Statistical/Probabilistic Learning Techniques in Developmental Psychology

Alyson Wong[1] and Fei Xu[1]

Department of Psychology, University of California, Berkeley

In press, Oxford Research Encyclopedia of Psychology

Summary

The world around us is a complex environment filled with useful data for learning as well as a lot of noise. In a simple, common setting such as a living room at home, there is a large amount of information. For example, there is visual information that includes the shapes and colors of furniture and movement of other people in the household, as well as auditory information such as parents talking to each other, the television playing, or birds chirping outside. The learner faces the challenging task of extracting relevant information in order to make sense of the world. This learning problem is especially acute for infants who have had very little exposure to their environment: how do they filter out irrelevant information and form representations of the world so quickly? For example, when infants hear a continuous stream of sounds from their parents with little knowledge of how language works, how are they able to parse that stream into words and map meanings onto these words? When a child sorts their toys into groups of animals, how do they know what perceptual information defines each category so they can form the categories and generalize to new animal toys? Substantial research has strived to answer these questions, providing evidence that from infancy on, many statistical learning mechanisms work in tandem to help young learners learn about the world. Statistical learning in development often refers to the implicit process of extracting structure from the environment. The research literature reviewed here explores statistical learning purely from exposure or observation without any explicit instruction, feedback, or reward. Over the past several decades, much work has explored a variety of different robust statistical learning mechanisms, each involving different and distinct computations. These mechanisms enable quick, efficient learning in a variety of different domains, such as category learning, language learning, causal learning, and learning about people and their preferences and goals. In this review, statistical learning and statistical inference mechanisms throughout the course of development, with a focus on the first few years of life, are discussed.

Keywords

Statistical learning; correlation learning; probabilistic inference; non-adjacent dependencies; transitional probabilities

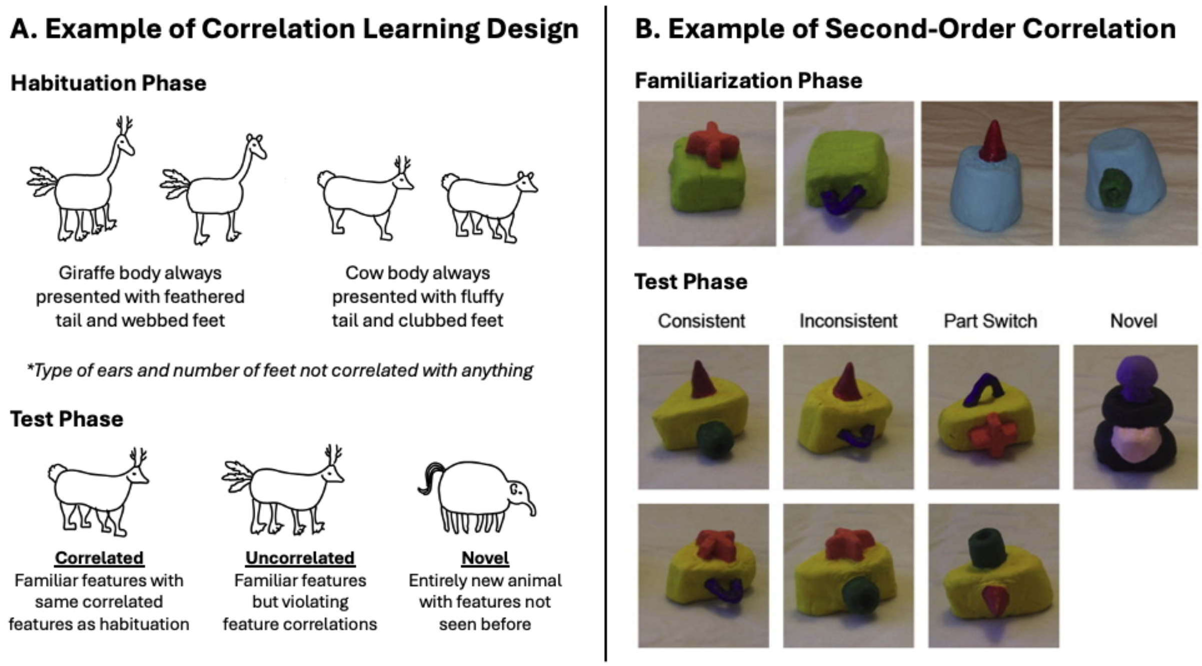
Correlation Learning

Many studies have explored one important statistical learning mechanism in infancy and early childhood – correlation learning. Historically, correlation learning was the first statistical learning mechanism explored across development. Most work on correlation learning focuses on simultaneous feature correlations, or that the presence of one feature simultaneously coincides with the presence of one or more other features, i.e., *first-order correlations* (Younger & Cohen, 1983). For example, animals with feathers typically also have beaks and wings, whereas animals with fur typically do not have those attributes. This indicates that the features of feathers, beaks, and wings are more highly correlated than fur, beaks, and wings. Love et al. (2004) provide a model that well fits category learning mechanisms based on correlations in infants and adults. Several studies have explored whether infants are able to form categories based on detecting feature correlations.

Between 4 1/2 and 7 months of age, infants begin to display an ability to represent correlations under specific circumstances (Mareschal et al., 2005; Younger & Cohen, 1986). By 10 months of age, infants are reliably able to perceive basic correlations between visual features. Younger and Cohen (1983) presented infants with various drawings of animals with several visual features that were correlated, such as the type of body, tail, and feet, as well as several features that were not correlated. When presented with novel animal exemplars with combinations of the same features as presented previously that either preserved previous feature correlations, violated previous correlations, or had completely novel features, thus serving as a test of generalization of these feature correlations and not purely recognition of the exact same stimuli, 10-month-old infants looked longer at animals that violated previous correlations or had completely novel features than to animals that preserved previous correlations, though 4- and 7-

month-olds failed to distinguish between animals that preserved previous correlations and animals that violated those correlations.

Figure 1. Depiction of First- and Second-Order Correlation Learning Tasks



Note: A) Depiction of design from Younger & Cohen (1986) correlation learning paradigm. B) Depiction of Yermolayeva & Rakison (2016) second-order correlation paradigm.

Not only are infants able to perceive correlations, but infants' correlational learning abilities are quite robust. By 10 months of age, infants are able to use correlational information to form categories (Younger, 1985) and are also sensitive to these correlations among features that may be useful in forming natural object categories, such as different types of animal tails and ears or antlers in photographs instead of drawings (Younger, 1990). Additionally, infants by 14 months are able to recognize form-function correlations (Madole & Cohen, 1995), such that, when shown objects that adhere to form-function correlations, such as an object with rubber black wheels that always function and spin whereas an object with yellow plastic wheels that never function, 14-month-old infants looked longer at a new object with rubber black wheels that were nonfunctional. Between 14 and 18 months, infants were able to detect correlations between different kinds of dynamic (e.g., two objects producing two distinct patterns of motion) and static features (Rakison & Poulin-Dubois, 2002).

Most recently, some research has suggested that infants can learn even more complex correlations, such as correlations between features that, in contrast to first-order correlations where features are presented simultaneously, are *not* observed simultaneously, commonly referred to as *second-order correlations*. Yermolayeva and Rakison (2016) familiarized seven- and 11-month-old infants to numerous 3D objects correlating the body of an object with external parts. Most importantly, multiple external parts correlated with the same body are never seen together during these trials, such as Part A appearing on Body 1 and Part B appearing on Body 1, but Part A and B never appearing on Body 1 simultaneously. During the test phase, multiple external parts were then presented simultaneously on a novel body, with some objects containing external parts that were previously correlated with the same body (e.g. Part A and Part B on Body 3 simultaneously) and some objects containing external parts that were previously correlated with different bodies (e.g. Part A from Body 1 and Part C from Body 2 appearing on Body 3). Both 7- and 11-month-olds (but not 9-month-olds) showed evidence of learning these second-order correlations by looking longer at objects with uncorrelated parts than objects with correlated parts.

Around 20 months of age, infants can learn second-order correlations between static and dynamic features, but only in category contexts (Rakison & Benton, 2019). That is, when correlating two static features (e.g. various red objects and a white cross) and separately correlating a static and dynamic feature (e.g. red objects move rectilinearly), infants looked longer when viewing a novel object that violates those correlations (e.g. a pink object with a white cross moving curvilinearly) compared to when viewing a novel object that preserves those correlations (e.g. a pink object with a white cross moving rectilinearly). However, infants struggle with this task when objects are presented in a non-category context, i.e., when presented with only one exemplar of a static and dynamic feature correlation.

Later in development, children are also able to learn second-order correlations between an objects' features and their causal efficacy. Benton et al. (2021) explored whether children are able to do this using a blicket-detector paradigm. In this paradigm, children are presented with a blicket-detector that lights up and plays music when either a specific object or specific combinations of objects are placed on the machine, thus establishing that some objects have causal powers to start the machine whereas others do not. Benton et al. (2021) presented two- and three-year-old children with two differently colored and shaped blocks (e.g. red cube and green cylinder), each with a different sticker placed on the front (e.g. yellow circle or purple diamond). Separately, children were also shown two new blocks as seen previously but without the stickers and demonstrated that only one of the blocks made a blicket detector machine light up and play music. When presented with two new blocks only differing in the attached stickers (i.e. yellow circle or purple diamond presented initially), both two- and -three-year-olds were able to correctly learn the second-order correlation and identify the block that had causal powers, provided that they correctly remembered the initial sticker-block pairings.

Transitional Probabilities

Statistical learning has often been considered synonymous with the detection and usage of transitional probabilities, an important mechanism that has been studied extensively. In contrast to correlation learning, where correlated elements are typically presented simultaneously, transitional probabilities typically refer to the probability between *successive* elements (Saffran et al., 1996), such as syllables in language. The computation for transitional probabilities can be defined as

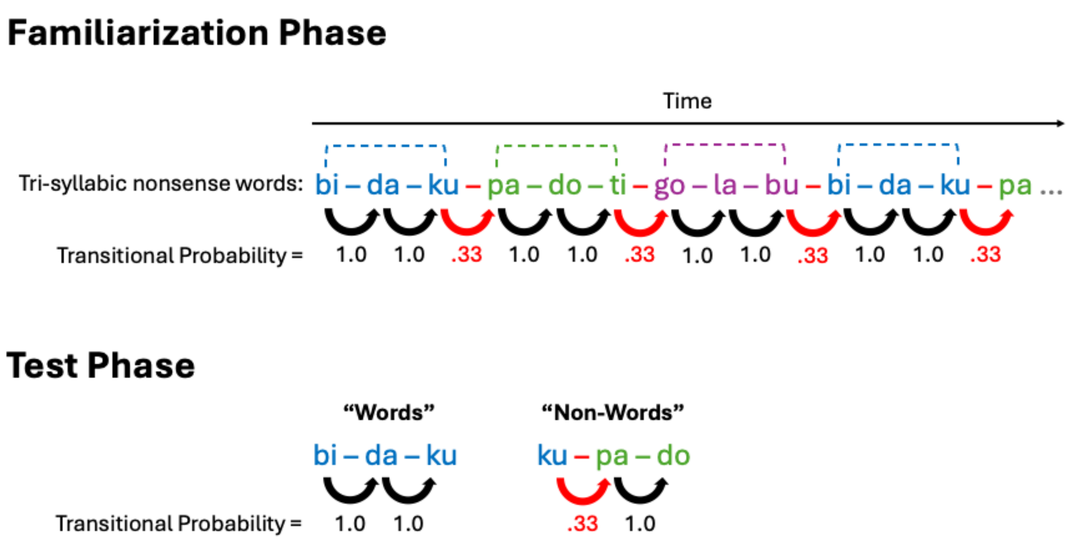
$$probability\ of\ Y|X\ =\ (frequency\ of\ XY)\ /\ (frequency\ of\ X)$$

These probabilities exist in language, such that some syllables are more likely to occur together *within* words, and thus have a higher transitional probability, than *across* word boundaries, which have lower transitional probabilities. For example, in a phrase such as *fluffy#bunny*, the co-occurrence of syllables within each word, such as *flu* and *fy* or *bu* and *ny*, have greater transitional probabilities than syllable combinations that span across word boundaries, such as *fy#bu*. In the last few decades, a rich body of research has explored the ability to extract statistical patterns based on temporal transitional probabilities throughout development in language and a variety of other domains (see Saffran & Kirkham, 2018 for a review).

Transitional Probabilities in Language Learning

Word Segmentation. Some of the first work on statistical learning explored word segmentation in the domain of language. A study by Saffran et al. (1996) investigated whether infants could use transitional probabilities to segment words in a continuous stream of speech. Infants heard continuous streams of speech consisting of three-syllable nonsense words (e.g. *bi-da-ku*, *pa-do-ti*). The transitional probabilities of syllables varied, such that syllables within the same nonsense word formed pairs that always occurred together (e.g. *bi-da, da-ku*) and thus had a transitional probability of 1.0, whereas other syllable pairs spanning across nonsense words only sometimes occurred together (e.g. *ku-pa*) and thus had a lower transitional probability. After listening to this continuous speech stream, 8-month-old infants listened longer to lower-transitional probability triplets and completely novel syllable triplets that never occurred together than the higher-transitional probability triplets, suggesting they were able to distinguish between 'words' and 'non-words' based on transitional probabilities. Furthermore, Aslin et al. (1998) confirmed infants' abilities to use transitional probabilities in word segmentation and ruled out a confounding variable of frequency of syllable triplets. Additionally, Pelucchi et al. (2009a) found that English-learning 8-month-olds are also sensitive to transitional probabilities from naturalistic Italian speech. Estes and Lew-Williams (2015) found that 7- to 8-month-old and 10- to 11-month-old infants were able to segment words based on transitional probabilities when there were multiple speakers. Another study tested newborn infants in a similar paradigm using EEG and found that sleeping 3-day-olds can distinguish between 'words' and 'non-words' based on their transitional probabilities (Flo et al. 2022).

Fiigure 2. Example of Saffran et al. (1996) Transitional Probability Manipulation



Other studies have explored the use of different types of transitional probabilities. Perruchet and Desaulty (2008) explored the role of backwards transitional probabilities on word segmentation. Adult participants listened to structured continuous speech streams consisting of

12 different syllables, with streams varying such that either backwards transitional probabilities within and across syllable pairs were kept constant (e.g. syllables A-C were always followed by syllable X, transitional probability of A, B, or C preceding X is .33) and forwards probabilities varied, or forwards transitional probabilities were kept constant (e.g. X always followed by A-C, transitional probability of A following X is .33) and backwards probabilities varied. Participants who listened to the speech stream varying backwards transitional probabilities were better able to distinguish high from low transitional probability pairs than participants who heard the speech stream varying forwards transitional probabilities. Similar patterns of results have also been found with infants: 8-month-old infants also displayed the ability to segment words based on backwards transitional probabilities alone (Pelucchi et al., 2009b).

<u>Limitations in Word Segmentation.</u> Although statistical learning is a powerful mechanism for word segmentation, several studies have also found constraints and limitations on statistical learning abilities. For example, the complexity of the language and structure may hinder statistical learning. When varying the number of syllables per nonsense word, such as some words containing two syllables and others containing three, both 5.5- and 8-month-old infants failed to show evidence of proper word segmentation (Johnson & Tyler, 2010). Additionally, word segmentation based on statistical learning may be restricted by the number of syllables per word. Benjamin et al. (2023) explored word segmentation in adults and neonates using four-syllable nonsense words instead of tri- or di-syllabic words. Despite previous studies demonstrating successful word segmentation based on transitional probabilities in adults and neonates (Flo et al., 2022), both age groups failed to distinguish previously heard words from part-words (words spanning boundaries) or shuffle-words (words with the order of syllables shuffled). However, adults' performance increased when pauses between words when listening

to the initial continuous stream could be used as a facilitating cue, though neonates' performance did not improve.

<u>Role in Language Acquisition.</u> Although there may be limitations on statistical learning, it is clear that even from a young age, infants are sensitive to transitional probabilities of syllables in language. However, transitional probabilities and statistical learning seems to play a differing role across the course of language acquisition.

Previous findings suggest that the use of statistical learning may be critical for language acquisition specifically. Saffran (2001) first habituated infants to continuous speech streams consisting of tri-syllabic nonsense words. Infants were then presented with either English or entirely nonsense sentences ending with either a high-transitional probability nonsense word or a low-transitional probability nonsense word heard in habituation. Infants who were presented with English sentences listened longer when sentences were completed with high-transitional probability nonsense words than low-transitional probability words, whereas infants who heard entirely nonsense sentences listened equally to both types of words. These findings suggest that infants' sensitivity to statistical properties in speech are useful for language acquisition and not merely for detecting properties of sound sequences.

Other work has extended these findings to explore whether meanings can be mapped onto words segmented through statistical learning. A study by Estes et al. (2007) first presented 17-month-olds with continuous speech streams consisting of two-syllable nonsense words varying in transitional probabilities. Then, infants were habituated to novel 3D objects paired with two verbal labels comprised of either high transitional probability, low transitional probability, or completely novel syllable pairs from the continuous stream. During the test phase, only infants who were presented with high-transitional probability word labels for objects looked significantly longer when object-label pairings swapped compared to when they were consistent with habituation, suggesting infants may have more ease mapping meanings onto words with stronger transitional probabilities. However, this advantage of high transitional probabilities on mapping meanings to words may be short-lived. A study by Lany et al. (2024) found that by 22 months of age, infants no longer displayed a difference in mapping word meanings onto high- versus low-transitional probability words. Additionally, they found that infants with smaller parent-reported vocabularies continued to demonstrate a high-transitional probability advantage, whereas infants with larger vocabularies learned low- and high-transitional probability words similarly.

Transitional probabilities in other domains

Statistical learning using transitional probabilities is not only limited to language learning. This mechanism can also be used to segment other domains, such as non-linguistic auditory sequences. Saffran et al. (1999) found that both adults and infants could successfully segment continuous streams of musical tone sequences using transitional probabilities. Jonaitis and Saffran (2009) additionally found that adults could learn statistical regularities of novel musical styles that had never been heard before.

Furthermore, many studies with adults have investigated visual statistical learning abilities. Fiser and Aslin (2001) presented adults with scenes consisting of base pairs of two shapes with high internal transitional probabilities (shown simultaneously), whereas the transitional probabilities between shapes across base pairs were weaker. They found that after this training exposure, adults rated high transitional probability pairs as more familiar than pairs with lower transitional probabilities when frequency was controlled for. Additionally, Fiser & Aslin (2002a) familiarized adults with triplets of shapes shown sequentially, like in word segmentation statistical learning tasks. They found that adults reliably selected the triplets with high transitional probabilities as more familiar compared to the triplets with lower transitional probabilities.

In addition to using statistical learning to segment visual sequences, transitional probabilities are also used to segment continuous action sequences. Using the general methods of Saffran et al. (1996), Baldwin et al. (2008) found that adults used transitional probabilities in judging the familiarity of action sequences. Stahl et al. (2014) found that 7- to 9-month-old infants also used transitional probabilities to segment continuous visual action events. Monroy et al. (2019) found neural evidence of infants' abilities to segment continuous visual action sequences using transitional probabilities, finding increased motor activity in 18-month-olds in anticipation only of familiar action sequences but not random sequences, suggesting infants seem to predict future actions based on learned statistical sequences. Monroy et al. (2017) found that 19-month-olds can use their learned knowledge from visual statistical learning of action sequences to perform those sequences themselves when playing with a toy requiring specific action sequences to produce causal effects (e.g. lights and music).

Some work has also explored statistical learning in the tactile domain. Conway and Christiansen (2005) explored tactile statistical learning, finding that adults segmented vibration sequences using transitional probabilities. Interestingly, when conducting the same experiment but converting the sequences from the tactile domain to visual and auditory domains, they found that participants performed better in the auditory task than in the tactile or visual tasks, and performance did not differ between the latter two.

Changes Across Development

Although much work suggests statistical learning abilities emerge very early in development, other work suggests that there may be differences in sensitivity to transitional probabilities across development, particularly in the visual domain. Kirkham et al. (2002) first presented 2-, 5-, and 8-month-old infants with a continuous stream of single shapes with shape pairs defined only by their transitional probabilities. When presented with either high transitional probability pairs in isolation or lower transitional probability pairs, they found that all three age groups looked significantly longer at the low-transitional probability pairs of shapes. Thus, by two months of age, infants show a sensitivity to transitional probabilities in visual sequences. Fiser and Aslin (2002b) conducted an additional study with 9-month-old infants, finding sensitivity to transitional probabilities when frequency of pairs was controlled for. In addition to using visual shapes, Mermier et al. (2022) found that 12-month-old infants could extract statistical regularities from sequences of faces expressing various emotions.

However, infants' visual statistical abilities may still be developing throughout this time. Both Kirkham et al. (2002) and Fiser and Aslin (2002b) only measured infants' statistical learning abilities when manipulating either temporal or spatial dimensions, never both. Kirkham et al. (2007) found that when objects appeared in a continuous stream of three structured location pairs (e.g. circle appeared on top middle of 2x3 grid, always followed by top left), only 11-month-old infants, not 8-month-olds, looked longer at a novel sequence than a familiar sequence. When color and shape were added as an aiding cue, 8-month-olds, but not 5-month-olds, displayed longer looking times to novel sequences. Furthermore, a study by Marcovitch and Lewkowicz (2009) showed that 4.5- and 8.5-month-old infants, but not 2.5-month-olds, were able to distinguish between pairs of shapes with differing frequency but similar transitional probabilities as well as pairs with differing transitional probabilities but similar frequencies.

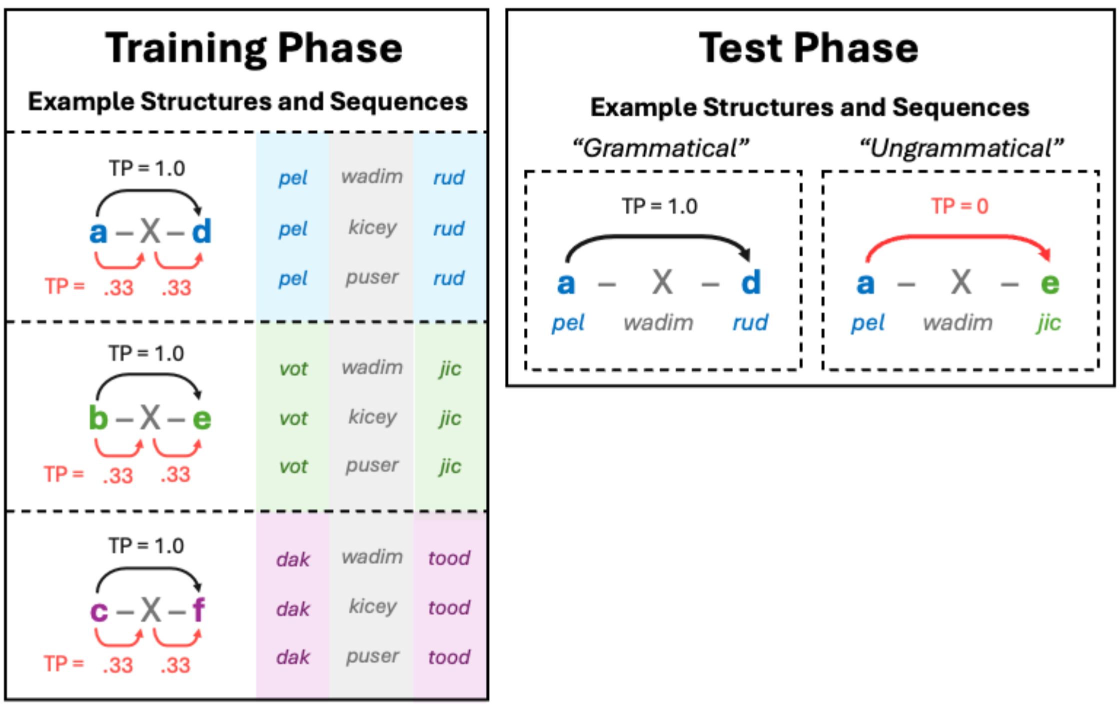
Lastly, Bulf et al. (2011) tested newborn infants and found they could compute transitional probabilities with two but not three pairs of visual shapes. Therefore, even newborns have some sensitivity to transitional probabilities in the visual domain, though their abilities are quite limited.

Non-Adjacent Dependencies

In addition to the large body of work on transitional probabilities between different types of items that are adjacent to one another, there has also been much work exploring transitional probabilities of non-adjacent elements. Although work on transitional probabilities typically refers to the probability of an element occurring given an immediately preceding or successive element (i.e. adjacent to one another), transitional probabilities also exist between elements that are non-adjacent. For example, in language, these probabilities often occur between auxiliaries and verb suffixes, such that the co-occurrence between the word *is* in English with the verb suffix *-ing* is frequent and thus has a relatively high transitional probability. However, unlike with the previously discussed transitional probabilities, these elements are separated by an intermediary element that varies and is unrelated to the transitional probability between the non-adjacent elements (e.g. <u>*is*</u> *eat<u>ing</u>*, <u>*is*</u> *walk<u>ing</u>*). Various studies have explored this type of transitional probability, typically referred to as non-adjacent dependencies (see Wilson et al. 2020 for a review).

A study by Gomez (2002) explored infants' and adults' abilities to learn non-adjacent dependencies within a continuous stream of artificial language. Infants and adults were first familiarized to a stream of sequences each containing three nonsense words following the form *AXB*. Elements *A* and *B* always occurred together within the same sequence, whereas element *X* would vary. This produces a transitional probability of 1.0 for elements *A-B*, but lower transitional probabilities between elements *A*-*X* and *X*-*B*, as well as the *B* element of one word and the *A* element of the following word. Listening times were measured for infants and adults were asked to identify whether new strings violated or conformed with the order. They found that both 18-month-old infants and adults were able to extract non-adjacent dependencies. Other studies have extended these findings to exploring learning of non-adjacent dependencies in natural language instead of artificial languages. Friederici et al. (2011) showed that non-Italian-learning 4-month-old infants learned non-adjacent dependences in Italian.

Figure 3. Example of Gomez (2002) Non-Adjacent Dependencies



Other work has explored whether these findings extend beyond the linguistic domain.Creel et al. (2004) found that adults could simultaneously learn both adjacent and non-adjacent dependencies in sequences of musical tones when additional perceptual cues, such as timbre or octaves, could be used to facilitate learning. Other work has additionally found that adults could learn non-adjacent dependencies in sequences of varying non-linguistic non-musical noises, such as chirps, beeps, and purrs (Gebhart et al., 2009). There have also been several studies exploring non-adjacent dependencies in the visual domain. A study by Bettoni et al. (2021) presented 9- to 12-month-old and 13- to 15-month-old infants with stimuli modeled after Gomez (2002), but replacing linguistic nonsense words with visual geometric shapes and arrays of dots. They found that the older age group displayed longer looking times to ungrammatical sequences. Extending these findings, Lu and Mintz (2021) showed that 9-month-old infants could extract non-adjacent dependencies in visual action sequences. A study by van den Bos et al. (2012) further suggested that simultaneously presenting visual cues along with auditory cues may aid with complex non-adjacent dependency learning in adults.

Changes Across Development

Although infants from a young age display some sensitivity to non-adjacent dependencies, various studies have found notable changes in these sensitivities across development. Culbertson et al. (2016) explored non-adjacent dependencies between subjects and verbs in French. Their results showed that 15-, 18-, 21-, and 24-month-old French-learning infants distinguished between grammatical and ungrammatical sentences, although the different age groups showed either a familiarity or novelty preference, perhaps due to rapid changes in robustness of perceptual representations in infants. Gomez and Maye (2005) also found that although 17-month-olds could learn non-adjacent dependencies in artificial speech streams and showed a novelty preference for ungrammatical strings than grammatical, 15-month-olds showed evidence of learning non-adjacent dependencies but through a familiarity preference for grammatical strings[1], and 12-month-olds failed at the task. With 2- and 3-year-old children, van

[1] There are several factors that may contribute to whether infants show a novelty or familiarity preference, such as the complexity of stimuli and experiment length. See the Caveats and Other Related Research section for a more thorough discussion of the influence of these methodological factors.

der Kant et al. (2020) found a shift in neural correlates of non-adjacent dependency learning, such that neural evidence suggested that 2-year-olds were sensitive to linguistic non-adjacent dependencies, though 3-year-olds were not. Interestingly, they found the opposite pattern with non-linguistic stimuli, with 3-year-olds displaying evidence of learning non-adjacent dependencies, though 2-year-olds did not.

Several studies revealed several limitations in infants' ability to learn non-adjacent dependencies. Gomez (2002) found that although infants were able to succeed in learning non-adjacent dependencies, they could only do so when there was sufficient variability in the *X* element in the *AXB* order. Adults also learned non-adjacent dependencies better with more variability, suggesting that learning of adjacent versus non-adjacent dependencies may depend on what statistical information is most salient.

Other research has also suggested that performance may depend on facilitating perceptual cues. Marchetto and Bonatti (2013) explored whether infants could discriminate words from possible morphological syllable combinations. They first presented infants with an artificial speech stream of three-syllable words following an *AXB* order but manipulated the frequency of words, resulting in some words with equally high adjacent and non-adjacent transitional probabilities and other words with low adjacent but high non-adjacent transitional probabilities. When presenting infants with test items consisting of both types of words, they found that 18-month-olds listened longer to words with high non-adjacent transitional probabilities and low adjacent probabilities, whereas 12-month-olds only did so when words in the training speech stream were separated by pauses. Indeed, even adults have difficulty with non-adjacent dependencies without these facilitating cues such as pauses (Wang & Mintz, 2018).

Additionally, other research has suggested phonological properties may interfere with non-adjacent dependency learning. Gonzalez-Gomez and Nazzi (2012) showed that 10-month-olds were more sensitive to the non-adjacent dependencies in streams consisting of words with more frequent phoneme order. Onnis et al. (2005) similarly found English-speaking adults were more likely to rate words with low non-adjacent transitional probabilities but frequent English phonological properties as familiar than words with high non-adjacent transitional probabilities but infrequent phonological properties. Lastly, Newport and Aslin (2004) found that adults were sensitive only to some non-adjacent regularities: they were better at learning non-adjacent dependencies between consonants or vowels than full syllables.

Adjacent vs. Non-Adjacent Dependencies

Several studies have also investigated whether adjacent and non-adjacent dependency learning can occur concurrently and whether they are governed by different mechanisms. When adults are presented with pseudowords, they are able to learn both adjacent and non-adjacent dependencies concurrently (Vuong et al. 2016; Romberg & Saffran, 2013; Frost & Monaghan 2016). Interestingly, knowledge of adjacent dependencies may also facilitate non-adjacent dependency learning. A study by Lany and Gomez (2008) presented 12-month-old infants with an artificial language consisting of nonsense two-word phrases with either high or low adjacent dependencies, followed by the same artificial language but with two-word phrases separated by a novel third element (i.e. dependencies between non-adjacent elements). After habituating to these languages, when presented with new sequences violating non-adjacent dependencies, infants initially exposed to the language with high adjacent dependencies showed stronger evidence of learning the subsequent non-adjacent dependencies.

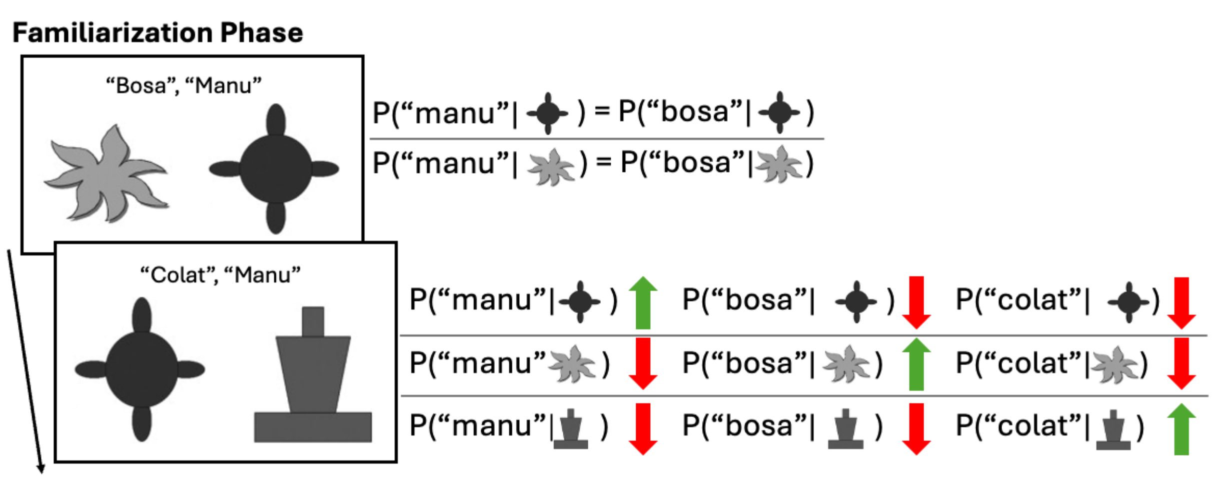
Cross-Situational Word Learning

Research has also suggested that learners can use statistical information across situations to map meanings onto words, what is commonly referred to as cross-situational word learning or cross-situational statistical learning. Similarly to correlation learning, this mechanism tracks correlations between features, though, in contrast to correlation learning, between features that occur sequentially rather than simultaneously. Taking the perspective of an infant learning language, there is a great deal of ambiguity in determining the referent of a heard word. If a parent says a novel word, "rabbit", and points generally to a scene of a backyard containing chairs, grass, flowers, trees, rabbits, and more, there are a number of possible meanings of that could be mapped onto that word, such as any of the individual objects in the scene (e.g. flower, tree), the entire scene itself, parts of individual objects (e.g. flower petal, rabbit's ear), etc. Cross-situational learning mechanisms address this issue and allow for tracking of statistical regularities across multiple situations to facilitate these word-meaning mappings. For example, after hearing the word "rabbit" in the backyard context, if an infant subsequently hears that same word in a new context, such as in a petting zoo, the infant could track the co-occurrences of the word and referent across situations to form the correct mapping.

A more straightforward example may be seeing images of a ball and a rabbit on a computer screen and hearing the words "ball" and "rabbit" for the first time. The word-meaning mappings are ambiguous in this scenario, since either word could map onto either referent. However, after subsequently seeing images of a rabbit and a mug and hearing "rabbit" and "mug", the co-occurrence of "rabbit" and the image of a rabbit can be calculated and is more frequent than any other word-referent pairing (i.e. "rabbit" and ball image), thus allowing the learner to make the correct mapping.

Although this is rather complex, various studies have found that infants, children, and adults use this mechanism to learn word-meaning mappings across situations (see Roembke et al. 2023 for a review). Yu and Smith (2007) presented adults with consecutive sets of four pictures and heard four nonsense words referring to the images, but without any information as to the specific word-picture pairings other than cross-situational statistics across trials. Adults were able to learn these word-referent mappings using cross-situational statistics. They also presented infants with several trials consisting of images of two objects simultaneously presented, followed by an audio stating two words. Both 12-month-olds and 14-month-olds looked longer towards the target referent corresponding with the verbal label, suggesting that cross-situational word learning develops early in infancy (Smith and Yu, 2008). Furthermore, by 12 months of age, infants and adults showed evidence of learning the meanings of minimal pairs with little phonetic discrimination between pairs (e.g. BON, DON) (Escudero et al, 2016a; 2016b), and preschool children can retain words learned through cross-situational word learning tasks (Vlach & DeBrock, 2019).

Figure 3. Illustration of Smith & Yu (2008) Paradigm



However, there also seem to be developmental changes in cross-situational word learning abilities. Vlach and Johnson (2013) found that 16-month-olds were less successful in cross-situational word learning than 20-month-olds, and especially when word-object pairs were interleaved. Yu and Smith (2011) probed further the mechanisms that underlie successful cross-situational word learning. Their findings suggest that infants who succeed at cross-situational word learning only store correct word-referent co-occurrences and form robust associations only for these pairs, whereas infants who have more difficulty with this task form weaker associations and split their attention between incorrect and correct pairs.

Other work with adults has also explored mechanisms underlying cross-situational word learning. Trueswell et al. (2013) aimed to explore whether adult learners consider a set of hypotheses regarding word meanings from one instance of learning to another, or whether they form only a single hypothesis and carry only that hypothesis forward. They found that adults' performance was only above chance when participants had previously guessed the correct word-referent mapping, but random when they had performed incorrectly on the previous instance, suggesting that indeed adults seem to form and carry forward only a single hypothesis for each potential word meaning. Yurovsky and Frank (2015) found that the degree to which learners can succeed at cross-situational word learning depends on the complexity of the task. When there were more possibilities for potential word meanings, participants encoded very little about candidate meanings other than the currently favored one, whereas when there were fewer possible meanings, they encoded information about other candidates as well.

Various studies have also explored the impact of other factors on cross-situational learning abilities. Suanda et al. (2014) explored the impact of contextual diversity of cross-situational word learning in children. They presented 5- through 7-year-olds with a standard cross-situational word learning task but manipulated the number of distractor items. They found that children could learn the meanings of words under all conditions of contextual diversity, though performance was better with increased contextual diversity. Dautriche and Chemla (2014) also explored the impact of context on mapping meanings to words through cross-situational word learning. They presented adults with a standard cross-situational word learning task with additionally introducing categories to facilitate word learning, such as presenting only distractor items from the same object category as the target (e.g. all clothes) or presenting some category distractors and some non-category distractors. They found that category information may serve as a guide for participants and aid in encoding word meanings, though it may also mislead participants to incorrectly select a same-category distractor as a word meaning. Another study by Fitneva and Christiansen (2017) explored the impact of initial accuracy on cross-situational word learning across development by presenting 4-year-olds, 10-year-olds, and adults with two images and asked them to identify which object corresponded with each label. They found that 4-year-olds benefited the most from initially accurate mappings and a greater number of accurate mappings, 10-year-olds benefited from any amount of initial accurate mappings, and adults benefited most from inaccurate initial mappings.

Several other studies have suggested that other cognitive systems may be related to cross-situational learning. Vlach and Debrock (2017, 2019) explored whether language and memory abilities predicted preschool children's cross-situational learning performance. In addition to presenting children with a standard cross-situational word learning task, children also completed several memory tasks, including object recognition, word recognition, and picture recognition tasks, as well as a vocabulary measure. They found that children's performance on both the language and memory tasks better predicted cross-situational learning performance than age, suggesting that other cognitive abilities may contribute to children's cross-situational learning abilities.

Previous studies have also found that cross-situational word learning is not limited to count nouns and objects. Scott and Fisher (2012) found that 2.5-year-olds formed correct action-label pairs using cross-situational statistics. Monaghan et al. (2015) also revealed that adults could learn noun-object pairings, verb-motion pairings, as well as both simultaneously through cross-situational word learning. Thus, they are capable of learning word meanings across multiple grammatical categories through cross-situational word learning. Rebuschat et al. (2021) found that adults could simultaneously learn noun, verb, and adjectives through cross-situational word learning, though they struggled to learn other grammatical categories. Lastly, Wang and Trueswell (2019) presented children with a set of novel images and asked children to select all the images that corresponded with a label at both the basic-level and subordinate-level. They found that children were able to learn subordinate level meanings for novel labels, though only when exemplars were from the same semantic domain (e.g. dalmatian vs. three other dogs) or when provided additional linguistic support.

Several studies have also explored whether cross-situational learning mechanisms are still effective when there is referential uncertainty. Smith et al. (2011) presented adults with multiple sets of six images and heard a nonsense word. Results suggested that when subsequently presented with 15 images, adults learned the correct word-object mappings. Vouloumanos and Werker (2009) explored whether infants were also capable of cross-situational word learning under uncertainty. Infants were presented with labeling events: some objects and labels always occurred together whereas others shared labels, with one label occurring at a high frequency and the other at a low frequency. Their results suggest that infants were able to learn word-referent pairs, and that they also kept track of low-frequency mappings.

Studies have also found that the presence of social cues may facilitate cross-situational word learning in scenarios with uncertainty. Macdonald et al. (2017) presented adults with sets of images and novel words, and also manipulated gaze cues, i.e., an agent stating the novel word either looked reliably or unreliably towards one of the objects. When eye gaze was reliable, adults used it to learn word meaning and stored less information about alternative word meanings. Furthermore, Frank et al. (2013) found that social or attentional cues, such as eye gaze or hand positioning, did not disambiguate between possible referents. Rather, they found that the combination of social cues and discourse continuity, or that a speaker continues to speak about the same thing as in their previous utterance, may be more useful than cross-situational statistics in some contexts.

Additionally, Yu and Ballard (2007) provide a statistical model for cross-situational learning, accounting for a set of words (*w*) and a specific number of words (*N)*, a set of meanings (*m*) and a specific number of meanings (*M*), and learning situation (*S*) and number of learning situations (s):

$$p(m_m|w_n) = \frac{\sum_{s=1}^{S} c(m_m|w_n, S_m^{(s)}, S_w^{(s)})}{\sum_{m=1}^{M} \sum_{s=1}^{S} c(m_m|w_n, S_m^{(s)}, S_w^{(s)})}$$

To demonstrate this procedure, take four utterances as four learning situations, such as “here is a cat”, “here is a dog”, “you like cat”, and “here is it”, and two objects with three possible referents across situations—cat, dog, and NON, which accounts for the possibility that a word may not have an extralinguistic referent. Taking the first learning situation, “Here is a cat” while presenting a toy cat, assuming none of these words have been heard before, the prior probabilities of the four possible word-referent mappings (i.e., “here”–cat, “is”–cat, “a”–cat,

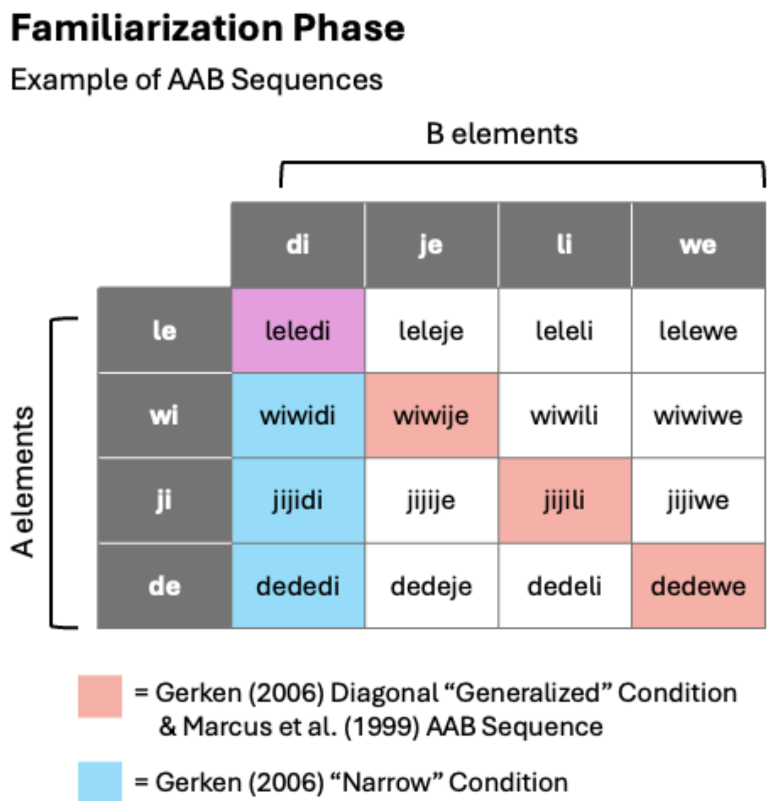
"cat"–cat) are considered to be uniform (.25). However, after the next learning situation, "Here is a dog" while presenting a toy dog, the probabilities are recalculated based on co-occurrences, such that "here", "is", and "a" have all co-occurred with both the cat referent and dog referent, thus resulting in a lower probability of each of those word-referent pairs. On the other hand, "dog" has only occurred with the dog referent and "cat" only with the cat referent, thus increasing the probability of those word-referent pairs. This process is repeated across situations to determine the highest probability word-referent pairs to determine the correct mappings.

Rule Learning

Rule learning is often contrasted with statistical learning, but rule learning mechanisms depend on the learner tracking distributional information in the input and extracting high level variables in order to generalize these abstract rules to novel exemplars. Not only can infants make inferences based on statistical information and learn statistical regularities in sequences of items, but they are also able to learn and generalize abstract rules. In contrast to transitional probabilities, rule learning refers to a more algebraic-like mechanism that allows for substitution of items that follow the same pattern, even if those items were never experienced before (Marcus et al., 1999). One common rule used to investigate this topic is an ABA rule. For example, one element, such as the syllable *wo*, may take the place of A in the sequence in one occurrence, and another element, *fe* may take the place of B in one occurrence (i.e. *wo-fe-wo*). However, an unlimited number of other elements could replace those same items in another occurrence yet still follow the same rule (e.g. *bu-pa-bu*, *to-di-to*, etc.). The transitional probabilities of entirely novel elements would be unknown since they were never experienced together and thus could not be used to determine statistical regularities, whereas rule learning allows for extraction of a more abstract, generalizable pattern to accommodate these substitutions.

Marcus et al. (1999) investigated whether infants can learn algebraic-like rules for which items can be substituted. They first familiarized infants to a two-minute sample of speech repeating three-word sentences that followed a specific rule (e.g. ABA - *wo fe wo*; ABB - *wo fe fe*). Then, they presented infants with entirely novel words that were never heard before either following the same rule as previously heard or following a new rule. They found 7-month-old infants looked longer towards sentences following a new rule, suggesting infants can learn rules and generalize them to novel elements. Gerken (2006) also explored whether rule learning depends on the distribution of category elements in exemplars that defines these rules. Similarly to Marcus et al. (1999), 9-month-old infants were first familiarized to speech samples of repeating three-word sequences following a specific rule. However, critically, some infants were first familiarized to syllable sequences that were generalized across varying exemplars, such as AAB sequences that all contained different elements but still adhered to the underlying pattern (e.g., *leledi*, *wiwije*, *jijili*, *jijiwe*), whereas other infants were familiarized to syllable sequencesthat were narrower in scope, such as AAB sequences that all ended with the same syllable (e.g., *leledi, wiwidi, jijidi, dededi*) where either a generalized underlying pattern (i.e. AAB pattern) could be extracted or a narrower pattern (i.e. sequences end with syllable *di*). Results from this study found that when infants were presented with test sequences that contained entirely novel syllables, only infants who were first familiarized to the generalized sequences showed evidence of distinguishing novel test sequences that violated the underlying rule presented in the familiarization phase from novel test sequences that adhered to this rule.

Figure 4. Example of Gerken (2006) and Marcus et al. (1999) Stimuli



Following this, several studies have explored whether infants can learn rules outside of speech. Saffran et al. (2007) presented 7-month-olds with sequences of color photographs of different dog breeds following one of two different sequences (e.g. malamute, cattle dog, malamute). When presented with new dog breeds either adhering to or violating the original sequence, 7-month-olds looked longer to the novel pattern than the familiar pattern, suggesting that infants were able to detect rules in these non-linguistic stimuli. Further, these results suggest that rule learning is not limited to language or speech. Bulf et al. (2015) also found that when presenting sequences of faces following specific patterns (e.g. face A, face B, face A), 7-month-old infants were able to generalize this pattern to a new set of faces and looked longer when the sequence of faces followed a new pattern compared to a previously observed pattern. In contrast, Johnson et al. (2009) presented visual sequences of shapes to 8- and 11-month-olds following several different patterns, and only 11-month-olds showed evidence of learning and distinguishing these patterns.

Dawson and Gerken (2009) additionally found conflicting evidence of rule learning in non-linguistic domains. They presented 4- and 7-month-old infants with continuous sequences of three-chord phrases following one of two different patterns. When presented with either new sequences that followed the previous pattern or followed a novel pattern, they found 4-month-olds looked longer to sequences that followed a novel pattern than a familiar pattern, though 7-month-olds failed to do so.

Another study by Marcus et al. (2007) presented 7.5-month-old infants with sound sequences, consisting of either sung syllables, musical tone sequences, or animal sounds. Infants only distinguished between familiar and unfamiliar patterns with syllables. But when infants were first familiarized to patterns in speech, they were able to generalize the learned rule to new stimuli consisting of non-speech sounds, suggesting speech may facilitate rule learning to new domains that infants may previously have had difficulty with. Rabagliati et al. (2012) sought to explore whether the advantage of speech in rule learning may stem from the communicative nature of speech compared to other sound or visual elements that are not communicative. Instead of presenting speech sequences to infants, they presented videos of a person performing sign-like gestures in specific patterns. Notably, infants were able to extract some rules in sign-like gestures, though failed to do so in others, suggesting that the sensitivity to patterns in speech may not be due to its communicative quality.

Statistical Inference/Probabilistic Inference

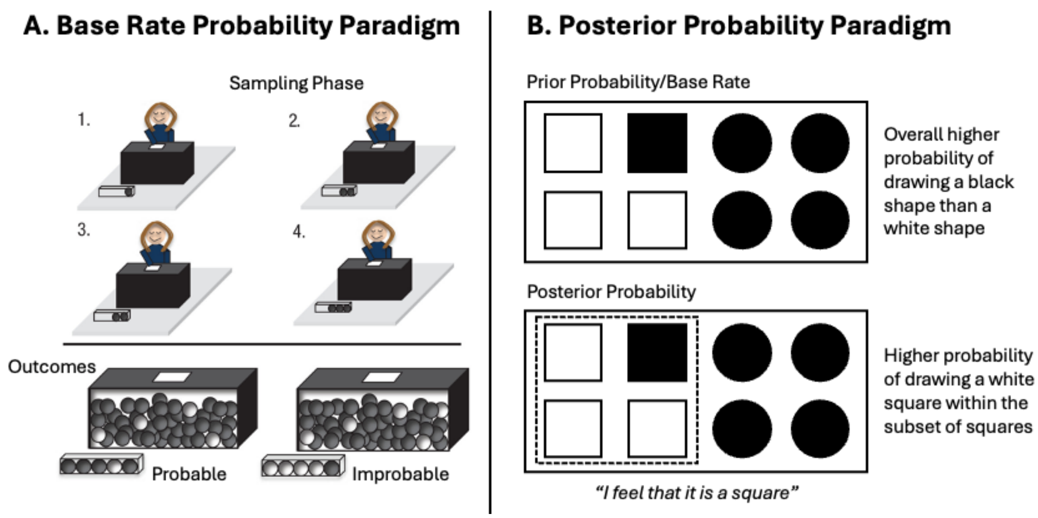
Although statistical learning often refers to the use of transitional probabilities, other forms of probabilistic information are also present in everyday life. For example, when buying a bag of mixed-flavor candies at the market, one might select the bag with the most amount of their favorite flavor visible, since this would have the highest probability of containing more of their desired flavor compared to the other less desirable flavors and to the ratio of flavors in other bags. This can be defined by the following formula, with *P(E)* representing the probability of an event occurring, *n(A)* representing the number of favorable outcomes and *n(S)* representing the number of total possible outcomes in the sample:

$$P(E) = \frac{n(E)}{n(S)}$$

This may seem intuitive, though this ability to extract probabilistic information, specifically making inferences about samples from populations and vice versa, and use this information is a powerful tool. Previous work has suggested that infants, children, and adults are sensitive to these types of statistical information and can utilize this information in statistical learning mechanisms. For example, much research to date has explored infants' and children's sensitivity to probabilistic information and their abilities to make judgements based on this information (see Denison & Xu, 2019 for a review).

Several studies have explored infants' abilities to make inferences about populations from the statistical information of a randomly drawn sample of objects. Xu and Garcia (2008) investigated whether 8-month-old infants were able to make predictions about a population of items based on a small sample. Infants were first familiarized with a large box containing either mostly red or mostly white ping pong balls visible to the infant. Then, with the contents hidden to the infant, five balls were randomly drawn from a similar box, drawing either a majority of red balls or majority of white balls. Lastly, the contents of the box were revealed to the infant. Results suggested that 8-month-olds looked longer when the randomly drawn sample of balls was improbable based on the revealed population of balls inside the box (e.g. if the sample was majority red but the box contained mostly white) than when the sample and population were consistent, indicating that these infants were able to make predictions about a population of items based on a randomly drawn sample. Additionally, Xu and Garcia (2008) found that infants were able to make predictions in the opposite direction, about a sample based on a population. Denison et al. (2013) explored the development of this ability and found that 6-month-olds, but not 4.5-month-olds, exhibited the same pattern of longer looking to improbable than probable populations given a randomly drawn sample.

Figure 5. Depiction of Probabilistic Inference Paradigms



Note: A) Denison and Xu (2019) illustration of Xu and Garcia (2008) paradigm B) example of posterior probability task from Girotto and Gonzalez (2008).

Infants can also make predictions about single events based on probabilistic information. A study by Téglás et al. (2007) presented 12-month-old infants with a 'lottery machine', containing three yellow objects and one blue object, and found that the infants looked longer when a blue object exited the machine (the improbable outcome) than when a yellow object did so (the probable outcome). Teglas and Bonatti (2016) provided further evidence that infants can reason about single-event probabilities, expecting a ball to exit from the probable side of a container (with three exits) than the improbable side (with one exit). Denison and Xu (2010b) found infants can also make predictions about single events with large set sizes. Given two locations, one of which is more likely to contain their preferred object randomly drawn from a population of two types of objects, 12- to 14-month-old infants searched more often at that location. This basic probabilistic reasoning ability does not seem to depend on formal education, e.g., Maya adults and children succeeded in probabilistic reasoning tasks as robustly as Italian children (Fontanari et al. 2014).

Other studies found that perceptual cues may influence infants' ability to reason about single-event probabilities, e.g., the total number of objects of each type in a scene, the physical distance of objects to the exit opening, as well as the duration that the contents of containers were occluded (Téglás et al. 2011). Lawson and Rakison (2013) similarly found that perceptual cues may impact infants' sensitivity to base rate information, or the probability of an event occurring. Their results suggest that 8-month-olds were more sensitive to congruency of outcome and its surrounding ball colors at the time of exit over probabilistic information, though this bias seems to subside by 12 months. Additionally, Teglas et al. (2015) found that 12-month-olds could only make probabilistic judgements for large set sizes when they could also use physical cues to assist their reasoning (e.g., distance to the exit and sizes of objects).

Although infants from an early age can make inferences based on statistical evidence, these abilities are still developing throughout early and later childhood. Teglas et al. (2007) found that 3- to 5-year-olds were able to predict that a ball was more likely to exit a 3-hole side than a 1-hole side of a container. Yet, Girotto et al. (2016) found that only 5-year-olds, but not 3- and 4-year-olds, were able to make correct predictions based on probabilistic information, even when all age groups succeeded in the task with deterministic information. A study by O'Grady and Xu (2020) provides additional evidence of children's emerging probabilistic reasoning abilities. When presenting 6- through 12-year-olds with bags containing different populations of marbles and asking children to select the bag that would yield a desired outcome, children were overall able to make quick and accurate predictions based on probabilistic information. However, performance varied across ages depending on other cues, such as area (e.g. an undesired color of marbles covering a larger area than desired marbles) or numeric cues (e.g. lower probability population containing larger number of desired marbles than higher probability population).

Furthermore, other more complex forms of probabilistic reasoning emerge in early childhood, such as posterior probabilities, or base rates that are updated based on new given information. Girotto and Gonzalez (2008) presented children with a bag of chips varying in shape (e.g. squares and circles) and color (e.g. black or white) and asked children to guess the color of a randomly sampled object. If children can reason about posterior probabilities, when told the shape of the sampled object, they should be able to incorporate this additional information into their probabilistic judgment. For example, if the majority of square chips are white despite the majority of the total population being black, children's judgments should shift upon learning this new information. Results suggest that 6- to 7-year-old children are able to successfully integrate this new information to form a posterior probability, though younger children failed to do so.

Probabilistic reasoning across domains

Physical and Psychological Reasoning. Although there are some constraints on infants' abilities to make probabilistic predictions, statistical inference/probabilistic reasoning mechanisms are quite robust, with infants and children being able to integrate other physical and psychological variables when making these judgments. Research has shown that infants can integrate physical properties of objects in their statistical inferences. Teglas et al. (2007), Teglas et al. (2011), and Denison and Xu (2010a) found that infants can integrate physical properties of objects in their probabilistic inference (e.g., if an object is blocked by a barrier; excluding objects who are not moveable from the computation).

Infants are also able to use statistical information to make inferences about the social world. Attisano and Denison (2020) and Xu and Denison (2009) explored whether infants' inferences differ when sampling is intentional versus random. Both 6- and 11-month-old infants integrated an agent's preference and their visual access to the population from which the sample was drawn in making predictions about which samples were probable.

Several studies focused on using non-random sampling to infer an agent's *preferences*. Wellman et al. (2016) found that infants infer preferences from statistical information. Infants either viewed an agent pull five blue balls from a box either containing a majority red balls or a majority blue balls. When the same agent was given the opportunity to choose between either a new bowl of entirely blue balls or new bowl of entirely red balls, 10-month-old infants only looked longer when the agent selected the bowl of red balls when they had previously selected all blue balls from a box where blue was the minority color. Using a similar method, Kushnir et al. (2010) introduced preschoolers and 20-month-olds to a puppet who selected five toys from a box, with the box containing either 18%, 50%, or 100% of the toys the puppet liked. Their results revealed that children's inferences regarding the puppet's preferences were graded: stronger for the 18% than the 50% than the 100% condition.

Ma and Xu (2011) investigated whether non-random sampling may teach infants to understand subject preference earlier. 16-month-olds and 2-year-olds initially assumed that the experimenter shared their preference for a more interesting type of toy. After watching the experimenter sample the boring type of boy from a jar (consisting of 13% boring toys and 87% interesting toys), they inferred that the experimenter liked the boring toy better even though they themselves continued to prefer the interesting toy. Two-year-olds succeeded in this task more robustly than 16-month-olds, suggesting that this ability to infer subjective preferences based on statistical evidence develops over time.

Other studies have also suggested that verbal framing may influence young children's ability to infer preferences from statistical information. When presenting 4- and 5-year-old children with an emotion-framed scenario (e.g. "some toys make her *happy*") compared to an action-framed scenario (e.g. "she *gets* some toys"), preschoolers were able to use statistical non-randomness to infer an agent's preference only when verbal framing was emotion-related (Garvin & Woodward, 2015). Children are also able to make even more sophisticated social inferences from statistical information. A study by Diesendruck et al. (2015) explored whether children can generalize preferences to others of the same group based on statistical information, finding that children could generalize preferences but only with sufficient statistical and social cues (e.g., multiple agents demonstrated the same preference that violated random sampling).

Not only are children able to make social inferences about preferences from probabilistic information, but they are also able to infer *emotional states*. Schlottmann (2001) found that children as young as 6 years of age judged that a puppet would be happier in response to a desired, improbable outcome than when the result was more probable, and, similarly, Schlottmann and Anderson (1994) found that children as young as 5 years correctly identified that a puppet would be happier playing a game with a higher probability of a desired outcome than a lower probability. Doan et al. (2018) investigated whether children could use probabilistic information to attribute surprise to others in response to an outcome. They presented children with two characters receiving gumballs from their respective gumball machines, one with a majority desired gumball or and one with a majority undesired gumball distribution. When asked which character was more surprised when receiving a desired outcome, 7-year-olds were reliably able to use probabilistic information to attribute surprise, 6-year-olds were inconsistent in their responses but able to use probabilistic information when given facilitating prompts, and 4- and 5-year-olds failed to do so. Other work has found infants can also make inferences in the opposite direction, with 12- through 18-month-olds looking longer when a probable outcome was revealed than an improbable outcome after an experimenter had initially exhibited surprise to the hidden outcome (Wu et al., 2024). Another study by Doan et al. (2020) found that by 5 years of age, children inferred an agent would be happier with a desired, improbable outcome than a desired, probable outcome.

Children can also infer goals from probabilistic information. Using a similar paradigm to Doan et al. (2018), Doan et al. (2022) again presented children with two gumball machines varying in their distributions of colored gumballs (either majority red or majority purple). When a character chose to press the machine with majority red, 3-year-olds inferred that her goal was to get a red gumball.

Probabilistic information can also be used to infer social relationships. Sehl et al. (2023) presented participants with diagrams of two people and their friendships with others in a group. They manipulated the number of friends each target person had, as well as the proportion of mutual friends between the two target people (e.g. three of nine friends were mutual, three of four friends were mutual, etc.). They found that both 5- to 7-year-olds and adults used the proportion of mutual friends between the two targets to infer the likelihood of them being friends, suggesting that from an early age, children can infer social networks using probabilistic information.

Causal Learning. Other research has suggested children are also able to make causal inferences from probabilistic information. Gopnik et al. (2001) presented 3- and 4-year-old children with a blicket-detector machine that illuminated and played music when blickets were placed on top. When children were shown two novel objects, with one object activating the detector 100% of the time and the other object only activating the detector 66% of the time, they identified both objects as being blickets, suggesting that they could draw causal conclusions even when observed evidence was probabilistic instead of deterministic.

Another study by Schulz and Gopnik (2004) similarly explored whether children could make causal inferences from independent and dependent probabilities of events. They presented 3- through 5-year-old children with various flowers either individually or in pairs, with pairs causing a puppet to sneeze either 100% of the time (e.g. flowers A and C together; flowers B and C) or 0% of the time (e.g. flowers A and B; though sneezing occurred with each flowers A and B 66% of the time), and individual flowers causing sneezing 100% of the time (e.g. flower C) or 66% of the time (e.g. flower A, flower B). Their results suggest that children who were presented with the flowers simultaneously most often selected the correct flower as causing the puppet to sneeze, whereas children who were only shown each flower individually selected all flowers at equal rates. This suggests that children were able to use conditional probabilities to make causal inferences.

Waismeyer et al. (2015) found that even younger children have this ability: 2-year-olds selected a toy that activated the machine 66% of the time more often than a toy that activated the machine 33% of the time. Kushnir and Gopnik (2005) explored whether children could infer the strength of causal relations based on probabilistic information. They presented 4-year-old children with a detector machine and provided either consistent evidence of which blocks activated the machine (e.g. Block A alone activated machine 100% of time, Block B alone only 33%), ambiguous evidence (e.g. Block A alone 100% of time, Blocks A and B together 66%), or conflicting evidence (e.g. Block A alone 100% of time, Block B 66%). Results suggested that children who were presented with consistent or ambiguous evidence were most likely to correctly identify the block with the highest probability of activating the machine, suggesting children could infer strength of causal powers from probabilistic information.

Kushnir and Gopnik (2007) also found that children's initial assumption about how to activate a machine – by placing a block on top of it – can be overridden with probabilistic information. That is, they choose a block with higher probability of activation when held above the machine over a block with lower probability of activation when placed on top of the machine. However, children sometimes struggle to use probabilistic information to infer causal structures under certain conditions. Kushnir et al. (2009) found that when children observe another person's action, they correctly infer causal relationships based on the strength of the evidence, but they rely more on their own actions (self-agency bias) and override probabilistic evidence in their inferences about causal structures.

Caveats and Other related research

It is important to note that several methodological differences may have contributed to failure or developmental differences in performance on these tasks exploring different mechanisms, especially for infants. For example, a meta-analysis conducted by Isbilen and Christiansen (2022) suggested that the specific type of tasks used to measure statistical learning mechanisms in children and adults significantly impacted effect sizes, with processing-based tasks (i.e. tasks that tap more directly into statistical learning mechanisms, such as variants on reaction time tasks) led to larger effect sizes than reflection-based measures (i.e. tasks that require participants to reflect on learned material, such as 2AFC tasks).

Similarly, there are also methodological variations for infant studies. With standard looking time measures used to test infants, vast majority of the time the prediction is that infants should show a novelty preference and look longer towards the novel stimulus than the familiar stimulus. However, infants sometimes look longer towards the familiar stimulus. One plausible explanation is that with complex stimuli may take more time to process so if infants are tested while they are still processing the stimli, they will show a familiarity preference. For example, Hunter et al. (1983) found that infants who were interrupted during a habituation phase continued to display a familiarity preference for a toy only if the toy was complex relative to the infants' age, whereas infants were not interrupted during the habituation phase shifted to a novelty preference. It is possible that using less complex stimuli, such as reducing the number of elements in the mini-corpus of stimuli elements, may have resulted in success for several of these tasks where infants failed to show evidence of sensitivity to statistical patterns.

Relatedly, with the habituation-dishabituation paradigms that many of studies employ, infants tend to shift their preference from familiar stimuli to novel stimuli with increased exposure to the familiar stimuli. It is possible that infants who failed to show evidence of distinguishing between stimuli based on statistical structures may begin to show a difference in preference with increased exposure to the stimuli. Additionally, much work has suggested that infants have rather poor working memory (see Reynolds & Romano, 2016 for a review). Many studies require infants to engage their working memory abilities to hold elements in memory in order to extract the underlying statistical structures. Thus, it is possible that infants' failure to show evidence of statistical learning in these paradigms may be due to limitations on working memory rather than limitations on statistical learning abilities.

Finally, there are several related research topics that are beyond the scope of this review. For example, research has explored learning of speech categories based on the statistical distribution of speech sounds. Maye et al. (2002) familiarized 6- and 8-month-old infants to speech streams consisting of speech sounds on a continuum between voiced and voiceless consonants (i.e. voiceless [t] vs. voiced [d]), with those voiced and voiceless consonants as the endpoints. Importantly, some infants were presented with speech streams that were bimodally distributed, such that speech sounds closer towards the endpoints were presented more frequently than sounds between those points, whereas other infants were presented with a unimodal distribution, with the endpoint sounds presented less frequently than the sounds between those points. During test trials, two different types of streams were presented, with one stream repeatedly playing only one single sound whereas the other stream alternated between two different sounds. Only infants familiarized to the bimodally distributed speech stream showed evidence of distinguishing between the two speech sound categories, suggesting they are sensitive to the statistical distribution of sounds when forming speech categories. Several other studies have expanded upon this work (Maye et al., 2002; Maye et al., 2008; Teinonen et al., 2008) and have developed computational models of this mechanism (McMurray et al., 2008; Schatz et al., 2021; Vallabha et al., 2007).

Additionally, a large body of work has expanded on rule learning to more complex forms of grammar learning. For example, Gomez and Gerken (1999) constructed an artificial grammar, with specific rules imposed on nonsense words such that only some nonsense words could follow others in sentences. Although this is similar to rule learning, these artificial grammars extend beyond single rules (e.g. AAB) to larger, more complex structures. Their results suggest that 12-month-old infants showed evidence of extracting grammatical structures that could also be generalized to novel exemplar. Other work has further explored grammar learning across development (see Gomez & Gerken, 2000 for a review). Additional work has explored adults' ability to extract generalizable grammatical structures based on statistical distributional information purely from exposure to an artificial grammar composed of nonsense words and categorize novel exemplars into these grammatical categories (Reeder et al., 2013).

Lastly, there is additional work exploring the role of memory and chunking in statistical learning (see Thiessen, 2017 for a review; Isbilen et al. 2020). These studies focused on how cognitive processes such as memory may significantly affect statistical learning.

Summary and Conclusions

In summary, a large number of statistical learning mechanisms enable infants, children, and adults to find regularities and structure in order to acquire knowledge of various domains and to understand their environment. Although statistical learning has traditionally been synonymous with learning based on transitional probabilities, various other mechanisms fall under the broader category of statistical learning, such as correlational learning and statistical inference. Our review provides an overview of the seminal and recent findings. These mechanisms are robust and explain learning in a vast array of domains, from language and visual sequences to emotion attribution.

From this review, several unexplored questions regarding statistical learning mechanisms have emerged. First, future work should continue to clarify the role of statistical learning in language acquisition across development, such as how transitional probabilities used in word segmentation may relate to word or syntax learning and how this may change with age. Second, future work should seek to determine how early in development these processes emerge. Third, future research may investigate further whether young learners understand other aspects of probabilistic inference. Lastly, future work should continue to explore the limitations of these statistical learning mechanisms, how they are used over the course of development, and how they interact with world knowledge that learners accumulate over time.

Further Reading

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Denison, S., & Xu, F. (2019). Infant Statisticians: The Origins of Reasoning Under Uncertainty. *Perspectives on Psychological Science*, *14*(4), 499–509. https://doi.org/10.1177/1745691619847201

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule Learning by Seven-Month-Old Infants. *Science*, *283*(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review*, *111*(1), 3–32. https://doi.org/10.1037/0033-295X.111.1.3

Yu, C., & Smith, L. B. (2007). Rapid Word Learning Under Uncertainty via Cross-Situational Statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Gómez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science*, *13*(5), 431–436. https://doi.org/10.1111/1467-9280.00476

Benton, D. T., Rakison, D. H., & Sobel, D. M. (2021). When correlation equals causation: A behavioral and computational account of second-order correlation learning in children. *Journal of Experimental Child Psychology*, *202*, 105008. https://doi.org/10.1016/j.jecp.2020.105008

Younger, B. A., & Cohen, L. B. (1983). Infant Perception of Correlations among Attributes. *Child Development*, *54*(4), 858–867. https://doi.org/10.2307/1129890

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van der Kant, A., Smith, K., & Rey, A. (2020). Non-adjacent Dependency Learning in Humans and Other Animals. *Topics in Cognitive Science*, *12*(3), 843–858. https://doi.org/10.1111/tops.12381

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

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063

Attisano, E., & Denison, S. (2020). Infants' reasoning about samples generated by intentional versus non-intentional agents. *Infancy*, *25*(1), 110–124. https://doi.org/10.1111/infa.12320

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407. https://doi.org/10.1016/j.cognition.2007.07.005

Benjamin, L., Fló, A., Palu, M., Naik, S., Melloni, L., & Dehaene-Lambertz, G. (2023). Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. *Developmental Science*, *26*(2), e13300. https://doi.org/10.1111/desc.13300

Benton, D. T., Rakison, D. H., & Sobel, D. M. (2021). When correlation equals causation: A behavioral and computational account of second-order correlation learning in children. *Journal of Experimental Child Psychology*, *202*, 105008. https://doi.org/10.1016/j.jecp.2020.105008

Bettoni, R., Bulf, H., Brady, S., & Johnson, S. P. (2021). Infants' learning of non-adjacent regularities from visual sequences. *Infancy*, *26*(2), 319–326. https://doi.org/10.1111/infa.12384

Bulf, H., Brenna, V., Valenza, E., Johnson, S. P., & Turati, C. (2015). Many faces, one rule: the role of perceptual expertise in infants' sequential rule learning. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01595

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132. https://doi.org/10.1016/j.cognition.2011.06.010

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119

Culbertson, J., Koulaguina, E., Gonzalez-Gomez, N., Legendre, G., & Nazzi, T. (2016). Developing knowledge of nonadjacent dependencies. *Developmental Psychology*, *52*(12), 2174–2183. https://doi.org/10.1037/dev0000246

Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 892–903. https://doi.org/10.1037/a0035657

Dawson, C., & Gerken, L. (2009). From domain-generality to domain-sensitivity: 4-Month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition*, *111*(3), 378–382. https://doi.org/10.1016/j.cognition.2009.02.010

Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology*, *49*(2), 243–249. https://doi.org/10.1037/a0028278

Denison, S., & Xu, F. (2010a). Integrating physical constraints in statistical inference by 11-Month-Old Infants. *Cognitive Science*, *34*(5), 885–908. https://doi.org/10.1111/j.1551-6709.2010.01111.x

Denison, S., & Xu, F. (2010b). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, *13*(5), 798–803. https://doi.org/10.1111/j.1467-7687.2009.00943.x

Denison, S., & Xu, F. (2019). Infant statisticians: The origins of reasoning under uncertainty. *Perspectives on Psychological Science*, *14*(4), 499–509. https://doi.org/10.1177/1745691619847201

Diesendruck, G., Salzer, S., Kushnir, T., & Xu, F. (2015). When choices are not personal: The effect of statistical and social cues on children's inferences about the scope of preferences. *Journal of Cognition and Development*, *16*(2), 370–380. https://doi.org/10.1080/15248372.2013.848870

Doan, T., Friedman, O., & Denison, S. (2018). Beyond belief: The probability-based notion of surprise in children. *Emotion*, *18*(8), 1163–1173. https://doi.org/10.1037/emo0000394

Doan, T., Stonehouse, E., Denison, S., & Friedman, O. (2022). The odds tell children what people favor. *Developmental Psychology*, *58*(9), 1759–1766. https://doi.org/10.1037/dev0001395

Escudero, P., Mulak, K. E., & Vlach, H. A. (2016a). Cross-situational learning of minimal word pairs. *Cognitive Science*, *40*(2), 455–465. https://doi.org/10.1111/cogs.12243

Escudero, P., Mulak, K. E., & Vlach, H. A. (2016b). Infants encode phonetic detail during cross-situational word learning. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01419

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: statistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260. https://doi.org/10.1111/j.1467-9280.2007.01885.x

Estes, K. G., & Lew-Williams, C. (2015). Listening through voices: Infant statistical word segmentation across multiple speakers. *Developmental Psychology*, *51*(11), 1517–1528. https://doi.org/10.1037/a0039725

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826. https://doi.org/10.1073/pnas.232472899

Fitneva, S. A., & Christiansen, M. H. (2017). Developmental changes in cross-situational word learning: The inverse effect of initial accuracy. *Cognitive Science*, *41*(S1), 141–161. https://doi.org/10.1111/cogs.12322

Fló, A., Benjamin, L., Palu, M., & Dehaene-Lambertz, G. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Scientific Reports*, *12*(1), 4391. https://doi.org/10.1038/s41598-022-08411-w

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, *9*(1), 1–24. https://doi.org/10.1080/15475441.2012.707101

Friederici, A. D., Mueller, J. L., & Oberecker, R. (2011). Precursors to natural grammar learning: Preliminary evidence from 4-month-old infants. *PLOS ONE*, *6*(3), e17920. https://doi.org/10.1371/journal.pone.0017920

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

Garvin, L. E., & Woodward, A. L. (2015). Verbal framing of statistical evidence drives children's preference inferences. *Cognition*, *138*, 35–48. https://doi.org/10.1016/j.cognition.2015.01.011

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, *16*(3), 486–490. https://doi.org/10.3758/PBR.16.3.486

Gerken, L. (2006). Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), B67–B74. https://doi.org/10.1016/j.cognition.2005.03.003

Girotto, V., Fontanari, L., Gonzalez, M., Vallortigara, G., & Blaye, A. (2016). Young children do not succeed in choice tasks that imply evaluating chances. *Cognition*, *152*, 32–39. https://doi.org/10.1016/j.cognition.2016.03.010

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition*, *106*(1), 325–344. https://doi.org/10.1016/j.cognition.2007.02.005

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436. https://doi.org/10.1111/1467-9280.00476

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135. https://doi.org/10.1016/S0010-0277(99)00003-7

Gómez, R. L. & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186. https://doi.org/10.1016/S1364-6613(00)01467-4

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206. https://doi.org/10.1207/s15327078in0702_4

Gonzalez-Gomez, N., & Nazzi, T. (2012). Acquisition of nonadjacent phonological dependencies in the native language during the first year of life. *Infancy*, *17*(5), 498–524. https://doi.org/10.1111/j.1532-7078.2011.00104.x

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620–629. https://doi.org/10.1037/0012-1649.37.5.620

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, *19*(3), 338–352. https://doi.org/10.1037/0012-1649.19.3.338

Isbilen, E. S., & Christiansen, M. H. (2022). Statistical learning of language: A meta-analysis into 25 years of research. *Cognitive Science*, *46*(9), e13198. https://doi.org/10.1111/cogs.13198

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*(7), e12848. https://doi.org/10.1111/cogs.12848

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345. https://doi.org/10.1111/j.1467-7687.2009.00886.x

Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., & Slemmer, J. A. (2009). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy*, *14*(1), 2–18. https://doi.org/10.1080/15250000802569611

Jonaitis, E. M., & Saffran, J. R. (2009). Learning harmony: The role of serial statistics. *Cognitive Science*, *33*(5), 951–968. https://doi.org/10.1111/j.1551-6709.2009.01036.x

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 1559–1571. https://doi.org/10.1111/j.1467-8624.2007.01083.x

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, *16*(9), 678–683. https:/doi.org/10.1111/j.1467-9280.2005.01595.x

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *43*(1), 186–196. https://doi.org/10.1037/0012-1649.43.1.186

Kushnir, T., Wellman, H. M., & Gelman, S. A. (2009). A self-agency bias in preschoolers' causal inferences. *Developmental Psychology*, *45*(2), 597–603. https://doi.org/10.1037/a0014727

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*(8), 1134–1140. https://doi.org/10.1177/0956797610376652

Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*(12), 1247–1252. https://doi.org/10.1111/j.1467-9280.2008.02233.x

Lany, J., Karaman, F., & Hay, J. F. (2024). A changing role for transitional probabilities in word learning during the transition to toddlerhood? *Developmental Psychology*, *60*(3), 567–581. https://doi.org/10.1037/dev0001641

Lawson, C. A., & Rakison, D. H. (2013). Expectations about single event probabilities in the first year of life: The influence of perceptual and statistical information. *Infancy*, *18*(6), 961–982. https://doi.org/10.1111/infa.12014

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

Lu, H. S., & Mintz, T. H. (2021). Learning non-adjacent rules and non-adjacent dependencies from human actions in 9-month-old infants. *PLOS ONE*, *16*(6), e0252959. https://doi.org/10.1371/journal.pone.0252959

Ma, L., & Xu, F. (2011). Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition*, *120*(3), 403–411. https://doi.org/10.1016/j.cognition.2011.02.003

MacDonald, K., Yurovsky, D., & Frank, M. C. (2017). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology*, *94*, 67–84. https://doi.org/10.1016/j.cogpsych.2017.02.003

Madole, K. L., & Cohen, L. B. (1995). The role of object parts in infants' attention to form-function correlations. *Developmental Psychology*, *31*(4), 637–648. https://doi.org/10.1037/0012-1649.31.4.637

Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology*, *67*(3), 130–150. https://doi.org/10.1016/j.cogpsych.2013.08.001

Marcovitch, S., & Lewkowicz, D. J. (2009). Sequence learning in infancy: the independent contributions of conditional probability and pair frequency information. *Developmental Science*, *12*(6), 1020–1025. https://doi.org/10.1111/j.1467-7687.2009.00838.x

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science*, *18*(5), 387–391. https://doi.org/10.1111/j.1467-9280.2007.01910.x

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Mareschal, D., Powell, D., Westermann, G., & Volein, A. (2005). Evidence of rapid correlation-based perceptual category learning by 4-month-olds. *Infant and Child Development*, *14*(5), 445–457. https://doi.org/10.1002/icd.415

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134. https://doi.org/10.1111/j.1467-7687.2007.00653.x

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x

Mermier, J., Quadrelli, E., Turati, C., & Bulf, H. (2022). Sequential learning of emotional faces is statistical at 12 months of age. *Infancy*, *27*(3), 479–491. https://doi.org/10.1111/infa.12463

Monaghan, P., Mattock, K., Davies, R. A. I., & Smith, A. C. (2015). Gavagai is as gavagai does: Learning nouns and verbs from cross-situational statistics. *Cognitive Science*, *39*(5), 1099–1112. https://doi.org/10.1111/cogs.12186

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology*, *157*, 14–28. https://doi.org/10.1016/j.jecp.2016.12.004

Monroy, C. D., Meyer, M., Schröer, L., Gerson, S. A., & Hunnius, S. (2019). The infant motor system predicts actions based on visual statistical learning. *NeuroImage*, *185*, 947–954. https://doi.org/10.1016/j.neuroimage.2017.12.016

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

O'Grady, S., & Xu, F. (2020). The development of nonsymbolic probability judgments in children. *Child Development*, *91*(3), 784–798. https://doi.org/10.1111/cdev.13222

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*(2), 225–237. https://doi.org/10.1016/j.jml.2005.02.011

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247. https://doi.org/10.1016/j.cognition.2009.07.011

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685. https://doi.org/10.1111/j.1467-8624.2009.01290.x

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, *36*(7), 1299–1305. https://doi.org/10.3758/MC.36.7.1299

Rabagliati, H., Senghas, A., Johnson, S., & Marcus, G. F. (2012). Infant rule learning: advantage language, or advantage speech? *PLOS ONE*, *7*(7), e40517. https://doi.org/10.1371/journal.pone.0040517

Rakison, D. H., & Benton, D. T. (2019). Second-order correlation learning of dynamic stimuli: Evidence from infants and computational modeling. *Infancy*, *24*(1), 57–78. https://doi.org/10.1111/infa.12274

Rakison, D. H., & Poulin-Dubois, D. (2002). You go ghis way and I'll go that way: Developmental changes in infants' detection of correlations among static and dynamic features in motion events. *Child Development*, *73*(3), 682–699. https://www.jstor.org/stable/3696244

Rebuschat, P., Monaghan, P., & Schoetensack, C. (2021). Learning vocabulary and grammar from cross-situational statistics. *Cognition*, *206*, 104475. https://doi.org/10.1016/j.cognition.2020.104475

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, *66*(1), 30–54. https://doi.org/10.1016/j.cogpsych.2012.09.001

Reynolds, G. D., & Romano, A. C. (2016). The development of attention systems and working memory in infancy. *Frontiers in Systems Neuroscience*, *10*. https://doi.org/10.3389/fnsys.2016.00015

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

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*(7), 1290–1320. https://doi.org/10.1111/cogs.12050

Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition*, *81*(2), 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://www.jstor.org/stable/2891705

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*(3), 669–680. https://doi.org/10.1016/j.cognition.2006.11.004

Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., & Dupoux, E. (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences*, *118*(7), e2001844118. https://doi.org/10.1073/pnas.2001844118

Schlottmann, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development*, *72*(1), 103–122. https://doi.org/10.1111/1467-8624.00268

Schlottmann, A., & Anderson, N. H. (1994). Children's judgments of expected value. *Developmental Psychology*, *30*(1), 56–66. https://doi.org/10.1037/0012-1649.30.1.56

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*(2), 162–176. https://doi.org/10.1037/0012-1649.40.2.162

Scott, R. M., & Fisher, C. (2012). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180. https://doi.org/10.1016/j.cognition.2011.10.010

Sehl, C. G., Friedman, O., & Denison, S. (2023). The social network: How people infer relationships from mutual connections. *Journal of Experimental Psychology: General*, *152*(4), 925–934. https://doi.org/10.1037/xge0001330

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480–498. https://doi.org/10.1111/j.1551-6709.2010.01158.x

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development*, *85*(5), 1821–1826. https://doi.org/10.1111/cdev.12247

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Téglás, E., & Bonatti, L. L. (2016). Infants anticipate probabilistic but not deterministic outcomes. *Cognition*, *157*, 227–236. https://doi.org/10.1016/j.cognition.2016.09.003

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, *104*(48), 19156–19159. https://doi.org/10.1073/pnas.0700271104

Téglás, E., Ibanez-Lillo, A., Costa, A., & Bonatti, L. L. (2015). Numerical representations and intuitions of probabilities at 12 months. *Developmental Science*, *18*(2), 183–193. https://doi.org/10.1111/desc.12196

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*(6033), 1054–1059. https://doi.org/10.1126/science.1196404

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855. https://doi.org/10.1016/j.cognition.2008.05.009

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2016.0056

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*(33), 13273–13278. https://doi.org/10.1073/pnas.0705369104

van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, *67*(4), 507–520. https://doi.org/10.1016/j.jml.2012.07.008

van der Kant, A., Männel, C., Paul, M., Friederici, A. D., Höhle, B., & Wartenburger, I. (2020). Linguistic and non-linguistic non-adjacent dependency learning in early development.

Developmental Cognitive Neuroscience, *45*, 100819. https://doi.org/10.1016/j.dcn.2020.100819

Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of Memory and Language*, *93*, 217–230. https://doi.org/10.1016/j.jml.2016.10.001

Vlach, H. A., & DeBrock, C. A. (2019). Statistics learned are statistics forgotten: Children's retention and retrieval of cross-situational word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 700–711. https://doi.org/10.1037/xlm0000611

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

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*(6), 1611–1617. https://doi.org/10.1037/a0016134

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent statistical learning of adjacent and nonadjacent dependencies. *Language Learning*, *66*(1), 8–30. https://doi.org/10.1111/lang.12137

Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: an action measure. *Developmental Science*, *18*(1), 175–182. https://doi.org/10.1111/desc.12208

Wang, F. H., & Mintz, T. H. (2018). Learning nonadjacent dependencies embedded in sentences of an artificial language: When learning breaks down. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *44*(4), 604–614. https://doi.org/10.1037/xlm0000483

Wang, F. H., & Trueswell, J. C. (2019). Spotting dalmatians: Children's ability to discover subordinate-level word meanings cross-situationally. *Cognitive Psychology*, *114*, 101226. https://doi.org/10.1016/j.cogpsych.2019.101226

Wellman, H. M., Kushnir, T., Xu, F., & Brink, K. A. (2016). Infants use statistical sampling to understand the psychological world. *Infancy*, *21*(5), 668–676. https://doi.org/10.1111/infa.12131

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van der Kant, A., Smith, K., & Rey, A. (2020). Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science*, *12*(3), 843–858. https://doi.org/10.1111/tops.12381

Wu, Y., Merrick, M., & Gweon, H. (2024). Expecting the unexpected: Infants use others' surprise to revise their own expectations. *Open Mind : Discoveries in Cognitive Science*, *8*, 67–83. https://doi.org/10.1162/opmi_a_00117

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*(1), 97–104. https://doi.org/10.1016/j.cognition.2009.04.006

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015. https://doi.org/10.1073/pnas.0704450105

Yermolayeva, Y., & Rakison, D. H. (2016). Seeing the unseen: Second-order correlation learning in 7- to 11-month-olds. *Cognition*, *152*, 87–100. https://doi.org/10.1016/j.cognition.2016.03.012

Younger, B. (1990). Infants' detection of correlations among feature categories. *Child Development*, *61*(3), 614–620. https://doi.org/10.2307/1130948

Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, *56*(6), 1574–1583. https://doi.org/10.2307/1130476

Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development*, *54*(4), 858–867. https://doi.org/10.2307/1129890

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, *57*(3), 803–815. https://doi.org/10.2307/1130356

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13), 2149–2165. https://doi.org/10.1016/j.neucom.2006.01.034

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180. https://doi.org/10.1111/j.1467-7687.2010.00958.x

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, *145*, 53–62. https://doi.org/10.1016/j.cognition.2015.07.013