by

Gwendolyn Flesher Price

2022

ABSTRACT OF THE DISSERTATION

Timing and Variability Support Children's Word Learning

by

Gwendolyn Flesher Price

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2022

Professor Catherine M. Sandhofer, Chair

Not all categories are made the same. Some categories have high within-category variability (e.g., "vehicles" can look very different) and some have low within-category variability (e.g., "cats" are pretty similar). Categories can also vary on their between-category variability where some categories are very similar to each other (e.g., "cats" and "dogs") and some are very different (e.g., "cats" and "vehicles"). Prior work has found that categories with high within and between variability are learned best in massed formats, and categories with low within and between variability are learned best in interleaved formats (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013). However, the unique contributions of each of these kinds of variability (i.e., within and between) have not previously been studied independently.

Two studies (*N*=128) investigated the unique contribution of within- and between-category variability to 2-year-old children's word learning in interleaved and massed presentations. These studies were conducted in a remote format through slide sharing in a video chat application. The first study investigated the impact of varying levels of between-category variability on word learning through massing and interleaving. This study found that higher between-category variability, compared to lower, led to higher performance at test (F(1,60)=13.51, p=.001, η_p^2=.184). These results show that being able to differentiate between categories is an important step in word learning. The second study investigated the impact of within-category variability. This study found no significant difference between high and low within-category variability (F(1,60)=2.89, p=.099, η_p^2=.045). In addition, neither study found significant differences between massed and interleaved conditions (Study 1: F(1,60)=0.64, p=.427, η_p^2=.011; Study 2: F(1,60)=0.05, p=.828, η_p^2=.001). These results suggest that different factors in the learning environment may impact how – and whether – children are able to learn new words in this online format. In sum, these two studies identify how between- and within-category variability impact word learning independently from each other at a stage in life where children are still developing their word learning capacities.

The dissertation of Gwendolyn Flesher Price is approved.

James W. Stigler

Idan Blank

Scott P. Johnson

Catherine M. Sandhofer, Committee Chair

University of California, Los Angeles

2022

Table of Contents

List of Figures

Acknowledgements

This is for my mother, father, stepmother, and sister, and those in my extended and adopted families. I want to express a deep and heartfelt thank you to everyone who supported me through this and made it all possible. My advisor, Cathy, and my fellow graduate students have been a wonderful bunch to spend 6 years with. Thanks to the LCD lab team: Natsuki, Christina, Michelle, Emily, Amalia, Siobhan, and Erjing, and especially my innumerable research assistants for helping me to create and refine my research, and to develop my teaching and mentoring skills. An extra special shoutout to Marissa, Jordan, and João for always helping me improve -- and stay sane. And of course, my cats Bear and Rust ought to be co-authors on this work for the amount of time they spend on my keyboard.

This work was completed during the COVID-19 pandemic, and as such I particularly want to thank the children and their caregivers who set aside time to participate in my studies. This was an incredibly difficult time for so many, and for so many reasons, but developmental science persisted due to their interest, dedication, and generosity. For that I will be forever grateful.

VITA

Education

University of California at Los Angeles **2017**
M. A. in Developmental Psychology
University of Hawaii at Mānoa **2015**
B. A. in Psychology, Summa Cum Laude

Fellowships, Grants, and Awards

Werner R. Scott Endowed Fellowship | *UCLA, Graduate Division*	2021 – 2022
Summer Mentored Research Fellowship | *UCLA, Graduate Division*	2021
Mini-Grant | *UCLA, Center for the Advancement of Teaching*	2021
TAC Educational Development Academy Fellowship | *UCLA, Center for the Advancement of Teaching*	2021
Werner R. Scott Endowed Fellowship | *UCLA, Graduate Division*	2020 – 2021
TAC Educational Development Academy Fellowship | *UCLA, Center for the Advancement of Teaching*	2020
Werner R. Scott Endowed Fellowship | *UCLA, Graduate Division*	2018 – 2019
Summer School Tuition Grant | *Swiss Graduate School for Cognition, Learning, and Memory*	2018
Werner R. Scott Endowed Fellowship | *UCLA, Graduate Division*	2017 – 2018
Graduate Summer Research Mentorship | *UCLA, Graduate Division*	2017
Distinguished University Fellowship | *UCLA, Department of Psychology*	2016 – 2017
Achievement Scholarship Award | *University of Hawaii at Mānoa*	2015

Academic Publications

Price, G. F., Ogren, M. & Sandhofer, C. M. (in press). Sorting out emotions: How labels influence emotion categorization. *Developmental Psychology*. https://doi.org/10.1037/dev0001391

Price, G. F., & Sandhofer, C. M. (2021). One versus many: Multiple examples in word learning. *Journal of Experimental Child Psychology*, *209*, 105173. https://doi.org/10.1016/j.jecp.2021.105173

Price, G. F. & Sandhofer C. M. (under review). Learning what words *do not* refer to: Using negation to learn words. *Journal of Cognition and Development.*

Selected Academic Presentations

Price, G. F. & Sandhofer, C. M. (2021) *Effects on word learning from spacing and category variability.* Poster presented at the Cognitive Science Society Annual Conference. 26 – 29 July, Vienna, Austria. (Online conference)

Price, G. F. & Sandhofer, C. M. (2021) *Parent talk: What's on the page matters.* Poster presented at the Biennial Society for Research in Child Development Conference. 7 – 9 April, Minneapolis, MN. (Online conference)

Price, G. F., Ogren, M. & Sandhofer, C. M. (2021) *Sorting out emotions: How labels influence categorization of emotions.* Poster presented at the Biennial Society for Research in Child Development Conference. 7 – 9 April, Minneapolis, MN. (Online conference)

Price, G. F. & Johnson, S. P. (2020) *Learning through interleaving and massing: Influence of memory development in infancy.* Poster presented at the International Congress on Infant Studies. 6-9 July, Glasgow, Scotland. (Online conference)

Price, G. F. & Sandhofer, C. M. (2020) *Interactions between Category Formation and Labels.* Poster presented at the Association for Psychological Science Annual Convention. 1 June, Chicago, USA. (Online conference)

Price, G. F. & Sandhofer, C. M. (2019) *Contrast in word learning: to not or not? The use of labels versus negation.* Poster presented at the Cognitive Development Society Biennial Conference. 17 – 19 October, Louisville, KY.

Price, G. F. & Sandhofer, C. M. (2019) *Timing shifts during learning: What changes and what persists.* Talk presented at the Annual Symposium on Cognitive and Language Development. 1 June, Los Angeles, CA.

Price, G. F. & Sandhofer, C. M. (2019) *The shape bias and the effect of comparison in word learning.* Talk presented at the Association for Psychological Science Annual Convention. 23 – 26 May, Washington D.C., USA.

Price, G. F. & Sandhofer, C. M. (2019) *Parent comparison and contrast speech is affected by variation of present visual display and child language comprehension.* Poster presented at the Cognitive Science Society Annual Conference. 24-27 July, Montreal, Canada.

Price, G. F. & Sandhofer, C. M. (2019). *The differential influence of comparison on children's word learning with an imposed delay.* Poster presented at the Society for Research in Child Development (SRCD) Biennial Conference. 21 – 23 March, Baltimore, MD.

Price, G. F. & Sandhofer, C. M. (2018). *Evidence of the effects of comparison and contrast in lab-based and semi-naturalistic contexts.* Talk presented at the Annual Symposium on Cognitive and Language Development. 26 May, Irvine, CA.

Price, G. F. & Sandhofer, C. M. (2018). *A disadvantage of comparison and contrast in object label learning.* Poster presented at the Cognitive Science Society Annual Conference. 25-28 July, Madison, WI.

General Introduction

Categorization is the grouping of objects, entities, and ideas into equivalence classes based on shared similarity (Medin, 1983; Rosch, 1978). Categorization is important for learning, prediction, and decision-making because it allows learners to organize information and apply it to novel situations. Categorization is at the core of word learning because both learning words and learning categories require generalizing about which properties, features, or aspects of situations matter. Children discover the meaning of words such that children not only learn the word for the specific instances they have experienced (e.g., using “cat” to label their family pet) but also generalize the meaning of words appropriately to new instances (e.g., pairing the label “cat” with unfamiliar cats). Research has shown that even young infants can generalize newly learned categories in laboratory word learning studies (Henderson & Woodward, 2012; Taxitari et al., 2020).

What defines a category? According to Mervis & Rosch (1981), “a category exists whenever two or more distinguishable objects or events are treated equivalently.”. However, categories involve more than equivalence classes; categories have an internal structure (i.e., related to their features, functions, or similarities) and boundaries that are not necessarily definite (Mervis & Rosch, 1981). Categorization and the labeling of categories is a compromise between classification accuracy (e.g., the label used is general enough to be true) and the predictive power of the label (i.e., the label is specific enough to be useful) (Rosch, 1978).

This push and pull of within-category features and between-category boundaries is a factor in the developmental progression of category learning (Callanan, 1985). Where a given category falls on the balance of these two principles can determine how early in development a child learns a word (Callanan, 1985; Mervis & Crisafi, 1982). For example, broad categories (e.g., animal) are learned later than categories that are less broad (e.g., cat). However, categories that involve more specificity (e.g., calico cat) can lead to boundaries between categories that are less distinct (e.g., the boundary between a calico versus tabby cats may be less clear) and therefore are learned later than the less specific categories of “animal” or “cats.” Therefore, the broadness, or within-category variability of a category, and the boundaries, or between-category variability, must be taken into account for understanding word learning.

Learning a word involves, at its essence, connecting a category to a label (Katerelos & Poulin-Dubois, 2011; Son et al., 2008). Providing labels for categories can help children form categories (Fulkerson & Haaf, 2003; Lupyan et al., 2007; Smith et al., 2002; Waxman & Markow, 1995) and make inferences about them (Dewar & Xu, 2009; Rips, 1975). For example, 12-month-old children who were given novel labels (e.g., “Look, a fep”), compared to no labels (e.g., “Look, here’s one”), for novel shape categories were able to generalize the category to new instances (Fulkerson & Haaf, 2006). The dynamic between categories and their labels is mutually augmenting, leading to vocabulary growth (Smith et al., 2002) over development.

Children learn words quickly and without explicit instruction (Carey & Bartlett, 1978; Dewar & Xu, 2007; Smith & Yu, 2008). Indeed, children can even learn – and generalize – new words given only a single example of a category (Price & Sandhofer, 2021). And children can remember words over a long period of time even when given few learning instances (Vlach et al., 2012; Wojcik, 2013). Early word learning doesn’t even need to happen through explicit instruction (Kloos & Sloutsky, 2008) or with speech directed at the child at all (Akhtar et al., 2001). Children are incredible word learners that can rapidly gain fluency in a language despite the stochastic and oftentimes erratic environment of childhood.

How information is presented to children has broad effects on word learning (Ankowski et al., 2013; Arnon & Clark, 2011; Atagi et al., 2016; Blaye & Jacques, 2009; Carey & Bartlett, 1978; Dixon et al., 2006; Flack & Horst, 2018; Golinkoff et al., 1992; Horst, 2013; Johanson & Papafragou, 2016; Kovack-Lesh & Oakes, 2007; LaTourrette & Waxman, 2020; Nagle & O'brien, 1987; Price & Sandhofer, 2021; Vlach et al., 2008; Vlach & Johnson, 2013; Yoshida & Smith, 2005). In this dissertation, I examine two aspects of the word learning environment – timing and variability – to better understand how children learn words with different category structures. These studies will examine how changes in the timing of examples interact with changes in category structure to impact word learning.

Timing in Word Learning

Multiple studies have demonstrated that timing plays a large role in word learning; Timing refers to the distance between presented examples of a category. For example, sometimes children hear the same word multiple times in a row; a child might hear, "Look at all these *cats*! This is a cat, and this is another kind of cat!" Other times, children hear words directed to a single referent, and some time passes before a child hears the word "cat" again. The timing between the instances of cat impacts how, and even whether, children can abstract the category structure and remember the concept over time.

Children typically see more than one example of a category, and they see these examples either simultaneously or at different points in time. Simultaneously comparing multiple examples of a category can aid in learning new words across the lifespan (Anderson et al., 2018; Gentner & Namy, 1999; Kotovsky & Gentner, 1996; Kurtz & Boukrina, 2004; Twomey et al., 2014). Simultaneous comparison is defined as viewing multiple items at the same time. Research studies differ in how they present children with multiple items to compare. In a typical study of simultaneous comparison, one or more examples are presented to a learner all at once. For example, in Graham et al.'s (2010) comparison study, when two examples of a category were presented and labeled, they were displayed simultaneously. However, in other studies, comparison is defined as massed presentations (i.e., one example presented directly after another) (Anderson, Chang, Hespos, & Gentner, 2018; Carvalho & Goldstone, 2014). Multiple examples can also be presented spaced apart in time (i.e., one after another with time in between each instance) (Childers & Tomasello, 2002; Kornell & Bjork, 2008). Whether the category examples are presented massed or spaced, using the same label makes clear that the examples share category membership (e.g., each round red object is labeled “apple”).

A large body of research has examined how spaced comparison can benefit learning. Spacing instances of learning in time often benefits memory for categories more than massing when tested after a delay (Birnbaum et al., 2013; R. A. Bjork, 2011; Kornell & Bjork, 2008). This learning benefit is termed “the spacing effect.” The spacing effect is a robust finding among adults, infants, and animal learning studies (Ebbinghaus, 1964/1885; Fanselow & Tighe, 1988; Linde et al., 1985; Toppino, 1991). Moreover, spacing examples of a category out over time can help support category learning (R. A. Bjork & Allen, 1970).

There has been some research examining how children’s word learning is affected by spacing. For example, in one study, children were given novel objects of the same shape labeled with novel words, and they were presented either simultaneously, sequentially, or with spacing (Vlach et al., 2012). The results indicated that simultaneous presentations led to greater generalization than spaced presentations at an immediate test. However, when there was a brief delay between learning the words and testing, spaced presentations led to greater learning. This is evidence that spacing learning events apart in time promotes the generalization of words and categories. Other research indicates that spacing effects are found in nouns and verbs (Childers & Tomasello, 2002), memorization of an object and generalization of a category (Vlach et al., 2008), and first and second language learning (Nakata & Elgort, 2021).

However, spacing does not always lead to higher performance than massing. The spacing effect seems to depend on children's vocabulary knowledge (Slone & Sandhofer, 2017). Some researchers have recently called for more work on the specifics of the spacing effect in early childhood to determine what developmental changes and individual characteristics may interact with the spacing effect (Knabe & Vlach, 2020), especially in light of the relative dearth of work in this area compared to that of the adult literature.

Although spacing has demonstrated effects on how well children learn words in the laboratory, it does not always translate well to non-laboratory settings. In everyday word learning contexts, children are unlikely to encounter a single labeling instance with periods of non-verbal time between events. Instead, a word is more likely to be heard in between instances of other words. When multiple examples of a word are interspersed, it is referred to as interleaving. Interleaving is similar to spacing such that there is time between examples of a word, but the interstimulus interval contains other learning events (Birnbaum et al., 2013). For example, if a child is learning the words “cat” and “dog,” they may hear cats referenced, then dog, then cats again, and finally the dog again. This alternating between instances of multiple words is at the core of interleaving.

Interleaving typically leads to higher rates of learning than massed presentations. Interleaving helps learners overcome context-dependent learning and fixation in problem-solving (R. A. Bjork, 2011; Storm, 2011). Interleaving also benefits learning by promoting discrimination between examples of words. That is, by drawing attention to the differences between the category of things that a word refers to (Kang & Pashler, 2012; Kornell & Bjork, 2008). When learning multiple words, the examples of the two words are sometimes similar enough that learners must focus on the differences to learn where to separate the two ideas. For example, when learning "tabby cat" and "calico cat," learners might notice many features in common between the two animals, so they must instead focus on the color and pattern of the cats' coats to differentiate the two categories. In these instances, interleaving shows increased benefits over massing.

Variability in Word Learning

The variability between and within categories has been shown to have large effects on word learning (Althaus & Plunkett, 2016; Childers et al., 2017; Hahn et al., 2005; Kovack-Lesh & Oakes, 2007; Perry et al., 2010; Thibaut & Witt, 2015). The referents for a word may vary in how similar they are to each other. For example, two "cats" have many features in common (e.g., similar in shape, sound, etc., with small differences in fur color and size). However, two "vehicles" may have fewer features in common (e.g., an airplane and a boat are both transport vessels, but their shape, mode of propulsion, and usual location are all different). In some research, the variability within examples of a category helps children learn the underlying commonalities necessary for category membership (Perry et al., 2010; Quinn & Bhatt, 2010). For example, category members often have variability in features that are unnecessary for category inclusion (e.g., the length – or even presence - of a cat's fur is irrelevant to being a cat). This kind of variability within the category leads children to generalize the category structure better than if they were given objects with no within-category variability (Twomey et al., 2014; Perry et al., 2010).

Moreover, children develop vocabulary earlier when taught categories with more within-category variability. In one study, 18-month-old children were taught new categories with examples that either had high or low within-category variability (Perry et al., 2010). For instance, for the category of “bucket,” children were either presented with buckets that varied in material, size, and color but were similar in shape, or children were presented with relatively similar buckets in material, size, color, and shape. The children who experienced high within-category variability had a significantly larger vocabulary at a one-month follow-up than those taught with low variability categories. This may be because children sift through the variation in examples of a category to find what it means to be part of the category, and they do so best when given multiple variable examples. Variability within a category teaches children when to attend to one feature (e.g., shape), and importantly when to not attend to a feature (e.g., color) to learn what it means for an object to be a bucket.

On the other hand, some evidence suggests that higher within-category variability is not always beneficial for word learning (Childers et al., 2017; Maguire et al., 2008; Vlach et al., 2008). Vlach et al. (2008) found a benefit for less within-category variability in massed and spaced presentations. The researchers found that children who were taught novel nouns with the *same* example presented multiple times performed better than those who were taught a category with multiple *different* examples. However, that study used a memorization test (i.e., asked children to pick out the same object they saw before) for the no-variability condition and used a generalization test (i.e., asked children to select a new member of the category) for the variable condition. Therefore, these tests are difficult to compare, as one was a test of recognition and the other a test of generalization. This is important because the variability in examples has been shown to aid category generalization, but variability effects have not been found when the test asks children to select an item they have seen during training. This indicates that there may be some limits to the benefits of within-category variability for learning new words, such that less within-category variability may be more helpful for memorizing specific items (Vlach et al., 2008), or when learning later learned and action-based parts of speech like verbs (Maguire et al., 2008). However, extending these ideas to everyday word learning where generalization is needed is difficult. Understanding how low and high within-category variability impacts children's ability to generalize new words and how this interacts with presentation timing is open for further research.

In addition, other studies suggest a middle ground approach; that some constancy paired with some within-category variability may help young children learn new words (Goldenberg & Sandhofer, 2013; Quinn & Bhatt, 2010; Raz et al., 2019). That is, too much within-category variability may lead children to fail to understand the properties of the category well enough to understand the category structure and generalize to new instances. Too little within-category variability, on the other hand, may lead children to be unclear on how much variation is acceptable within the category. For example, one study by Goldenberg & Sandhofer (2013) presented 2-year-old children with a word learning task that included a background context (e.g., a cloth that the object was presented on) that is either always the same, always different, or with some variation *and* some consistency. Children are therefore more likely to generalize when the examples are presented with some variation and some consistency.

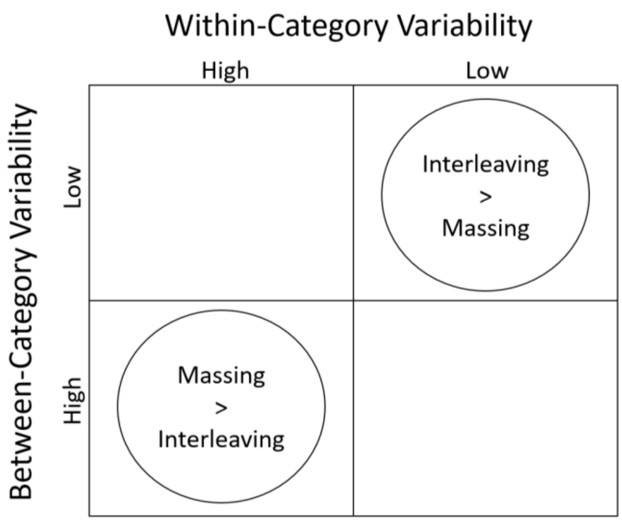
Variability and Spacing

Combining presentation timing and object variability makes between-category variability particularly salient. Between-category variability measures how different two categories are from each other. For example, cats and birds differ in many features (e.g., coat, shape, size, sound, ability to fly). These two animals would have high between-category variability. On the other hand, cats and dogs are somewhat similar-looking animals (e.g., general size, fur texture, number of legs) but still differ in a few ways (e.g., ear and face shape, sound). Based on their shared visual and textural features, these two animals would have lower between-category variability.

Massing and interleaving present a learner with two or more categories in different timing schedules. Thus, massing and interleaving necessarily combine two (or more) categories, making the perceptual variability between two (or more) categories an essential part of the learning event. As a result, massed and interleaved presentation styles affect category learning differently depending on the variability between categories in addition to the within category variability discussed above. Massing draws attention to the features that indicate category inclusion (Carvalho & Goldstone, 2017). That is, the features of the category that matter for category inclusion may be more readily noticed when multiple examples are presented one after another, highlighting the shared and consistent features. Thus, when the relevant features for a category are difficult to identify, or when there is more within-category variability, massed presentation helps to aggregate that information better than interleaving (Carvalho & Goldstone, 2014; Sorensen & Woltz, 2016). In Figure 1, this is represented by the bottom left quadrant. When two categories are more difficult to differentiate, or when there is less between-category variability, interleaving can help the learner discriminate between categories more so than massing. For example, when two different categories had low between-category variability, that is, when it was difficult to determine what makes the categories distinct, adults performed better on tests of generalization when the examples of the categories were presented in an interleaved format than when they were presented in a massed format (Zulkiply & Burt, 2013). In Figure 1, this is represented by the top right quadrant. Variability can interact between massing and spacing due to differing demands on understanding what instances belong to the category and how categories differ from one another.

Figure 1

Prior Literature on Between- and Within-Category Variability



Note. This chart represents results from the prior literature. Each cell represents one combination of Between- and Within-Category Variability.

The Current Studies

Separately, spacing and variability have been shown to increase category learning in children and adults (Kornell & Bjork, 2008; Vlach & Sandhofer, 2012). In studies that combine spacing and variability, variability appears to influence the effectiveness of spacing in learning studies (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013). However, these studies simultaneously varied between- and within-category variability, leaving researchers unable to tease apart their unique contributions to this interaction.

The current studies used shape categories to understand how variability in combination with the spacing effect may affect generalization. Shape words, and nouns in general, make up the majority of the early vocabularies for English-learning children (Gentner, 1982). Children as early as two years of age are biased to attend to shape when labeling new objects (Landau et al., 1988); that is, children who learn the label for a new object will extend the label to other objects of the same shape. In addition, previous spacing studies found the effect for shape categories, but not other early learned category types like color or texture, which may be due to the predominance of shape words in children's early vocabularies (Slone & Sandhofer, 2017). Lastly, previous studies investigating spacing and between- and within-category variability have used shape-based categories (Carvalho & Goldstone, 2015a; Slone & Sandhofer, 2017; Vlach et al., 2012; Zulkiply & Burt, 2013). Thus, we used shape categories in these studies to compare this work to prior literature and ensure the two-year-old children would be able to learn the words through a spacing presentation.

Children were particularly important to study in this context for several reasons. First, this interaction between example variability and spacing has been studied in adults but not in children, so there is a dearth of evidence to show how this category variability and spacing interaction effects may play out with children who do not have robust vocabulary and experience learning and using language. Second, children are less likely to rely on deterministic features for categorization (e.g., all wugs have an orange dot) and are more likely to categorize by the overall similarity of the examples (Deng & Sloutsky, 2016). Because children categorize objects based on holistic similarity more than adults, variability and spacing may differentially impact learning. Third, studies on adult learning frequently identify effortful practices contribute to learning (E. L. Bjork & Bjork, 2011); however, given the developing vocabulary, executive attention, working memory, and other cognitive processes in children's (Frank et al., 2016; Gathercole et al., 2004; Jones et al., 2003), learning practices that may be helpful to adults may be simply too difficult to be helpful to children. Therefore, it is important to study children's learning to examine how the results from the adult literature generalize to developing learners.

The current studies examine how the variability between categories (Study 1) and within categories (Study 2) interact with presentation timing in 2-year-old children. Investigating these ideas will further the understanding of the utility of spacing in early word learning and how category structure may pose different challenges in word learning.

Study 1

Discriminating between different categories is an important aspect of word learning and has been studied across the age span (see Murphy, 2002, for a review). Some categories are easy to learn, even for young infants (Anderson et al., 2018; Eimas & Quinn, 1994); however, others are difficult, even for adults (Zulkiply & Burt, 2013). Several factors affect learnability. However, one reason some categories may be easier or harder to learn is the variability between the categories being learned simultaneously. When two categories are perceptually similar, for example, cats and dogs, greater attention to the differences may be required to discriminate between them. However, categories that are perceptually different, for example, cats and birds, may require less emphasis on category discrimination.

The interaction between between-category variability and spacing may be an important factor in category discrimination. Spacing allows for deeper understanding and generalization of categories. One type of spacing, interleaving, also provides discriminative contrast. When categories are interleaved, the learner switches between the two categories each time until all examples are shown. This switching back and forth between the categories draws the learner's attention to the differences between the categories and contrasts the two categories to help the learner discriminate between them (Kang & Pashler, 2012; Kornell & Bjork, 2008). Another type of presentation, called massing, is when the learner sees all the examples of one category before seeing the examples of another category. This type of presentation does not draw attention to the differences between the categories. Therefore, learners may benefit more from interleaving than massing when there is a larger need to discriminate between categories, that is, when there is low between-category variability.

The present study assessed two-year-old children's ability to learn novel shape categories using a novel noun generalization task. Children were tasked with learning categories with either high or low between-category variability. Additionally, children learned these categories in either an interleaved or massed presentation. Because discriminating between categories is essential to learning new words, I hypothesized that children would have higher performance at test in the high between-category variability condition than in the low between-category variability condition overall. Second, because massing leads to higher test performance than interleaving when tested soon after learning, I hypothesized that children would have higher test performance in the massed than the interleaved condition when given categories with high between-category variability. Third, when children were presented with categories with low between-category variability, we hypothesized that they would have difficulty discriminating between categories. As such, I hypothesized that children would have higher performance in the interleaved condition compared to the massed condition when they were presented with categories with low between-category variability.

Method

Participants

The participants were 64 two-year-old children (*M*=30.03 months, *SD*=3.06; 36 Female, 28 Male). In this sample, 14 of the children were currently in preschool, two had previously attended before the COVID-19 pandemic, and 48 had never enrolled in preschool. Preschool enrollment is important to note due to the significant impacts it has on children's language development, especially in the short term (Yoshikawa et al., 2016), and may have other social learning implications such as the ways children expect to learn and whom they expect to learn from. Of the 64 children in the sample, 38 were white, 15 mixed, 4 Asian, 1 Black/African American, and 5 chose not to report. This sample was highly educated, with 44 reporting at least one parent with a graduate or professional degree, 19 with at least one parent with a 4-year college degree, and 1 chose not to report.

Before participating in the study, informed consent was obtained from each child's parent or guardian, and verbal assent was obtained from each child. Participants were recruited from a birth records database and online science platforms (e.g., https://childrenhelpingscience.com/). All participants were learning English as a primary language.

Design and Study Overview

The study used a 2 x 2 design, with between-category variability (high or low) and spacing (interleaved or massed) as between-subjects' factors. In the study children were randomly assigned to one of the four conditions: High between-variability and Interleaved (*N*=16), High between-variability and Massed (*N*=16), Low between-variability and Interleaved (*N*=16), Low between-variability and Massed (*N*=16). There were no significant differences between the conditions in terms of children's age ($F(3,60) = 1.49$, $p = .225$, $\eta_p^2 = 0.07$), gender ($F(3,60) = 2.11$, $p = .109$, $\eta_p^2 = 0.10$), nor preschool attendance ($F(3,60) = 0.80$, $p = .499$, $\eta_p^2 = 0.04$).

Timing Conditions

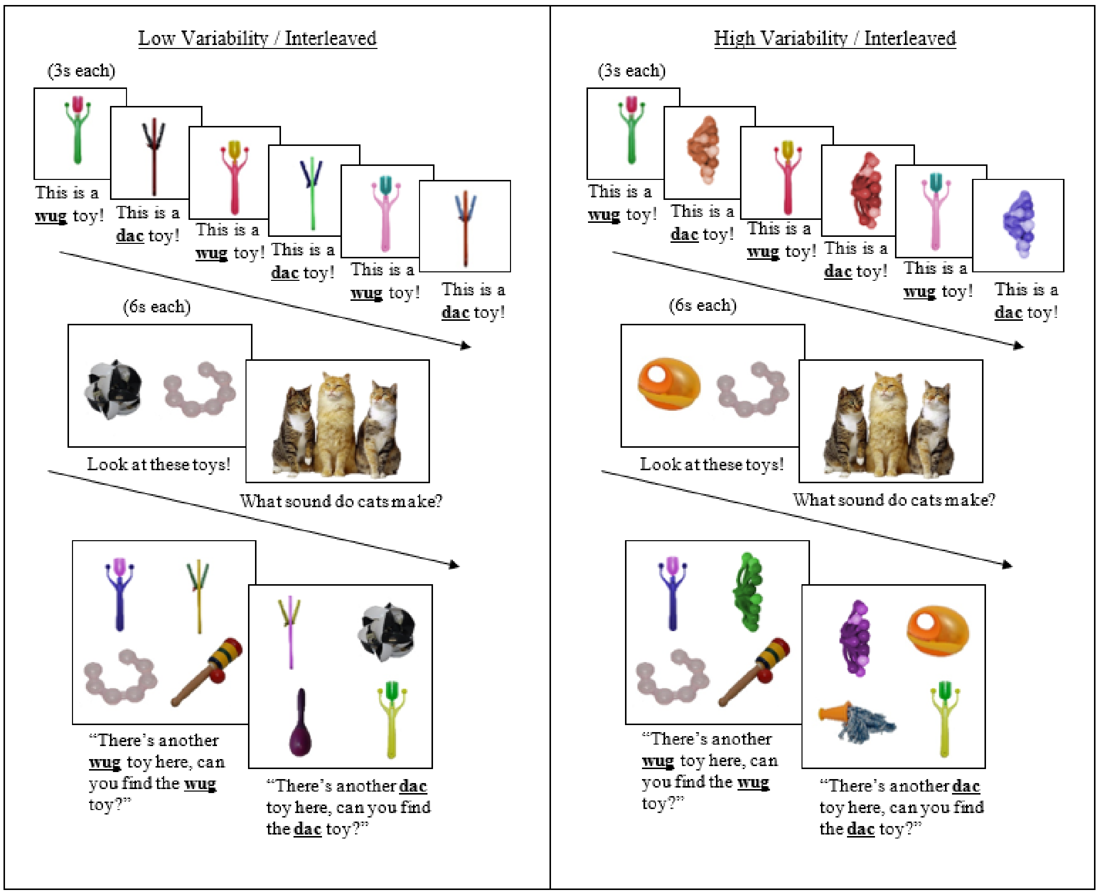
The interleaved and massed manipulations were created by changing which items were presented in which order, whether all items of a category were presented one by one in a row (i.e., massed) or interspersed between examples of the other category (i.e., interleaved). The study took place entirely on a video chat service (i.e., Zoom), and the stimuli were presented to children on PowerPoint slides through screen sharing. The data collected were a demographic survey sent to parents before the start of the study and participants' behavioral responses by pointing to the screen.

High and Low Between-Category Variability Stimuli

The NOUN Database (Horst & Hout, 2014) was used to create pairs of categories that were high and pairs that were low in between-category variability. The NOUN Database includes ratings for between-category variability independent of within-category variability. The images were rated by independent adult raters using a similarity sorting task and sorted using a multidimensional scaling analysis (Horst & Hout, 2016). Each set of images we created consisted of 6 category slides (3 for each of the two categories), one distractor slide that displayed two images (one for each category), one delay slide, and two test slides that displayed four test images (one test set for each category). The examples within each category were all rated as having low within-category variability (i.e., the category instances were highly similar). Different categories were paired to create high and low between-category variability pairs. Four category pairs (i.e., eight total categories) were selected for each variability condition. The category pairs with high between-category variability exceeded the distance rating threshold of .87 in multidimensional space (M = 1.17, SD = 0.05). Category pairs selected as having low between-category variability were below that threshold (M = 0.56, SD = 0.12). Figures 2 and 3 show examples of the low and high between-category variability pairs selected for this study.

Figure 2

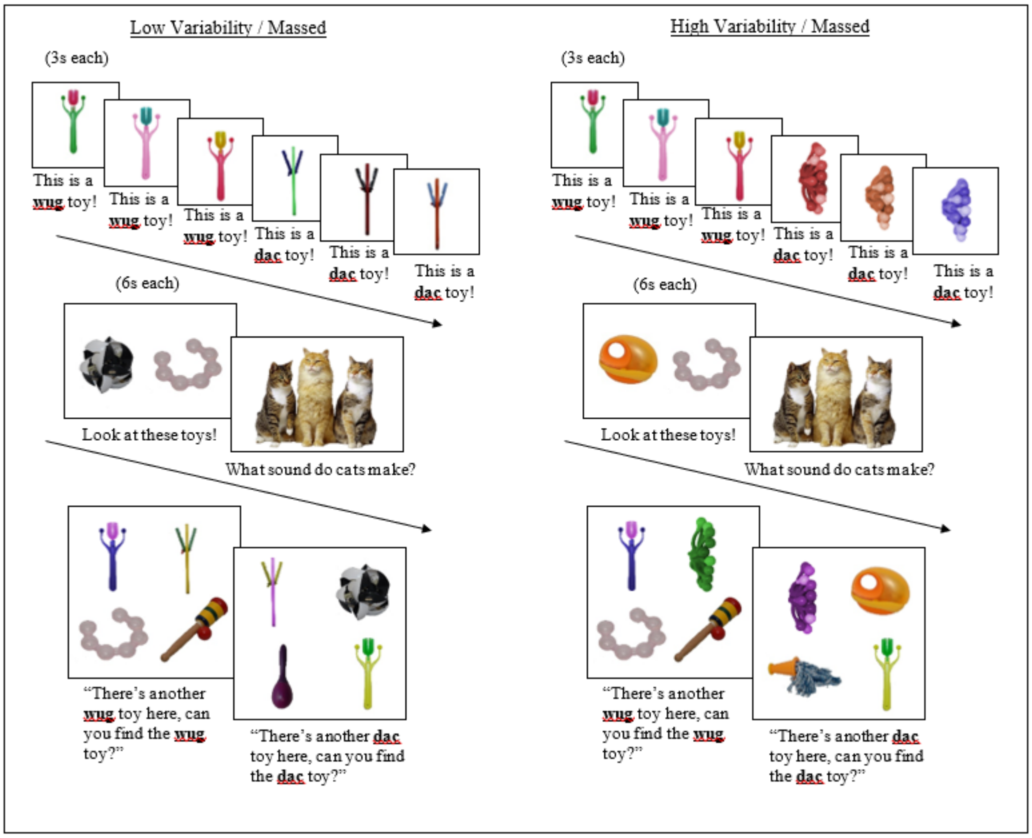
Study 1 Interleaved Method for Low and High Variability



Note. The left panel is an example of the Low Variability condition, and the right panel is an example of the High Variability condition. They both depict these conditions in the Interleaved format. In each panel, the top row is the learning phase, the middle row is the distracter phase, and the bottom row is the test phase.

Figure 3

Study 1 Massed Method for Low and High Variability



Note. The left panel is an example of the Low Variability condition, and the right panel is an example of the High Variability condition. They both depict these conditions in the Massed format. In each panel, the top row is the learning phase, the middle row is the distracter phase, and the bottom row is the test phase.

Distractor Stimuli

The distractor items were selected to have one of the highest distance ratings from each tested category. If the item with the highest distance rating was unavailable due to being in use for another trial, or if the item was from another category within the experiment, then the item with the next highest distance rating was selected to ensure each distractor item was unique.

Test Stimuli

Across the study, there were four sets of two tests, each of the four learning phases was followed by two test slides. Each test had four images. One image was a new instance of the tested category. A second image was a new instance of the other category from the learning set. A third image was one of the distractor items. The fourth image was the close item. The close item at test was selected to have one of the lowest distance ratings from the tested category. If the item with the lowest distance rating was used for another trial within the experiment, then the item with the next lowest distance rating was selected to ensure each close item was unique. This object was selected to ensure that it was as close as possible to create a reasonable option at test. Because the variability in this study was between the categories, the test sets were different for high and low between-category variability conditions due to the change in which categories were paired.

All stimuli for the experiment are displayed in Appendices A (high between-category variability) and B (low between-category variability).

Procedure

Children were randomly assigned to one of the four learning conditions. Each condition began with a familiarization task. Then children proceeded through four trials. Each trial included a learning, distracter, and test phase.

Familiarization

The study began with a familiarization task. See Figure 4 for the four pictures of objects used in the task (i.e., shoe, cup, cat, and ball). Children were asked to point to each item in turn (i.e., “Can you point to the cat?”). The parent then said aloud the corresponding number (e.g., they were told, “if your child points to the object at the top right of the screen, say ‘two’”) to ensure the experimenter accurately understood where the child was pointing. Children were given two chances to point to each of the four items. If children pointed to each of the four items accurately, they moved on to the next task.

Figure 4

Familiarization image

Note. All four images represent words known on average by 89% of children at the lowest age in this study, 24 months, and by 97.25% of children by the mean age in this study, 30 months (Frank et al., 2016).

Novel Noun Generalization Task

The novel noun generalization task exposed children to eight different categories across four trials. Each trial had a learning, distracter, and test phase. See Figures 2 (interleaved conditions) and 3 (massed conditions) for an example of a trial of each of the four conditions.

Learning Phase. In the learning phase of a trial, children were shown six slides with one image each. Three slides showed an example from one category, and three slides showed an example from a second category. The two paired categories had either low or high between-category variability. The category examples were either presented in a massed format (i.e., the three examples from category 1 were presented before the three examples from category 2) or an interleaved format (i.e., each example from one category 1 was followed by an example from category 2). Each example was presented for 3 seconds and was labeled with a novel word (e.g., "This is a wug toy").

Distracter Phase. Immediately following the learning phase was the distracter phase. The distracter phase consisted of two images on one slide. They were presented for 6 seconds and were not labeled (e.g., "Look at these!").

Delay. Children were then presented with a delay slide (i.e., cats, oranges, balls, or dogs) for 6 seconds following the distracter phase. These objects were not labeled with a novel word. Instead, they were discussed generally (e.g., "What are these?", "Can you count how many cats there are?", "What sound do cats make?"). This functioned as a delay between when children saw the distracter phase items and when they were tested on the categories, as the test included the distracter items. See Figure 5 for one of the slides shown to participants.

Figure 5

Delay slide

Note. This is one of the four delay slides used in the study. The other slides contained (1) dogs, (2) oranges, and (3) balls.

Test Phase. After the delay, two test trials were presented in a randomized order, one for each of the two learned categories. Each test was prompted with the phrase, “Can you point to the (novel noun) toy?” The picture was presented for as long as it took the child to respond. The child’s response was recorded, and the next trial began immediately.

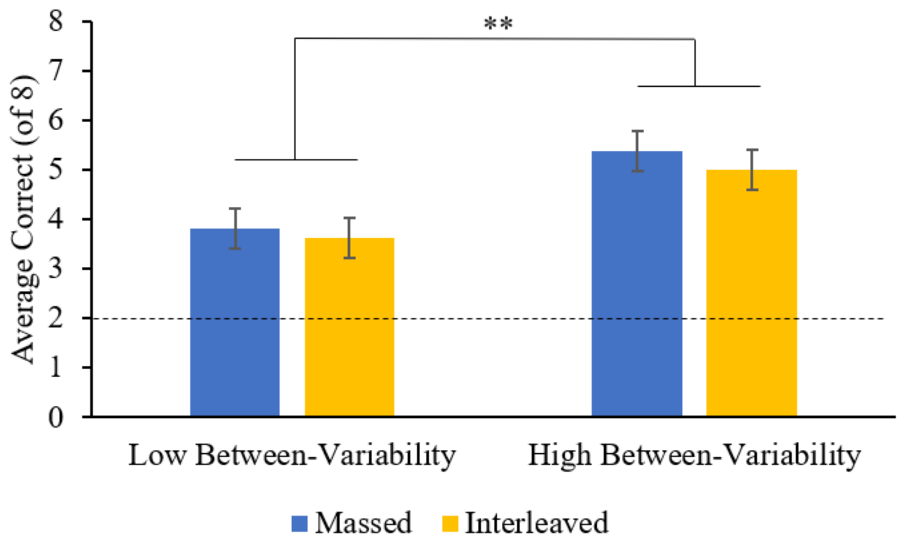
Results

I first asked whether the four conditions individually led children to learn new words. To answer this question, I performed one-sample t-tests on each of the four conditions and compared them against chance. All four conditions were above chance: low variability/massed ($t(15)=4.12$, $p=.001$, $d=1.03$), low variability/interleaved ($t(15)=3.81$, $p=.002$, $d=0.95$), high variability/massed ($t(15)=11.21$, $p<.001$, $d=2.80$), and high variability/interleaved ($t(15)=7.70$, $p<.001$, $d=1.93$).

I then addressed the primary goal of this study to examine whether between-category variability and the timing of presentations of examples (e.g., interleaved or massed) affected children’s category learning. To examine how the four learning conditions affected performance, I conducted a 2x2 between-subjects Analysis of Variance (ANOVA). See Figure 6 for a graphical representation of the results. The two-way ANOVA showed that test performance was higher on average if there was more ($M=5.16$, $SD=1.37$) rather than less ($M=3.72$, $SD=1.71$) between-category variability ($F(1,60)=13.51$, $p=.001$, $\eta_p^2=.184$). The ANOVA also showed that there were no significant differences between the massed ($M=4.59$, $SD=1.68$) and interleaved ($M=4.28$, $SD=1.73$) conditions ($F(1,60)=0.64$, $p=.427$, $\eta_p^2=.011$), nor was there an interaction between level of between-category variability and the spacing conditions (*F*(1,60)=0.10, *p*=.750, η_p^2=.002).

Figure 6

Study 1 Results



Note. Average selection of target object at test out of 8 trials: Low Variability & Massed (3.81), Low Variability & Interleaved (3.63), High Variability & Massed (5.38), and High Variability & Interleaved (5.00). Error bars represent standard error.

A third set of analyses was performed to determine whether the between-category variability led to higher confusability between the two categories in each trial set and whether that differed based on the spacing conditions. To answer this question, I conducted a 2x2 between-subjects ANOVA. The dependent variable was how often children chose the object from the other category they learned in that set (e.g., if they chose the wug toy when asked for the dac toy). The results showed that children chose the other word at test more often if there was less (*M*=2.59, *SD*=1.50) rather than more (*M*=1.61, *SD*=1.09) between-category variability

(F(1,60)=9.51, p=.003, η_p^2=.137). The ANOVA also showed that for this dependent variable, there were no significant differences between the massed (M=2.03, SD=1.53) and interleaved (M=2.13, SD=1.29) conditions (F(1,60)=0.08, p=.780, η_p^2=.001), nor was there an interaction between level of between-category variability and the spacing conditions (F(1,60)=0.08, p=.780, η_p^2=.001).

A fourth set of analyses was performed to determine whether between-category variability or spacing impacted children's confusion on what features (i.e., shape, texture, or color) to pay attention to during learning. To answer this question, I conducted another 2x2 between-subjects ANOVA, and this time the dependent variable was how often children chose the close match object at test. The results showed no significant differences based on level of between-category variability (F(1,60)=0.85, p=.361, η_p^2=.014), spacing (F(1,60)=0.85, p=.361, η_p^2=.014), nor the interaction between between-category variability and spacing (F(1,60)=1.91, p=.172, η_p^2=.031).

The final set of analyses was performed to determine whether the four pairs of high between-category variability objects, and separately the four pairs of low between-category variability objects, were equally important in driving the between-category effect. To answer this question, I conducted a 2x2 within-subjects ANOVA to determine whether there were significant differences between the high between-category variability pairs. The results showed no significant differences between the four high between-category variability pairs (F(3,93)=0.15, p=.927, η_p^2=.005). I also conducted a separate within-subjects ANOVA to determine whether there were significant differences between the low between-category variability pairs. The results also showed no differences between the four low between-category variability pairs ($F(3,93)=2.58$, $p=.058$, $\eta_p^2=.077$).

Discussion

This study directly targeted the between-category variability aspect of learning multiple words simultaneously. Categories with higher between-category variability led to higher performance, but this did not differ based on the type of spaced presentation that the examples appeared in. Increased test performance for high between-category variability sets may be due to the ease of discriminability. Categories with low between-category variability may be easily confusable, as children who learned categories with low between-category variability were more likely to choose the other object they learned in that trial.

Interestingly, children in this study did not show significant differences at test between interleaving and massing in the low between-variability condition. This deviates from prior research showing that interleaved presentations help learners discriminate between categories with many overlapping features (Carvalho & Goldstone, 2014; Kang & Pashler, 2012; Zulkiply & Burt, 2013). This may be due to the short time frame in which each trial took place in the current study. While studies have found the spacing effect in young children even with rapid presentation of examples (Toppino, 1991), many more studies have shown a delay between learning and test is necessary for the full effect to show. In the current study, the delay between learning and test was about 12 seconds, much shorter than previous research. Previous research with adults has instituted longer delays of 5 to 20 minutes between learning and test (Carvalho & Goldstone, 2014, 2017; Kang & Pashler, 2012; Zulkiply & Burt, 2013). In related studies with children, delays of 30 seconds to 2 minutes have been used and lead to above chance performance on tests of word learning (Rea & Modigliani, 1987; Vlach et al., 2012). However, those studies were conducted in person, which may have helped keep young children's attention on task, thus allowing the children to participate in a longer task successfully than they could when interacting with an experimenter through a screen. Indeed, some work has shown that conducting studies with children under the age of 3 on a screen can lead to lower test performance than in in-person studies (Jing & Kirkorian, 2020).

Additionally, the test itself was different from previous studies on between- and within-category variability due to the age of the participants. In previous research, adult participants decided which category an item belonged to on 36 successive test trials, and each item belonged to one of the categories (Carvalho & Goldstone, 2014, 2015, 2017), or the participants were asked to select the category's label from a set of 12 labels for each of the test items (Kang & Pashler, 2012; Zulkiply & Burt, 2013). For the current study's 2-year-old participants, such tests would not be feasible for them to complete due to their inability to read and lowered ability to attend to a single task for as long as it would take to complete 36 test trials. These test differences may account for some of the differences in results between the current and prior research.

The lack of a difference in the test results between massing and interleaving in the conditions with high between-category variability may be due to ease of differentiation and ability to identify the category structure. Although some studies have argued that massing leads to lowered performance than interleaving (Birnbaum et al., 2013; Vlach et al., 2012), other studies have argued that massing leads to higher performance (Carvalho & Goldstone, 2015a; Sorensen & Woltz, 2016). Massing leads to higher performance than interleaving when the category structure is difficult to determine, that is, when there is high within-category variability. In this study, the within-category variability is low for every condition, perhaps making the benefits of massing unneeded. Researchers have argued that interleaving is most beneficial when the categories are hard to differentiate (Birnbaum et al., 2013). Thus, in the present study, the conditions with high between-category variability may not have needed extra help to differentiate between the categories.

Future studies should examine how changes in between-category variability impact learning when within-category variability is held at a high rate across conditions. This may further realize the role of massing in learning when categories are difficult to learn and generalize. If these results hold when within-category variability is high, then it would indicate that between-category variability is an independent factor in word learning worthy of note in studies that teach children multiple words at once. In addition, increasing the amount of spacing in the study and adding in a longer delay than was instituted here may further realize the spacing effect in this interaction.

Study 2

Understanding the structure of a category is necessary for word learning and generalization. A child needs to understand that "bird" can refer to an emperor penguin, a pigeon, and a seagull. Importantly, not all categories are created equal in terms of their amount of within-category variability. A category like "bird" may have substantially more variability in shape, size, texture, color, etc., than a category like "apple." The amount of variability within a category impacts how well a child can generalize the label to a new instance of the category (Carvalho et al., 2021; Higgins & Ross, 2011; Lewis & Frank, 2018; Maguire et al., 2008; Monaghan et al., 2017; Perry et al., 2010; Quinn & Bhatt, 2010).

In many previous studies, the variability within a category impacted children's word learning. Learning a category with high within-category variability can help children generalize novel shape categories (Perry et al., 2010), but for other word types, like verbs, it can prevent initial understanding of the category structure (Childers et al., 2017; Maguire et al., 2008). Higher within-category variability in the examples of a shape category can result in more robust learning in terms of reduced dependence on any one particular example to generalize the category (Monaghan et al., 2017) and acceptance of a more extensive range of possible features into the category set (Lewis & Frank, 2018; Perry et al., 2010). Therefore, within-category variability is an important feature of categories that impacts children's acquisition.

As discussed in the previous study, some research has been done with adult learners to show the interaction between within- and between-category variability and spacing (Carvalho & Goldstone, 2014, 2015b; Zulkiply & Burt, 2013), but the studies have always manipulated between- and within-category variability at the same time. Some studies on children's learning suggest that the variability in examples helps children's category learning (Perry et al., 2010; Quinn & Bhatt, 2010), and in separate studies, the spacing of examples in time helps children's category learning (Toppino, 1991; Vlach et al., 2012). However, studies examining within-category variability in children's word learning have not manipulated the timing of presentations (Maguire et al., 2008; Perry et al., 2010). Moreover, studies examining the effect of spaced presentations have not manipulated the within-category variability of the examples presented to children, except to show the effects of a memorization test where there was no variability in category examples (Vlach et al., 2008).

Difficulties like interleaving and within-category variability can be desirable for learning (R. A. Bjork & Kroll, 2015; Perry et al., 2010; Rost & Mcmurray, 2009; Singh, 2008). As the name would suggest, desirable difficulties in learning are situations or procedures that are difficult yet counterintuitively benefit learning (E. L. Bjork & Bjork, 2011; R. A. Bjork & Kroll, 2015; Knabe & Vlach, 2020). Interleaving is one such desirable difficulty (Kornell & Bjork, 2008). Variability in category exemplars may also be a desirable difficulty in word learning that helps specifically with the generalization of the category (Perry et al., 2010). The interaction between interleaved presentations and high within-category variability may further increase learning. However, it is also possible that having both spacing and high within-category variability in the same learning event may instead be too difficult to be desirable.

In this study, I asked how 2-year-old children's ability to learn novel shape categories was affected by the within-category variability and the spacing of the learned examples. Children were tasked with learning categories with either high or low within-category variability. They learned these categories in either an interleaved or massed presentation. I first hypothesized that because within-category variability aids generalization (Estes & Burke, 1953; Perry et al., 2010), there would be a main effect of within-category variability, such that high within-category variability would lead to higher test performance than low within-category variability. Second, we hypothesized that massing would lead to higher test results than interleaving for categories with high within-category variability because massing draws learners' attention toward the similar features within a category (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013). Third, we hypothesized that interleaving would lead to higher test results than massing in the low within-category variability condition due to the increased discrimination between categories with interleaving that would help learners differentiate the similar categories.

Method

Participants

The participants were 64 two-year-old children (*M*=29.08 months, *SD*=3.40; 30 Female, 34 Male). In this sample, 20 of the children were currently in some form of preschool, four had previously attended before the pandemic, and 40 had never enrolled in preschool. As with the previous study, preschool enrollment is important to note due to the significant impacts on children's language development (Yoshikawa et al., 2016). Of the 64 children in the sample, 40 were white, 13 mixed, 6 Asian, 2 Black/African American, 1 Hispanic, 1 American Indian/Alaskan Native, and 1 chose not to report. This sample was highly educated, with 43 reporting at least one parent with a graduate or professional degree, 20 with a 4-year college degree, and 1 with some college experience.

Before participating in the study, informed consent was obtained from each child's parent or guardian, and verbal assent was obtained from each child. Participants were recruited from a birth records database and online science platforms (e.g., https://childrenhelpingscience.com/). All participants were learning English as a primary language.

Design and Study Overview

The study used a 2 x 2 design, with within-category variability (high or low) and spacing (interleaved or massed) as between-subjects' factors. Children were randomly assigned to one of four conditions: High within-variability and interleaved (N=16), high within-variability and massed (N=16), low within-variability and interleaved (N=16), and low within-variability and massed (N=16). There were no significant differences between the conditions in children's age ($F(3,60) = 1.37$, $p = .261$, $\eta_p^2 = 0.06$), gender ($F(3,60) = 0.90$, $p = .446$, $\eta_p^2 = 0.04$) nor preschool attendance ($F(3,60) = 0.14$, $p = .935$, $\eta_p^2 = 0.01$).

To manipulate within-category variability, we used images of novel objects. Novel objects were created to ensure a standard level of within-category variability within each category for low and high within-category variability. The interleaved and massed manipulations were created in the same format as in Study 1 by changing which items were presented in which order, all examples of one category in a row (i.e., massed), or examples of both categories interspersed in between each other (i.e., interleaved). As with Study 1, this study took place entirely on Zoom, a video chat service, and the stimuli were shown to the participants via screen sharing a PowerPoint presentation. The data collected were a demographic survey filled out by the parents and the behavioral responses of the child participants as they pointed to the screen.

Stimuli

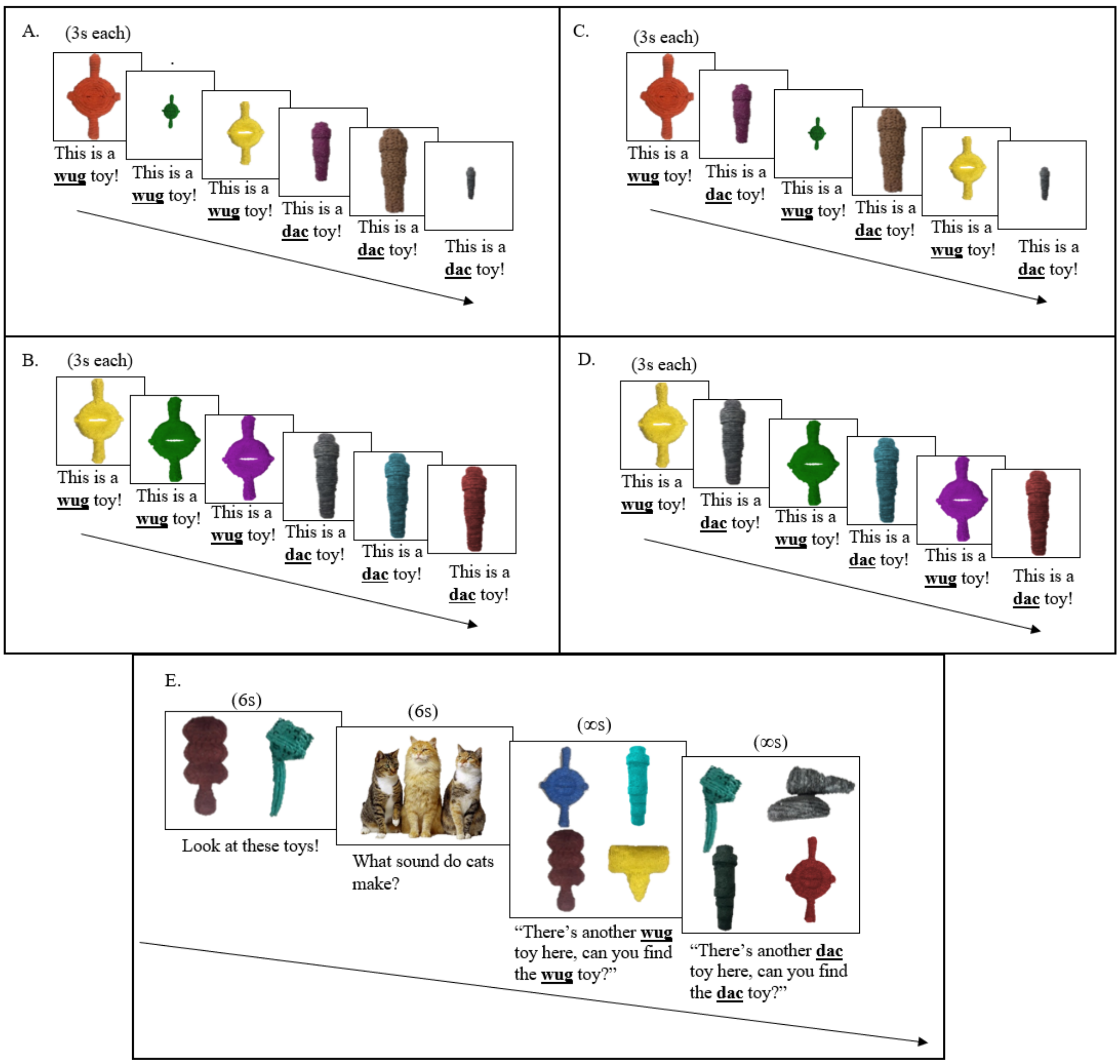
The stimuli were pictures of novel objects. Figure 7 shows an example of a trial showing the Interleaved and Massed conditions and the two Variability conditions. In each trial, there were six slides depicting the two categories (3 pictures depicting a single object per category), one slide depicting the two distractor objects, one slide of the delay items, and two test slides (one for each of the categories) each with four test phase images.

The novel categories were defined by their shape, meaning that the objects' shapes did not differ within a category in any condition. Each category had either low or high within-category variability. In the low within-category variability condition, each category example was the same texture and size but differed in color. In the high within-category variability condition, each category example differed in texture, size, and color. Regardless of within-category variability, each pair of categories learned had the same amount of between-category variability (i.e., the categories always differed in color and texture but were similar in size).

The distractor slide showed two different images from the categories being learned. These images had different shapes and colors but had the same size as two of the learning phases images. As with Study 1, there were two test slides, and each test slide had four images. One image was a new instance of the tested category. The second image was a new instance of the other category learned in the set. A third image was one of the distractor items. The fourth image was the close item. The close item was different from the tested category on shape but was similar in texture and color to one of the tested category's learning phase items. Because the variability in this study was *within* the category, the test sets were the same across all four conditions.

Figure 7

Study 2 Stimuli



Note. Panel A is an example of the High Variability condition, and panel B is an example of the

Low Variability condition. Panels A and B both depict these conditions in the Massed format. Panel C (High Variability) and Panel D (Low Variability) depict these conditions in the Interleaved format. Panel E depicts the distractor and test phases that would follow any one of the four conditions in Panels A through D.

Procedure

The procedure was the same as it was for Study 1. Children were randomly assigned to one of the four learning conditions. Each condition began with a familiarization task. Then children proceeded through four trials. Each trial included a learning, distracter, and test phase.

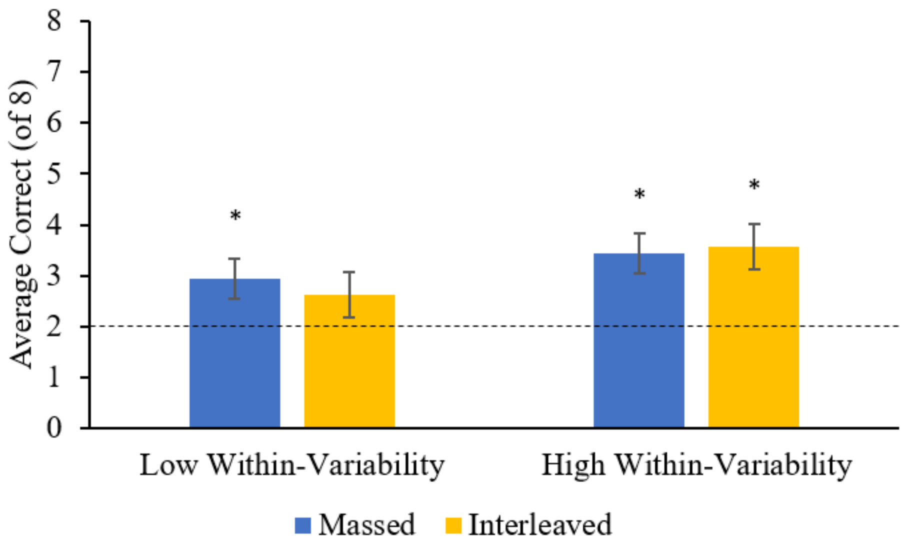
Results

I first asked whether children learned the words in each of the conditions. To answer this question, I performed one-sample t-tests on each of the four conditions to compare them individually against chance. Three conditions were above chance: low variability/massed ($t(15)=2.61$, $p=.020$, $d=0.65$), high variability/massed ($t(15)=3.29$, $p=.005$, $d=0.82$), and high variability/interleaved ($t(15)=3.13$, $p=.007$, $d=0.78$). The low variability/interleaved condition was the only condition not to be significantly different from chance ($t(15)=1.54$, $p=.145$, $d=0.38$).

I then addressed the main goal of this study, that is, to determine whether within-category variability impacted category learning and whether the timing of presentations of examples (i.e., interleaved or massed) interacted with this type of variability. To examine how the learning conditions affected performance, I conducted a 2x2 Analysis of Variance (ANOVA). See Figure 8 for a graphical representation of the results. The two-way ANOVA showed no significant differences between the high ($M=3.50$, $SD=1.85$) and low ($M=2.78$, $SD=1.52$) within-category variability conditions (F(1,60)=2.89, p=.099, η_p^2=.045). The ANOVA also showed that there were no significant differences between the massed (M=4.59, SD=1.68) and interleaved (M=4.28, SD=1.73) conditions (F(1,60)=0.05, p=.828, η_p^2=.001), nor was there an interaction between level of within-category variability and the spacing conditions (F(1,60)=0.26, p=.612, η_p^2=.004).

Figure 8

Study 2 Results



Note. Average selection of target object at test out of 8 trials: Low Variability & Massed (2.94), Low Variability & Interleaved (2.63), High Variability & Massed (3.44), and High Variability & Interleaved (3.56). Error baThes represent standard error.

A third set of analyses were performed to determine whether within-category variability led to higher confusability between the two categories in each trial set and whether that differed based on the spacing conditions. To answer this question, I conducted a 2x2 ANOVA; the dependent variable was how often children chose the object from the other category they learned in that set (e.g., if they chose the wug toy when asked for the dac toy). The results showed no significant differences based on level of within-category variability (F(1,60)=0.96, p=.332, η_p^2=.016), spacing (F(1,60)=2.29, p=.136, η_p^2=.037), nor the interaction between variability and spacing (F(1,60)=0.20, p=.658, η_p^2=.003).

A fourth and final set of analyses were performed to determine whether between-category variability or spacing impacted children's confusion on what features (i.e., shape, texture, or color) to pay attention to during learning. To answer this question, I conducted a 2x2 ANOVA; the dependent variable was how often children chose the close match object at test. The results showed no significant differences based on level of between-category variability (F(1,60)=0.04, p=.844, η_p^2=.001), spacing (F(1,60)=0.35, p=.556, η_p^2=.006), nor the interaction between variability and spacing (F(1,60)=0.35, p=.556, η_p^2=.006).

Discussion

This study examined how within-category variability impacted children when learning multiple words in two different spaced presentation styles. Children did not significantly differ in any of the four learning conditions in this study. The lack of a difference between conditions based on choice for the other object they learned could be because the between-category variability was held at a constant level across all four conditions, making it equally easy to differentiate between categories, mirroring the high between-category variability condition from Study 1. The children in this study also did not seem to struggle more with understanding the category structure when there was higher within-category variability, as evidenced by the non-significant change in choice for the other category and close match objects at test.

One condition did seem to struggle more than the others: the low within-category variability and interleaved condition. A lack of variability and the extra difficulty of interleaving examples seemed to lead children to have lower test performance. This aligns with prior research in that children seem to have lower scores on generalization tests when there is less variability within the category (Perry et al., 2010). Literature on interleaving considers it a desirable difficulty (R. A. Bjork & Kroll, 2015); however, this may be one instance where the difficulty was too high to be desirable.

The lack of a difference between high and low within-category variability differs from prior literature. Category variability literature has previously shown the benefits of within-category variability on children's word learning and generalization for variability in the objects themselves (Perry et al., 2010; Thibaut & Witt, 2015), in the speakers labeling the objects (Rost & Mcmurray, 2010), and in the context surrounding the objects (Goldenberg & Sandhofer, 2013). In this study, differences in variability did not seem to impact learning, suggesting a difference in methodology (i.e., online vs. in-person), or type (i.e., size, color, and perceived texture) and amount of variability (i.e., measured as three features instead of a multi-dimensional scaling task as with Study 1) may have impacted word learning.

The dimensions in which within-category variability was presented may have impacted the results. For variability in size, previous work has shown that even three-year-old children have a difficult time learning and generalizing a new category when presented with irrelevant size changes in category examples for color categories (Ankowski et al., 2013, Study 2), but size changes have shown to help children learn shape categories (Perry et al., 2010). However, more often than not, studies that vary within-category features do not vary examples in size but vary other features such as texture, color, shape, and speaker while keeping the size of the objects the same (Althaus & Plunkett, 2016; Eimas & Quinn, 1994; Graham et al., 2004; Rost & Mcmurray, 2009). In addition, for the online 2-D presentation of images, the size differences could have been perceived as differences in distance between the learner and the object, and not indeed that the object itself was of a different size. For variability in texture, the stimuli used in the current study were presented on a screen. Thus, the texture could only be noticed visually, which may have changed the impact of that dimension on children's conceptualization of the category. These two dimensions, size and texture, were the main within-category variability differences between the low and high within-category variability conditions. For the full effect of these within-category variability differences, especially concerning the texture changes, this study may show different results if replicated in an in-person format.

This study deviates as well from prior literature on spacing. Previous literature has shown benefits of either interleaving or massing depending on task specifics (Carvalho & Goldstone, 2015a; Kornell & Bjork, 2008). In this study, differences in spacing of category examples did not seem to impact learning in either direction. This could suggest that the instantiation of spacing in this study did not meet the requirements for the spacing effect. That is, the timing of examples may have been too quick, or the delay before the test was too short. In addition, children in each condition correctly chose the target object at test in less than half of the trials on average, perhaps suggesting that the task was too difficult for this age range.

Future studies should explore age differences and methodological changes to this study. If these results were due to task difficulty, it is possible that older children would have higher test performance on this task. This would suggest there was too much difficulty incorporating both spacing and category variability for 2-year-old children. If these results were due to methodological choices, it is possible that instituting a longer delay before the test may reveal the spacing and variability interaction as predicted. This would suggest that the benefits of the spacing effect were not given time to materialize in this study.

General Discussion

This is the first set of studies to piece apart between- and within-category variability in spaced word learning. Previous research has examined between- and within-category variability in spaced word learning, but they have not often been studied with children and have not separated the two kinds of variability to examine the individual impacts they may have on learning (Carvalho et al., 2021; Kloos & Sloutsky, 2008; Zulkiply & Burt, 2013). Indeed, other researchers have explicitly called for studies examining systematically manipulated differences in within- and between-category features, for example, Hammer (2015) wrote, “I do suggest that in order to properly determine the soundness and robustness of a hypothesized feature preference or an intrinsic attentional bias, it is necessary to test subjects in several scenarios where the respective saliency of within-category and between-categories differences is systematically manipulated.” (p.451). The current dissertation has addressed this call and uncovered how the variability of within- and between-category features impacts children’s word learning in the context of massed and interleaved learning presentations.

The first study in this dissertation examined how between-category variability impacted children’s word learning through massed and interleaved learning formats. Using the Horst and Hout (2014) Novel Noun Database, pairs of categories were created to present children with higher and lower between-category variability sets based on the multi-dimensional scaling tests. The results showed a higher learning performance at test for category sets for high compared to low between-category variability. However, there were no differences in spacing conditions or interactions between spacing and variability. This suggests that discriminability between categories that are learned together significantly impacts novel word learning.

The second study in this dissertation examined how within-category variability impacted children’s word learning through massed and interleaved learning formats. This study used images of novel objects with unique shapes, textures, and colors to control for within-category differences systematically. The results showed no differences in learning performance based on within-category variability, spacing, or the interaction of these variables. This could suggest a few things, including that the amount, type, or presentation style of variability does not impact children’s learning or that within-category variability overall does not impact learning as much as between-category learning does for this age group.

Two-year-old children were particularly important to study in this format because they categorize new objects differently than adults (Kloos & Sloutsky, 2008) and have developing cognitive capacities (Frank et al., 2016; Gathercole et al., 2004; Jones et al., 2003) that differentiate them from adults. Children categorize based on similarity and attentional factors (e.g., salient features), and adults categorize based more on conceptual information (e.g., a given label such as “animals”) (Deng & Sloutsky, 2016; Sloutsky et al., 2015). In addition, some things that are desirable difficulties for adults, like spacing (R. A. Bjork & Kroll, 2015), may play out differently due to children’s developing cognitive capacities. Therefore, the role of within- and between-category variability changes studied with adults (Carvalho & Goldstone, 2015a) may differ in children, especially regarding the interaction between variation and spacing.

Many studies have examined instances in which within- and between-category variability is helpful for learning. Differentiating between categories is necessary to understand the boundaries of a category (Andrews et al., 2005), but category differentiation is difficult for young children (Mandler et al., 1991). High between-category variability, as shown in Study 1 of this dissertation and other work (e.g., Mandler et al., 1991), may help children form categories and learn words. Within-category variability has also previously been found to be useful for learning words, specifically for instances where the categories being learned are more abstract (Higgins & Ross, 2011), or with fewer instead of multiple points of variation between examples (Goldenberg & Sandhofer, 2013; Waxman & Klibanoff, 2000). However, within-category variability did not help in the current studies, which leads to the question: why?

Multiple studies have reported conditions that show little benefit of within-category variability (Childers et al., 2017; Goldenberg & Sandhofer, 2013; Maguire et al., 2008; Raz et al., 2019) such that reducing the amount of within-category variability leads to greater attention to the relevant features of the category. Although within-category variability can be helpful for generalizing categories from a young age (Perry et al., 2010; Twomey et al., 2014), studies have shown that less within-category variability can lead to higher rates of word learning in situations such as learning fewer objects at one time (Raz et al., 2019) and increased learning for verbs when one actor acts out the verb examples instead of many actors (Childers et al., 2017; Maguire et al., 2008). The results of the current study show no significant difference between high and low within-category variability, so while the present study does not provide support for the benefits of low within-category variability, it also does not provide support for the benefits of high within-category variability.

The findings from these studies may not be directly applicable to children's everyday word learning because the everyday categories children encounter may have different distributions of within-category variability. Carvalho and colleagues (2021) have argued that real-world categories are organized around skewed featural distributions (i.e., a small set of items are seen more often, with rare items seen less often) and that our lab studies generally create categories around normal distributions. Other studies have shown that skewed featural distributions lead to narrower generalizations (Lewis & Frank, 2018), which may help children separate categories before learning what all can belong in a given category. Indeed other researchers have found that these lab-created categories with a normal distribution and high within-category variability are not beneficial even for adults, as they lead to slower responses and higher error rates during generalization, even as they have some learning "benefit" in leading learners to accept more distant (i.e., variable) items as part of the category (Hahn et al., 2005). Therefore, in the current studies, how a child learned "wug" may differ from how they first learned the real-world category "bike." This difference in item distribution may lead to a large gap between lab-based studies and real-world learning.

Previous studies, reviews, and meta-analyses on the spacing effect show it to be arguably the most replicable and robust effects in experimental psychology (Bjork, 2011; Vlach & Sandhofer, 2012). One explanation for why spacing instances in time increases memory is that it acts as a desirable difficulty. A difficulty is desirable when a learner's response to it supports learning, understanding, and memory recall (R. A. Bjork & Kroll, 2015). Spacing is a desirable difficulty because the forgetting that occurs over time between learning instances is beneficial to memory for the learned items, even if this recall is more difficult when there is more space in between. Through the effortful process of repeated memory retrieval, the memory trace is strengthened, much like the strengthening of a muscle (R. A. Bjork, 2011; R. A. Bjork & Allen, 1970). For example, in one study, (Kornell & Bjork, 2008), learners were tasked with learning artists' painting styles. The results indicated that spacing examples of a single artist out in time led to higher generalization rates over time. That is, each time an example of a given artist was presented on the screen, the difficult process of forgetting and subsequent retrieval of the memory trace was repeated, resulting in a stronger memory of the artist's style. The stronger the memory, the more likely it was that the learner would identify a new instance of a given artist's work after a long delay. Spacing allows for some amount of forgetting, which makes the retrieval event upon subsequent presentation of the items to be learned a much more powerful learning event (Jacoby, L. L., Bjork, R. A., & Kelley, 1994).

Interleaving is a powerful learning tool, and much more needs to be learned about how it impacts children's language learning. The current studies examined the difference between interleaving and massing on children's word learning as the variability changed both between and within categories. The difference between interleaving and massing is twofold; interleaving introduces spacing between examples of a category (i.e., spacing effect), and it adds the contrast between categories by interspersing the two categories together (i.e., discriminative contrast) (Birnbaum et al., 2013; Carvalho & Goldstone, 2015a). Studies have shown that depending on the level of within- and between-category difficulty, adults will show differentiation in learning between interleaved and massed formats, with higher variability both within- and between-categories being learned at higher rates in massed formats (Sorensen & Woltz, 2016; Zulkiply & Burt, 2013) and lower variability being learned at higher rates in interleaved formats (Kornell & Bjork, 2008).

An important question to ask then is why there might have been no statistical differences between the choices at test for children in the interleaved and massed conditions. One possible reason is the video deficit: the idea that learning through video leads to poorer learning outcomes than learning with 3D objects for object name learning, among other domains. This deficit is often found up to 3 years of age and often persists throughout toddlerhood, depending on the task

(Jing & Kirkorian, 2020). In addition, this presentation style of interacting with 2-year-olds through a video chat may have also been less socially engaging, leading to poorer attention and subsequent learning outcomes.

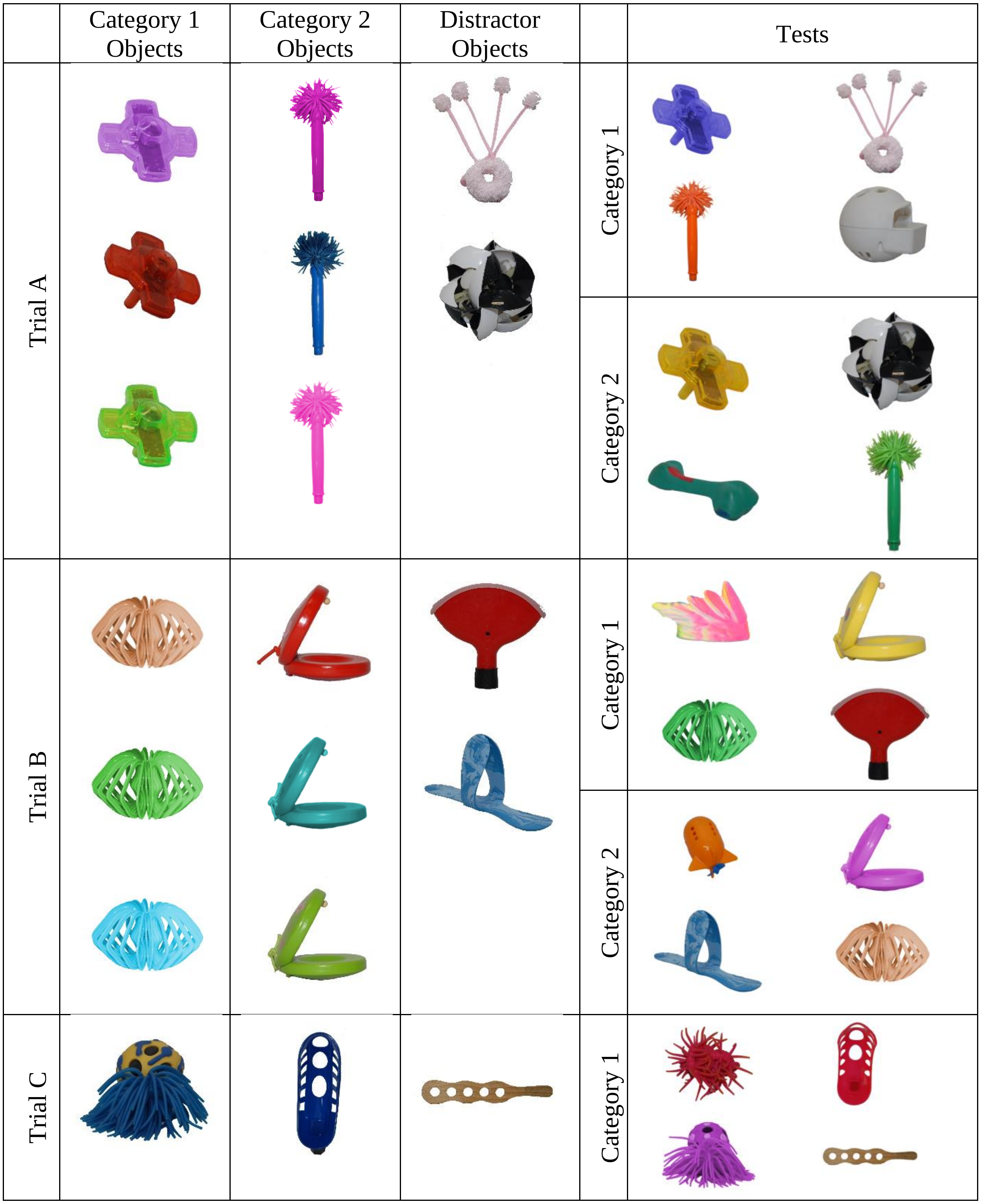
The non-significant difference between massed and interleaved presentations has been found before in younger (Vlach & Johnson, 2013) and older (Benitez et al., 2020) children's cross-situational word learning tasks. In Benitez et al. (2020), the 4- to 7-year-old children performed equally well on a cross-situational word learning task when given massed and interleaved presentations. In contrast, adults in their study did show a significant difference between the massed and interleaved conditions that was in line with prior spacing literature. However, these cross-situational studies may differ from the ostensive labeling in the current dissertation due to the higher memory and attention demands of cross-situational tasks. Nevertheless, they do mirror the current work, showing that children are capable of learning in both massed and interleaved presentation styles, but that differences between these conditions are usually seen in studies with adults. These studies suggest that the spacing effect may not be as robust in children due to their attention and memory development and that the spacing effect may be closely tied to task set up in the early years of word learning.

I will not overstate or overinterpret the null results found in these studies. Future research may want to examine these conditions in in-person contexts after the pandemic has abated. Children may have an easier time attending to and learning from experimenters who are not testing them through online video calls. Future research may also want to tweak the methodology, which would be easier to do in an in-person format, to increase the time spent on the study overall. The spacing between examples in the current studies was kept minimal to keep 2-year-old's attention on the screen. For future online studies, the methodology could instead incorporate the learning into a storybook or similar context to support children's attention while still increasing the amount of spacing between examples. While the current studies cannot state much about children's differential learning through interleaving and massing, they set the groundwork for future studies to examine other ways these conditions may help or hinder children's word learning. To echo Knabe & Vlach (2020), the spacing effect literature in young children's word learning is an important area that needs further research to understand its developmental limitations and trajectory.

In conclusion, the two studies in this dissertation explored the unique contributions of between- and within-category variability to children's word learning. These results provide valuable insight into how children understand a category's structure and differentiate between similar categories. Ultimately, these results are informative for our growing understanding of children's word learning development through spacing and the two kinds of variability.

Appendix A.

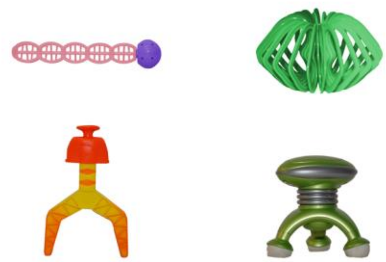
Study 1 High Between-Category Variability Stimuli

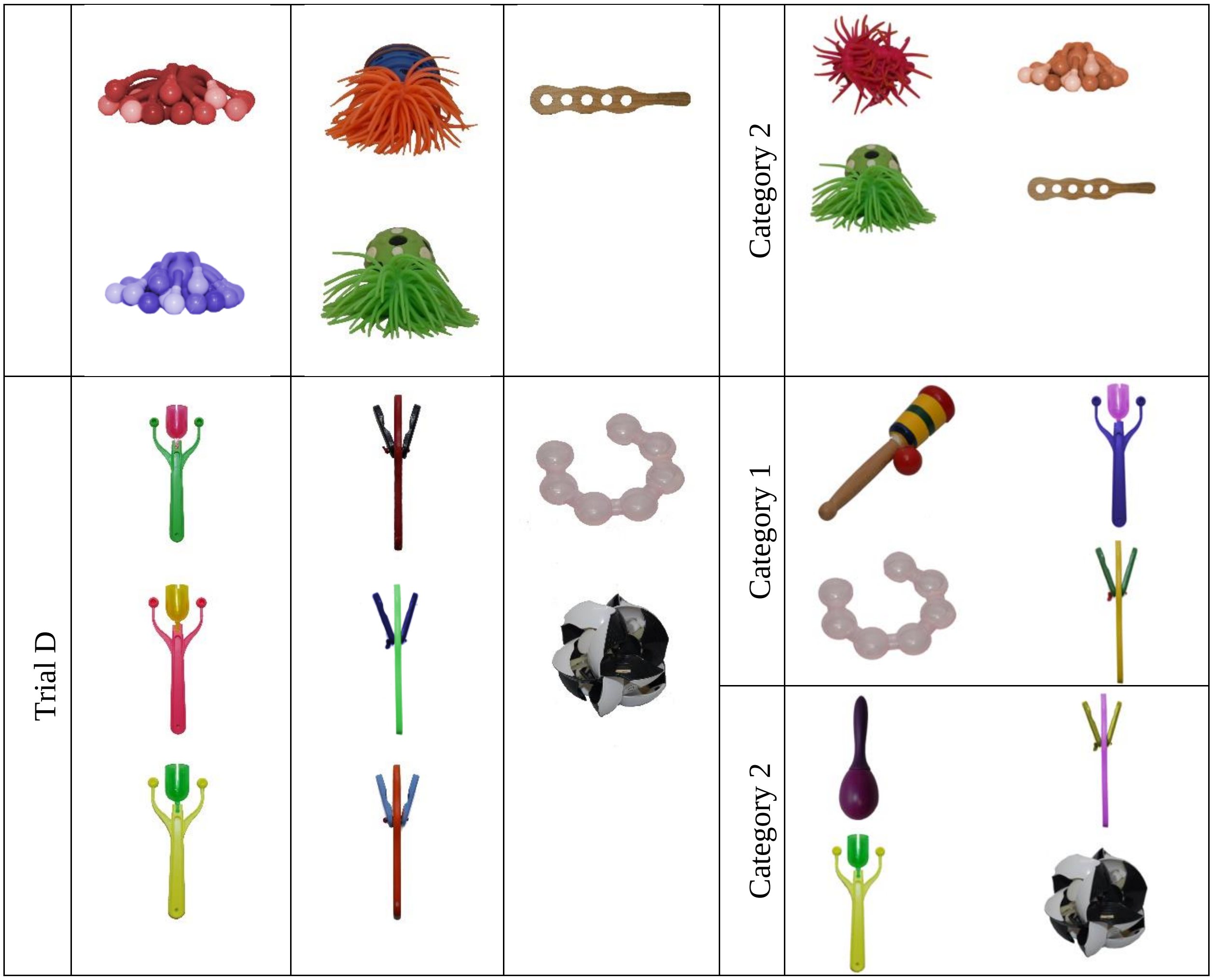



Trial D
Category 1
Category 2
Category 2

Appendix B.

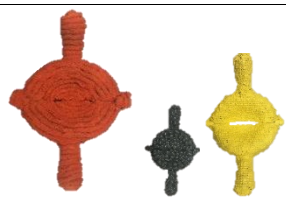
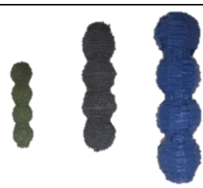
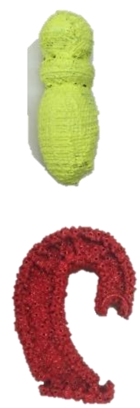
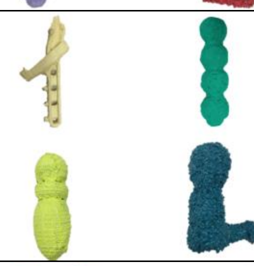
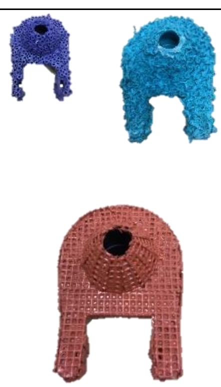
Study 1 Low Between-Category Variability Stimuli

	Category 1 Objects	Category 2 Objects	Distractor Objects		Tests
Trial A				Category 1	
				Category 2	
Trial B				Category 1	
				Category 2	
Trial C				Category 1	


Trial D
Category 2
Category 1
Category 2

Appendix C.

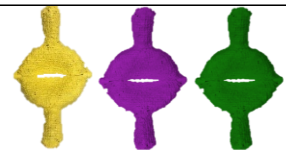
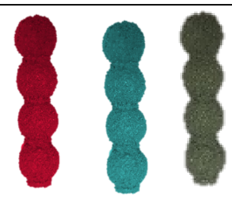
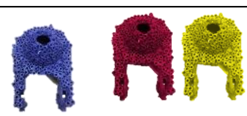
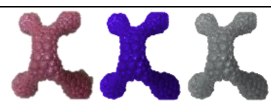
Study 2 High Within-Category Variability Stimuli

	Category 1 Objects	Category 2 Objects	Distractor Objects		Tests
Trial A				Category 1	
				Category 2	
Trial B				Category 1	
				Category 2	
Trial C				Category 1	
				Category 2	


Trial D
Category 1
Category 2

Appendix D.

Study 2 Low Within-Category Variability Stimuli

	Category 1 Objects	Category 2 Objects	Distractor Objects		Tests
Trial A				Category 1	
				Category 2	
Trial B				Category 1	
				Category 2	
Trial C				Category 1	
				Category 2	


Trial D
Category 1
Category 2

References

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416–430. https://doi.org/https://doi.org/10.1111/1467-8624.00287

Althaus, N., & Plunkett, K. (2016). Categorization in infancy: labeling induces a persisting focus on commonalities. *Developmental Science*, *19*(5), 770–780. https://doi.org/10.1111/desc.12358

Anderson, E. M., Chang, Y. J., Hespos, S., & Gentner, D. (2018). Comparison within pairs promotes analogical abstraction in three-month-olds. *Cognition*, *176*(April 2017), 74–86. https://doi.org/10.1016/j.cognition.2018.03.008

Andrews, J. K., Kurtz, K. J., & Livingston, K. R. (2005). Improving category learning through the use of context items: Compare or contrast? *Proceedings of the 27th …*. http://csjarchive.cogsci.rpi.edu/proceedings/2005/docs/p121.pdf

Ankowski, A. A., Vlach, H. A., & Sandhofer, C. M. (2013). Comparison versus contrast: Task specifics affect category acquisition. *Infant and Child Development*, *3*, 1–23. https://doi.org/10.1002/icd

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth - children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129. https://doi.org/10.1080/15475441.2010.505489

Atagi, N., Goldenberg, E. R., & Sandhofer, C. M. (2016). Children's use of linguistic information when learning in a bilingual context. *Journal of Experimental Child Psychology*, *144*, 199–208. https://doi.org/10.1016/j.jecp.2015.11.005

Benitez, V. L., Zettersten, M., & Wojcik, E. (2020). The temporal structure of naming events differentially affects children's and adults' cross-situational word learning. *Journal of Experimental Child Psychology*, *200*, 1–18. https://doi.org/10.1016/j.jecp.2020.104961

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*, 392–402. https://doi.org/10.3758/s13421-012-0272-7

Bjork, E. L., & Bjork, R. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, *56*(64), 1–13.

Bjork, R. A. (2011). On the symbiosis of learning, remembering, and forgetting. In *Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork* (pp. 1–21). Psychology Press.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differrential encoding? *Journal of Verbal Learning and Verbal Behavior*, *9*, 567–572.

Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *American Journal of Psychology*, *128*(2), 241–252. https://doi.org/10.5406/amerjpsyc.128.2.0241

Blaye, A., & Jacques, S. (2009). Categorical flexibility in preschoolers: Contributions of conceptual knowledge and executive control. *Developmental Science*, *12*(6), 863–873. https://doi.org/10.1111/j.1467-7687.2009.00832.x

Callanan, M. A. (1985). How Parents Label Objects for Young Children: The Role of Input in the Acquisition of Category Hierarchies. *Child Development*, *56*(2), 508–523.

Carey, S., & Bartlett, E. (1978). Acquiring a Single New Word. *Papers and Reports on Child*

Language Development, *15*, 17–29.

Carvalho, P. F., Chen, C., & Yu, C. (2021). The distributional properties of exemplars affect category learning and generalization. *Scientific Reports*, *11*, 1–11. https://doi.org/10.1038/s41598-021-90743-0

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*, 481–495. https://doi.org/10.3758/s13421-013-0371-0

Carvalho, P. F., & Goldstone, R. L. (2015a). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, *22*, 281–288. https://doi.org/10.3758/s13423-014-0676-4

Carvalho, P. F., & Goldstone, R. L. (2015b). What you learn is more than what you see : what can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6*(April), 1–12. https://doi.org/10.3389/fpsyg.2015.00505

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(11), 1699–1719. https://doi.org/10.1037/xlm0000406

Childers, J. B., Paik, J. H., Flores, M., & Lai, G. (2017). Does variability across events affect verb learning in English, Mandarin, and Korean? *Cognitive Science*, *41*, 808–830. https://doi.org/10.1111/cogs.12398

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*(6), 967–978. https://doi.org/10.1037/0012-1649.38.6.967

Deng, W. (Sophia), & Sloutsky, V. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, 24–62. https://doi.org/10.1016/j.cogpsych.2016.09.002

Dewar, K., & Xu, F. (2007). Do 9-month-old infants expect distinct words to refer to kinds? *Developmental Psychology*, *43*(5), 1227–1238. https://doi.org/10.1037/0012-1649.43.5.1227

Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science*, *20*(2), 252–257. https://doi.org/https://doi.org/10.1111/j.1467-9280.2009.02278.x

Dixon, W. E., Salley, B. J., & Clements, A. D. (2006). Temperament, distraction, and learning in toddlerhood. *Infant Behavior and Development*, *29*(3), 342–357. https://doi.org/10.1016/j.infbeh.2006.01.002

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*, 903–917. https://doi.org/https://doi.org/10.1111/j.1467-8624.1994.tb00792.x

Estes, W. K., & Burke, C. J. (1953). A theory of stimulus variability in learning. *Psychological Review*, *60*(4), 276–286. https://doi.org/https://doi.org/10.1037/h0055775

Fanselow, M. S., & Tighe, T. J. (1988). Contextual Conditioning With Massed Versus Distributed Unconditional Stimuli in the Absence of Explicit Conditional Stimuli. *Journal* *of Experimental Psychology: Animal Behavior Processes*, *14*(2), 187–199. https://doi.org/10.1037/0097-7403.14.2.187

Flack, Z. M., & Horst, J. S. (2018). Two sides to every story: Children learn words better from one storybook page at a time. *Infant and Child Development*, *27*(1), 1–12. https://doi.org/10.1002/icd.2047

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9- and 15-month-olds' object categorization. *Infancy*, *4*(3), 349–369. https://doi.org/10.1207/S15327078IN0403_03

Fulkerson, A. L., & Haaf, R. A. (2006). Does object naming aid 12-month-olds' formation of novel object categories? *First Language*, *26*(4), 347–361. https://doi.org/10.1177/0142723706059217

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought, and culture* (pp. 301–334). Lawrence Erlbaum Associates.

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive*

Development, *14*(4), 487–513. https://doi.org/10.1016/S0885-2014(99)00016-7

Goldenberg, E. R., & Sandhofer, C. M. (2013). Same, varied, or both? Contextual support aids young children in generalizing category labels. *Journal of Experimental Child Psychology*, *115*(1), 150–162. https://doi.org/10.1016/j.jecp.2012.11.011

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Children and Adults Use Lexical Principles to Learn New Nouns. *Developmental Psychology*, *28*(1), 99–108. https://doi.org/https://doi.org/10.1037/0012-1649.28.1.99

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-Month-Olds Rely on Shared Labels and Shape Similarity for Inductive Inferences. *Child Development*, *75*(2), 409–427. https://doi.org/https://doi.org/10.1111/j.1467-8624.2004.00683.x

Graham, S. A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology*, *107*(3), 280–290. https://doi.org/10.1016/j.jecp.2010.04.017

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, *33*(2), 289–302.

Hammer, R. (2015). Impact of feature saliency on visual category learning. *Frontiers in Psychology*, *6*(MAR), 1–6. https://doi.org/10.3389/fpsyg.2015.00451

Henderson, A. M. E., & Woodward, A. L. (2012). Nine-month-old infants generalize object labels, but not object preferences across individuals. *Developmental Science*, *15*(5), 641–652. https://doi.org/10.1111/j.1467-7687.2012.01157.x

Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to compare for what. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*(33), 1388–1393.

Horst, J. S. (2013). Context and repetition in word learning. *Frontiers in Psychology*, *4*(April), 1–11. https://doi.org/10.3389/fpsyg.2013.00149

Horst, J. S., & Hout, M. C. (2014). Novel object & unusual name (NOUN) database: A collection of novel images for use in experimental research. *Unpublished Manuscript*.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*, 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of Comprehension, Competence, and Remembering. In *Learning, Remembering, Believing: Enhancing Individual and Team Performance* (pp. 57–80).

Jing, M., & Kirkorian, H. (2020). Video deficit in children's early learning. *The International Encyclopedia of Media Psychology*, 1–8. https://doi.org/10.1002/9781119011071.iemp0239

Johanson, M., & Papafragou, A. (2016). The influence of labels and facts on children's and adults' categorization. *Journal of Experimental Child Psychology*, *144*, 130–151. https://doi.org/10.1016/j.jecp.2015.11.010

Jones, L. B., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental Science*, *6*(5), 498–504. https://doi.org/https://doi.org/10.1111/1467-7687.00307

Kang, S. H. K., & Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. https://doi.org/10.1002/acp.1801

Katerelos, M., & Poulin-Dubois, D. (2011). A cross-linguistic study of word-mapping in 18- to 20-month-old infants. *Infancy*, *16*(5), 508–534. https://doi.org/10.1111/j.1532-7078.2010.00064.x

Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General*, *137*(1), 52–72. https://doi.org/10.1037/0096-3445.137.1.52

Knabe, M. L., & Vlach, H. A. (2020). When are difficulties desirable for children? First steps toward a developmental and individual differences account of the spacing effect. *Journal of Applied Research in Memory and Cognition*, *9*(4), 447–454. https://doi.org/10.1016/j.jarmac.2020.07.007

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the '"Enemy of Induction"'? *Psychological Science*, *19*(6), 585–592. https://doi.org/https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kotovsky, L., & Gentner, D. (1996). Comparison and Categorization in the Development of Relational Similarity. *Child Development*, *67*(6), 2797–2822. https://doi.org/10.2307/1131753

Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorization. *Journal of Experimental Child Psychology*, *98*(2), 69–93. https://doi.org/https://doi.org/10.1016/j.jecp.2007.05.001

Kurtz, K. J., & Boukrina, O. (2004). Learning relational categories by comparison of paired examples. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 756–761.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The Importance of Shape in Early Lexical Learning. *Cognitive Development*, *3*, 299–321. https://doi.org/https://doi.org/10.1016/0885-2014(88)90014-7

LaTourrette, A. S., & Waxman, S. R. (2020). Naming guides how 12-month-old infants encode and remember objects. *Proceedings of the National Academy of Sciences*, *117*(35), 21230–21234. https://doi.org/10.1073/pnas.2006608117

Lewis, M. L., & Frank, M. C. (2018). Still Suspicious: The Suspicious- Coincidence Effect Revisited. *Psychological Science*, *29*(12), 2039–2047. https://doi.org/10.1177/0956797618794931

Linde, E. Vander, Morrongiello, B. A., & Rovee-Collier, C. (1985). Determinants of Retention in 8-Week-Old Infants. *Developmental Psychology*, *21*(4), 601–613. https://doi.org/10.1037/0012-1649.21.4.601

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

Maguire, M. J., Hirsh-pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628–634. https://doi.org/10.1111/j.1467-7687.2008.00707.x

Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats:

Differentiating global categories. *Cognitive Psychology*, *23*(2), 263–298.

Mervis, C. B., & Crisafi, M. A. (1982). Order of Acquisition of Subordinate-, Basic-, and Superordinate-Level Categories. *Child Development*, *53*(1), 258–266.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89–115.

Monaghan, P., Brand, J., Frost, R. L. A., & Taylor, G. (2017). Multiple variable cues in the environment promote accurate and robust word learning. *Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)*, 817–822. https://mindmodeling.org/cogsci2017/papers/0164/index.html

Nagle, K. J., & O'brien, M. (1987). Parents' speech to toddlers: The effect of play context. *Journal of Child Language*, *14*(2), 269–279. https://doi.org/10.1017/S0305000900012927

Nakata, T., & Elgort, I. (2021). Effects of spacing on contextual vocabulary learning: Spacing facilitates the acquisition of explicit, but not tacit, vocabulary knowledge. *Second Language Research*, *37*(2), 233–260. https://doi.org/10.1177/0267658320927764

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902. https://doi.org/10.1177/0956797610389189.Learn

Price, G. F., & Sandhofer, C. M. (2021). One versus many: Multiple examples in word learning. *Journal of Experimental Child Psychology*, *209*, 1–19. https://doi.org/10.1016/j.jecp.2021.105173

Quinn, P. C., & Bhatt, R. S. (2010). Learning perceptual organization in infancy: The effect of simultaneous versus sequential variability experience. *Perception*, *39*, 795–807. https://doi.org/10.1068/p6639

Raz, H. K., Abney, D. H., Crandall, D., Yu, C., & Smith, L. B. (2019). How do infants start learning object names in a sea of clutter? *CogSci ... Annual Conference of the Cognitive Science Society. Cognitive Science Society (U.S.). Conference*, *2019*, 521–526.

Rea, C. P., & Modigliani, V. (1987). The spacing effect in 4- to 9-year-old children. *Memory & Cognition*, *15*(5), 436–443.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*, 665–681. https://doi.org/https://doi.org/10.1016/S0022-5371(75)80055-7

Rosch, E. (1978). Principles of categorization. In *Cognition and categorization* (pp. 27–48). Lawrence Erlbaum Associates.

Rost, G. C., & Mcmurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 1–3. https://doi.org/10.1111/j.1467-7687.2008.00786.x.Speaker

Rost, G. C., & Mcmurray, B. (2010). Finding the Signal by Adding Noise : The Role of Noncontrastive Phonetic Variability in Early Word Learning. *Infancy*, *15*(6), 608–635. https://doi.org/10.1111/j.1532-7078.2010.00033.x

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*, 833–870. https://doi.org/10.1016/j.cognition.2007.05.002

Slone, L. K., & Sandhofer, C. M. (2017). Consider the category: The effect of spacing depends on individual learning histories. *Journal of Experimental Child Psychology*, *159*, 34–49. https://doi.org/10.1016/j.jecp.2017.01.010

Sloutsky, V. M., Sophia Deng, W., Fisher, A. V., & Kloos, H. (2015). Conceptual influences on induction: A case for a late onset. *Cognitive Psychology*, *82*, 1–31. https://doi.org/10.1016/j.cogpsych.2015.08.005

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*(1), 13–19. https://doi.org/10.1111/1467-9280.00403

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

Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition*, *108*, 626–638. https://doi.org/10.1016/j.cognition.2008.05.002

Sorensen, L. J., & Woltz, D. J. (2016). Blocking as a friend of induction in verbal category learning. *Memory and Cognition*, *44*(7), 1000–1013. https://doi.org/10.3758/s13421-016-0615-x

Storm, B. C. (2011). The Benefit of Forgetting in Thinking and Remembering. *Current Directions in Psychological Science*, *20*(5), 291–295. https://doi.org/10.1177/0963721411418469

Taxitari, L., Twomey, K. E., Westermann, G., & Mani, N. (2020). The Limits of Infants' Early

Word Learning. *Language Learning and Development*, *16*(1), 1–21. https://doi.org/10.1080/15475441.2019.1670184

Thibaut, J., & Witt, A. (2015). Young children's learning of relational categories: Multiple comparisons and their cognitive constraints. *Frontiers in Psychology*, *6*(May), 1–10. https://doi.org/10.3389/fpsyg.2015.00643

Toppino, T. C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition*, *19*(2), 159–167.

Twomey, K. E., Ranson, S. L., & Horst, J. S. (2014). That's more like it: Multiple exemplars facilitate word learning. *Infant and Child Development*, *122*(October 2013), 105–122. https://doi.org/10.1002/icd.1824

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(1), 246–254. https://doi.org/10.1037/a0025260

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382. https://doi.org/10.1016/j.cognition.2013.02.015

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*(4), 1137–1144. https://doi.org/10.1111/j.1467-8624.2012.01781.x

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology*, *36*(5), 571–581. https://doi.org/10.1037//O012-1649.36.5.571

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257–302. https://doi.org/10.1006/cogp.1995.1016

Wojcik, E. H. (2013). Remembering new words: Integrating early memory development into word learning. *Frontiers in Psychology*, *4*, 1–12. https://doi.org/10.3389/fpsyg.2013.00151

Yoshida, H., & Smith, L. B. (2005). Linguistic Cues Enhance the Learning of Perceptual Cues. *Psychological Science*, *16*(2), 90–95. https://doi.org/https://doi.org/10.1111/j.0956-7976.2005.00787.x

Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *The Future of Children*, *26*(2), 21–35. https://www.jstor.org/stable/43940579%0AJSTOR

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*, 16–27. https://doi.org/10.3758/s13421-012-0238-9