of about 10 μg dl−1 during infancy and early childhood. The estimated mean IQ for that group, based on the combined evidence from the longitudinal cohort studies, is approximately 95. As illustrated in Figure 5, one consequence of this 5-point (5%) downward shift in IQ for the exposed group is a disproportionate (57%) increase in the number of children with IQ scores in the extremely low range (<70). An IQ test score less than 70 is consistent with the need to place a child in a special education program—unfortunate necessity for the child and an approximate doubling of the cost for his or her education. Similarly, this 5% downward shift in the average IQ of a population would cause a disproportionate 40% reduction in the number of children who score in the very superior range (IQ > 130). An IQ score of 130 is often a requirement for access to accelerated courses in high school. Thus, a small effect of lead can be very costly for individuals and for society as a whole.

Summary

Despite the long and cheerless history of lead’s detrimental effects on human health, the efforts of public health advocates over the past half-century have resulted in an impressive reduction in childhood exposure throughout most of the industrialized world. However, although lead exposure is almost entirely preventable, millions of children throughout the world continue to suffer the adverse effects of excessive exposure to this potent neurotoxin. The neurobehavioral effects of lead are subtle, but they can be detected in children with blood concentrations below the CDC definition of an elevated BLL, and they appear to be lasting. Moreover, efforts to restore cognitive function in lead-exposed children through medical treatment suggest little reason for optimism. The shared perspective of developmental psychologists and medical and public health specialists is that only through primary prevention of lead exposure during gestation, infancy, and childhood is it possible to protect children from this ancient threat to human health.

See also: Anger and Aggression; Attention; Brain Development; Brain Function; Cognitive Development; Developmental Disabilities: Cognitive; Emotion Regulation; Neurological Development; Social-Emotional Development Assessment; Safety and Childproofing.

Suggested Readings


Relevant Websites

http://www.asmalldoseof.org – A Small Dose of Toxicology by Steven Gilbert, A Small Dose of Lead.

Learning

R L Gómez, The University of Arizona, Tucson, AZ, USA 2008 Elsevier Inc. All rights reserved.

Glossary

Classical conditioning – A process of behavior modification by which a learner comes to respond to a previously neutral stimulus. When the neutral stimulus is repeatedly paired with an unconditioned stimulus, it begins to elicit the desired response even in the absence of the unconditioned stimulus. Discrimination – After familiarization with novel information, infants are tested for learning in terms of their ability to differentiate a stimulus consistent with their learning experience versus a stimulus.
### Introduction

Learning is a relatively stable change in behavior that results from exposure to a novel stimulus. Developmentalists have long been interested in learning because of its potentially important role in cognitive development. Learning is a fundamental process that operates in concert with other perceptual and cognitive processes, but the extent of its contribution to early cognition (in contrast to the contribution of biological constraints) is largely unknown. One goal of current research is to identify core infant learning mechanisms in an effort to better characterize the infant’s initial state. Before detailing the research to date, the author briefly reviews the history of learning as it has pertained to issues in child development.

The impetus behind human development, whether caused by nature or nurture, has been debated for centuries. John Locke (1632–1704) espoused the view that children were born with a tabula rasa (blank slate) and must therefore acquire all knowledge through experience. In contrast, Jean-Jacques Rousseau (1712–78) emphasized the role of innate knowledge in development. An emphasis on learning re-emerged in the last century with Pavlov (1849–1936) who showed with classical conditioning that an unconditioned response (a reflex such as salivation) could be trained to respond to a conditioned stimulus (CS; the sound of a bell) if the CS was paired with an unconditioned stimulus (US) such as food. Eventually, the food could be taken away to show that the CS was sufficient for producing the response. Based on these principles, John B. Watson (1878–1958) advocated a psychology of learning called behaviorism. This movement dominated theories of human development in the early half of the twentieth century. In particular, B. F. Skinner (1904–90) developed a theory of learning based on principles of operant conditioning, emphasizing the role of reinforcement and punishment in shaping specific behaviors. His work culminated in the book *Verbal Behavior*, which attempted to explain how a complex skill (such as language) could be acquired by principles of operant learning, where learners receive reinforcement for particular linguistic behaviors. However, in 1957, Noam Chomsky argued that Skinners’ theory was inadequate for explaining the complex rules that underlie human language, rules that enable children to understand sentences they have never heard and generate novel sentences. Chomsky also rejected the notion that the processes advocated by Skinner could explain how human children can acquire language so rapidly. Chomsky rejected the simple processes proposed by Skinner in favor of the notion that children are biologically prepared to acquire language such that fundamental principles of language are hard-wired into the circuitry of the brain.

Although Chomsky’s arguments were directly relevant to language acquisition, they reflected a paradigm shift

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>The number of times a unit occurs.</td>
</tr>
<tr>
<td>Generalization</td>
<td>A process involving abstraction away from specific stimulus materials that leads to perception of higher-order regularities or rules. Generalization enables learners to recognize new examples that are similar, but not identical, to previously encountered examples.</td>
</tr>
<tr>
<td>High-amplitude operant sucking procedure</td>
<td>Infants in this procedure are tested in a reclining seat, facing forward toward a colorful display. The infant sucks on a blind nipple (one without a hole) connected by a rubber hose to a pressure transducer that produces a signal on a polygraph machine. High-amplitude sucks (the top 33% of responses) are reinforced to a particular stimulus. After time, the infant’s sucks will diminish as the infant becomes familiar with the stimulus. At that point a new stimulus is introduced. An increase in the quantity of high-amplitude sucks is interpreted as the infant having noticed a difference in the old and new stimulus.</td>
</tr>
<tr>
<td>Joint probability</td>
<td>The probability of two units occurring together.</td>
</tr>
<tr>
<td>Observational learning</td>
<td>A process by which learners acquire behaviors by observing others then imitating what they have observed.</td>
</tr>
<tr>
<td>Operant conditioning</td>
<td>A process of behavior modification in which the probability of a specific behavior is increased by applying positive reinforcement after the occurrence of the behavior. The occurrence of a behavior can be decreased through negative reinforcement.</td>
</tr>
<tr>
<td>Statistical learning</td>
<td>The discovery of structure in perceptual information in terms of the statistical properties of perceptual units. In language the perceptual units can be phonetic segments, syllables, or words. Taking English as an example, certain phonetic segments (e.g., the -ng sound in any word ending in ‘-ing’) are statistically probable at the ends of words, but are nonexistent at the beginnings of words, whereas this segment can occur with high statistical probability at the beginnings of words in some languages.</td>
</tr>
<tr>
<td>Transitional probability</td>
<td>The probability of the occurrence of one unit given another (the probability of event B given event A is the joint probability of event A and event B divided by the probability of event A).</td>
</tr>
</tbody>
</table>

*Encyclopedia of Infant and Early Childhood Development* (2008), vol. 2, pp. 213-224
spreading across the field of psychology at that time. The notion that human behavior could be explained in terms of associative learning principles was widely abandoned in favor of theories emphasizing underlying cognitive processes. Other scientists also argued that simple associative learning principles were inadequate for explaining human behavior. For instance, Lashley (1890–1958) pointed out that many human endeavors, such as language, involve hierarchically organized behaviors as opposed to chains of stimulus response associations. Additionally, in 1961, Albert Bandura showed that children could learn through vicarious observation (observational learning) without need for direct reinforcement. In the area of language acquisition, learning theories were replaced by theories emphasizing inborn principles and parameters (where principles are rules common to all languages and parameters are rules capturing the possible differences between languages).

In recent years the tide has turned back toward an emphasis on learning. This has been driven by the success of computer models demonstrating that simple associative systems can accomplish fairly complex learning, as well as broad acceptance of the fact that the perceptual input to learners is rich in statistical information. Finally, experimental work with infants has led to exciting discoveries regarding early learning abilities suggesting that human learning is more powerful than originally conceived in behaviorist theory. This experimental work, and precursors to it, is the focus of this article.

The remaining sections summarize some of the early work on learning before detailing recent experimental work on infants’ ability to detect statistical structure. The article ends by posing open questions and challenges for learning research.

**Early Work on Learning**

Much of the earliest research on infant learning was methodological in nature. Driving questions were whether human infants were even capable of learning and if so, what types of learning they would exhibit. Researchers were also interested in determining when learning might begin to take place in development. Before these questions could be answered, fundamental methodological details had to be worked out. For instance, it was important to rule out the possibility that a change in behavior reflected acclimatization to a stimulus or a response to a particularly appealing reward, behaviors that mimic but do not reflect learning. Additionally, a number of early studies failed because researchers did not know how much or how little exposure to a stimulus would promote learning or which types of rewards were needed.

The earliest studies investigated whether infants could be classically conditioned to associate a conditioned stimulus with an unconditioned response by pairing a CS with an UCS. In 1920, John B Watson and Rosalie Raynor exposed an 11-month-old child named Albert to an UCS (a loud sound) each time he touched a CS (a white rat). The loud sound caused Albert to cry and withdraw his hand, so that subsequently, merely seeing the rat led to the same behavior, demonstrating memory of a learned stimulus–response pairing.

Infants just a few hours old can be classically conditioned also. In 1984, E M Blass and colleagues followed a stroke on the forehead (the CS) by immediate oral delivery of sucrose to infants through a glass pipette (the UCS). A control group was exposed to the same number of CS–UCS pairings, but the time interval between the pairings varied across trials. Another control group received only the UCS in order to rule out the possibility that repeated deliveries of sucrose itself would result in a change in behavior. During an extinction phase where infants received the CS but not the sucrose delivery, the experimental group cried when the sucrose was not delivered whereas infants in the two control groups did not. All three groups experienced withdrawal of sucrose, so this in itself could not explain the experimental group's behavior. Their behavior could only be explained in terms of their having learned the predictive CS–UCS relationship. This study, and others in this general vein, established that infants could be classically conditioned from birth.

Researchers were also interested in determining whether infants were capable of learning using operant principles. Operant conditioning involves reinforcing a naturally occurring response to increase or decrease the rate of that response. For instance, positively reinforcing a response leads to an increased rate of that behavior. This is in contrast to classical conditioning, which involves a learned association between a CS and UCS. The first study to obtain evidence for operant conditioning in newborns was conducted in 1966 by Siqueland and Lipsett. They paired differential auditory stimuli with an unconditioned tactile stimulus (stroking the infant's cheek). The stroking produces an unconditioned rooting reflex to that side and by necessity a head turn. In their study, when the pairing of the positive auditory stimulus (a buzzer) and the tactile stimulus produced a head turn, this was always followed by a positive reinforcer (administration of sugar water). However, a head-turn in response to the pairing of the tactile stimulus and the negative auditory stimulus (a tone) was never reinforced. Responding to the presence of the positive auditory stimulus increased over time whereas responding to the negative auditory stimulus decreased suggesting that infants were discriminating between the two types of auditory stimuli based on operant conditioning.

An operant paradigm that has been used extensively to study learning and memory in infants since then, is the mobile conjugate reinforcement procedure developed by Carolyn Rovee-Collier in 1969. In this procedure, an infant is placed on his back in a crib beneath a mobile.
A ribbon runs from a suspension hook on the mobile to the infant's ankle (in the reinforcing condition) or from a hook that will not move the mobile (the nonreinforcing condition). First, a measure of baseline kicking is obtained by observing the number of kicks in the nonreinforcing condition (a pretest). After obtaining a baseline measure, the ribbon is tied to the suspension hook so that when the infant kicks he moves the mobile, resulting in a high rate of kicking and attention to the mobile. Memory of the learned experience can then be assessed after various delays by positioning the infant beneath the mobile, attaching his ankle to the nonreinforcing hook (to prevent additional learning), and measuring the number of kicks he produces. Learning is evidenced by a greater rate of kicking relative to baseline when infants are tested with the same mobile as compared to a different one. This procedure was used to show that 2-month-old infants could remember a learning experience occurring 24 h earlier.

Infants and young children do not necessarily need reinforcement to learn. In a seminal study of observational learning reported in 1961, Bandura showed that 4-year-olds who simply observed an adult beating up a Bobo doll were more likely to direct similar behaviors toward the doll in subsequent play than were children in a control condition who did not see an adult exhibiting such behavior. This type of learning is referred to as observational because learners imitate an observed behavior with no stimulus–response pairing or any kind of reinforcement. Observational learning does not appear to have a lower age limit. In deferred imitation, another type of observational learning, an experimenter models a sequence of actions and the infant is later tested on the ability to reproduce the behavior. In 1992, Patricia Bauer and Jean Mandler showed that infants as young as 11.5 months of age can learn and later imitate novel actions in an event sequence like making a rattle. Events were simple, consisting of two or three actions, but infants readily reproduced the actions in their correct order. Interestingly, the type of sequence matters, such that arbitrary sequences involving a series of events that are not causally related (like banging, turning, and stacking a ring on a dowel) are much more difficult to learn than causally related actions that require actions to be performed in a certain order (such as making a rattle by putting a ball in a paper cup, joining the mouth of the cup with the mouth of another paper cup, and shaking). In a different experimental paradigm demonstrating observational learning, Rachel Barr and colleagues in 1996 showed that infants as young as 6 months of age can remember and imitate portions of a sequence they have observed being modeled. The sequence, consisting of removing a mitten from a puppet's hand, shaking the mitten (causing a bell inside the mitten to sound), and placing the mitten back on the puppet, was imitated by the infants after a 24 h delay. In most cases, exposure was very brief. The event sequences were modeled just twice with 11.5-month-olds and six times with the 6-month-olds. Additionally, infants as young as 2 months of age can engage in learning without feedback as shown by Naomi Wentworth and Marshall Haith in a study reported in 1992. Infants were exposed to an alternating left–right pattern of visually presented pictures. On one side (left or right) the picture was always the same and on the other side the picture varied. Infants showed learning of the visual content of the picture as evidenced by their tendency to anticipate the location of the stable picture and to respond to it more quickly as compared to the location of the unstable picture.

Given these findings it becomes important to ask just how early learning occurs in human development. Given the results detailed above, it is reasonable to think that learning may occur as soon as infants are able to process sensory information, indeed that they might even begin learning in utero.

One of the earliest indications of fetal learning was the finding that newborns prefer their mother's voice to that of another female speaker. They also prefer sentences from their native language to sentences from another language. Passages read in French produced higher sucking rates in French newborns than passages read in Russian. Other studies have shown that these preferences are not specific to French. Therefore such preferences must be shaped by prenatal experience with maternal speech. What might infants be learning? We know from intrauterine recordings that low-frequency components of maternal speech, including its rhythmic qualities are audible in utero and infants born prematurely at 24 weeks are able to react to sounds, raising the possibility that learning may begin this early.

In 1986, Anthony DeCasper and Melanie Spence showed that newborns, whose mothers read a passage aloud each day during the last 6 weeks of pregnancy, were able to discriminate the passage from an unfamiliar one at birth. Two-day-olds were tested, using a high-amplitude operant sucking procedure, to see whether the familiar passage would be more reinforcing than an unfamiliar one, even when read in another woman’s voice. It was, suggesting that infants had learned features from their training passage involving its prosodic (or rhythmic) qualities over and above features specific to their mother's voice. The passages were not read aloud before the newborns were tested, and thus learning must have occurred in utero. A later study, by DeCasper and colleagues, used heart rate as a dependent measure to test learning in 37-week-old fetuses. Mothers recited one of two rhymes out loud, once a day, over 4 weeks. At 37 weeks' gestational age the fetuses were stimulated with recordings of the familiar and unfamiliar rhymes. The familiar rhyme elicited a decrease in fetal heart rate, whereas the unfamiliar one did not, suggesting discrimination of the two passages, and hence learning.
Statistical Learning

Despite the success of these early studies, learning was not studied in its own right for many years. In the past 10 years, however, researchers have begun to document learning of statistical regularities in perceptual input. Statistical learning, the ability to track probabilistic patterns, is an important mechanism because probabilistic structure abounds in the information available to our senses.

Statistical structure can take many forms including the frequency of individual units, joint probability, or the transitional probability of one unit given another. Joint probability is defined as the probability of two units occurring together. Transitional probability is the probability of the occurrence of one unit given another (the probability of event B given event A is the joint probability of event A and event B divided by the probability of event A). There are other forms of statistical structure but common to all forms is the requirement that units occur with some regularity that lends itself to mathematical description (and presumably also to computation).

As an example of statistical learning, a problem infants must solve in speech perception is identifying words in running speech. This is a daunting task because words are rarely demarcated by pauses. One popular proposal is that infants might learn a few words spoken in isolation and use the boundaries of these known words to discover the boundaries of words that co-occur. However, useful information is conveyed in the higher transitional probabilities occurring between syllables in words as compared to the lower probabilities of syllables spanning words. In a landmark study reported in 1996, Jenny Saffran, Richard Aslin, and Elissa Newport showed that very young infants are able to capitalize on this kind of information. For instance, in the phrase 'pretty baby', the syllables 'ba' and 'by' occur within a word, whereas 'ty' and 'ba' span words. As such 'by' is more highly predicted by 'ba' than 'ba' is by 'ty'. Saffran and colleagues implemented this idea experimentally by exposing 8-month-old infants to continuously running syllables such as tupirotiladogolabudabikutupirodabiku... where tupiro, dabiku, tilado, and golabu were words that were presented in random order. For instance, in the first 4 syllables in this particular example 'tu pi' and 'pi ro' are syllables within words, whereas 'ro' and 'ti' are syllables spanning words.

Although statistical learning occurs in multiple modalities, much of the research has been conducted in the context of language acquisition. In these studies infants are exposed to artificial language materials they have never heard before. In natural language cues are correlated making it difficult to pinpoint the locus of learning, but artificial languages are devised with particular learning cues in mind, enabling more precise control over the input to learners. Artificial languages also control for prior exposure and therefore provide insights into learning capabilities at a given point in time instead of tapping a developing sensitivity midstream.

A typical study involves a familiarization phase followed by a test. Length of familiarization in infant studies varies from 2 to 3 min. Most studies counterbalance stimulus materials so that half of the infants are exposed to one version of the language and half to another version. At test, infants are exposed to strings from both versions so that what is grammatical for one group is ungrammatical for the other (e.g., version A strings violate the constraints of version B and vice versa). This ensures that the structure of the language, instead of something idiosyncratic about the sound tokens, is responsible for learning. Infants are tested using procedures for recording the amount of time they attend to different stimulus types. If learning has occurred, a group of infants should listen differentially to strings conforming to their training language versus strings that do not conform.

Statistical learning appears to play a role in the formation of speech categories, in the identification of word-like units in running speech, in the ability to track adjacent and nonadjacent word dependencies, and in generalizations involving the acquisition of categories and their relationships in speech. Learning research has also investigated whether infants can learn in the presence of noisy input and whether they can use prior experience to bootstrap learning of more difficult patterns from simpler ones. That research, as well as research conducted in the visual modality, will be summarized with the goal of describing a range of statistical learning studies including research conducted in the author's laboratory.

The Role of Frequency in the Formation of Speech Categories

Infants are sensitive to a wide range of speech contrasts very early in development. An example is the contrast between the consonants /b/ and /p/. Very young infants even discriminate contrasts that do not occur in their native language but by 8–10 months of age, discrimination only occurs for those contrasts in their native language. This finding is widely recognized as experience dependent, but until recently we did not have a very good candidate for the process underlying this change. One possibility, found in the literature on statistical learning, is that infants use statistical information to home in on phoneme categories. Although phonemes may vary acoustically along a dimension, such variation is not random. It patterns bi-modally such that the most frequent tokens of one category occur at one end of a dimension and tokens from another category occur at the other end. This is in contrast to a unimodal distribution in which the most frequent tokens occur between two ends of a continuum. Will distributions with these different characteristics...
influence infants’ ability to distinguish speech contrasts such that exposure to bi-modal distributions results in discrimination of a speech contrast, whereas exposure to a unimodal distribution prevents discrimination?

In the earliest infant experiment exploring this hypothesis in 2002, Jessica Maye, Janet Werker, and LouAnn Gerken familiarized 6- and 8-month-olds with one of two distributions of eight speech sounds on a /da/-/ta/ continuum (the voiced unaspirated /d/ in day and the voiceless unaspirated /t/ in stay) (Figure 1). Infants this age can make this discrimination, but if perception of speech sounds is malleable then exposure to a unimodal distribution should interfere with discrimination whereas exposure to a bi-modal distribution should preserve it. This pattern of findings would support the proposal that sensitivity to the frequency distributions of speech sounds is instrumental in learning. Infants in both unimodal and bi-modal conditions heard the same eight speech sounds along the continuum, but those in the bi-modal distribution condition heard speech sounds near the end (2 and 7) most frequently whereas infants in the unimodal distribution condition heard the middle speech sounds (4 and 5) most often. After familiarization, infants were tested on their ability to discriminate alternating speech sounds (the endpoints 1 and 8) from nonalternating ones (repeats of speech sounds 3 or 6). Only infants in the bi-modal condition discriminated alternating from nonalternating speech sounds, supporting the idea that exposure to the bimodal distribution led to preservation of two categories of speech sounds whereas exposure to the unimodal distribution resulted in the formation of one. In subsequent studies Jessica Maye and colleagues have shown that exposure to a bi-modal frequency distribution can also enable detection of an initially undetectable speech sound contrast. Thus, the ability to learn the frequency characteristics of speech sounds can blur distinctions between previously known categories or they can enable the formation of new ones.

The Role of Joint Probability and Transitional Probabilities in Visual Statistical Learning

Infants are also able to keep track of joint probabilities in learning the most frequent associations between objects in visual sequences. Natasha Kirkham, Jonathan Slemmer, and Scott Johnson in a study reported in 2002 familiarized infants, ages 2, 5, and 8 months, to a continuous series of objects, presented one at a time in sequence. The stimuli were six shapes (e.g., turquoise square, blue cross, yellow circle, pink diamond, green triangle, red octagon, where the square was always followed by the cross, the circle by the diamond, and so on). The infants were then tested to see if they would discriminate legal pairs of objects from illegal ones (e.g., turquoise square followed by the pink diamond) by manifesting longer looking times to the illegal combinations. All three age groups showed discrimination, demonstrating early sensitivity to joint probability in visually presented stimuli (Figure 2).

In a study reported the same year, Josef Fiser and Richard Aslin showed that infants can track transitional probabilities in objects co-occurring in visual scenes by 9 months of age, but because these studies have not yet

Figure 1  Bimodal versus unimodal distributions of [da]–[ta] stimuli during familiarization in a study of speech-category formation. The continuum of speech sounds is shown on the abscissa with Token 1 corresponding to [da] and Token 8 corresponding to [ta]. The ordinate axis plots the number of times each stimulus occurred during the familiarization phase. The presentation frequency for infants exposed to a bimodal presentation is depicted by the dotted line, and for the unimodal presentation by the solid line. Reproduced from Maye J, Werker J, and Gerken LA (2002) Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82: B101–B111, with permission from Elsevier.

Figure 2  A depiction of stimuli used in a visual learning study to assess learning of joint probability. Reproduced from Kirkham NZ, Slemmer JA, and Johnson SP (2002) Visual statistical learning in infancy: Evidence for a domain general learning mechanism. Cognition 83: B35–B42, with permission from Elsevier.
The Role of Transitional Probabilities in Segmenting Words in Continuous Speech

As described briefly earlier, infants can also keep track of more complex statistics, such as transitional probabilities in sequences of syllables, and they can use this information to discover word boundaries. Such learning has been tested in 7- and 8-month-olds by exposing them to continuous streams of four randomly ordered three-syllable words (e.g., tupiro, dabiku, tilado, golabu in a string such as tupirotiladogolabudabikutupiroadabikugolabu...). Although syllable pairs within words occurred with identical joint frequency to those occurring between words, higher transitional probabilities for syllables within words versus those spanning words, can provide cues to word boundaries. Take a phrase like ‘naughty puppy’. The syllable transition in ‘naugh-ty’ has a higher transitional probability than the transition ‘ty-pu’ because ‘naugh’ in the word ‘naughty’, is more likely to predict ‘ty’ than ‘ty’ is to predict ‘pu’. Jenny Saffran has shown in a series of studies that infants are able to use the differences in transitional probabilities within-words versus between-words to identify word boundaries in running speech such that they show longer listening times to part-words (words that span word boundaries, e.g., bikuti) than to words (e.g., tupiro). This pattern of listening makes sense if we assume that once infants have learned the words, they will find them less interesting, and hence will listen to them for a shorter time than they will listen to novel part-words. Infants can also track transitional probabilities in tone sequences, showing that such learning is not confined to linguistic stimuli.

Even though such learning is not specific to a particular context if it is relevant for language, one might predict that when infants segment syllable sequences (as opposed to tone sequences) they should treat these as candidates for words. This would be evidenced by a greater likelihood of discriminating words and part-words in the context of English sentence frames. If the segmented syllable strings have a word-like status, then we might expect infants to listen longer when words (as opposed to part-words) are embedded in English frames (“I like my tupiro” vs. “I like my bikuti”). This is because if syllable strings are ‘candidate words’ they should sound more natural in the context of familiar natural language frames than part-words, and infants should listen longer. However, it could be that infants are doing nothing more then discriminating legal from illegal patterns of syllables (tupiro vs. bikuti). Thus, an important control is one where the words and part-words are embedded in nonsense frames (such as “Zy fike ny tupiro”). If infants are treating words and part-words alike, then performance should be the same in English and nonsense frame types: as in English frames they should listen longer to nonsense strings with embedded ‘words’ as compared to part-words. Jenny Saffran found that 8-month-olds only discriminated words and part-words in English sentence frames suggesting that they treat the words they segment as more language-like than the part-words. Another clue that infants treat segmented nonsense words as candidates for real words is the fact that older 17-month-olds more readily map newly segmented words onto word meanings than part-words onto word meanings. Katharine Graf Estes, Julia Evan, Martha Alibali, and Jenny Saffran first exposed infants to a continuous stream of nonsense words, with transitional probabilities as the only cue to word boundaries. For half of the infants, two of the nonsense words were paired with visually presented objects. For the remaining infants the word forms used to label the objects were part-words. If the segmented units are treated as candidates for real words, infants should be more likely to map the word labels to objects than the part-word labels, and they were.

How Infants Might Tune in to Long Distance Dependencies

The research summarized early on shows that infants are adept at tracking sequential dependencies between adjacent elements. Indeed, this tendency occurs across species (for human, nonhuman primates, and rats), across development (in infants and adults), and even under incidental learning conditions, suggesting that it may be a default. However, many dependencies occur across longer distances, especially in language. Some examples are dependencies between auxiliaries and inflectional morphemes (e.g., is quickly running), and between nouns and verbs in number and tense agreement (The boys in the tree are laughing). If the tendency to track adjacent structure is pervasive, how might learners begin to track more remote dependencies in sequential structure?

Rebecca Gómez investigated this question in 2002 by familiarizing infants with one of two artificial languages. Language 1 sentences followed the patterns aXb or cXd (e.g., pel-wadim-jic, vot-kicey-rud). In Language 2 the relationship between the first and third elements was reversed to aXd or cXb such that pel sentences ended with rud, and vot sentences ended with jic (pel-wadim-rud, vot-kicey-jic). The a, b, c, d, and X elements were restricted to the same positions in the two languages and adjacent dependencies were identical (aX occurred in both languages as did Xd) so that strings could only be distinguished by learning the relationships between the nonadjacent first and third words. The size of the pool from which the middle element was drawn was also
manipulated (set-size = 3, 12, or 24) while holding frequency of exposure to particular nonadjacent dependencies constant (see Figure 3). The purpose of this manipulation was to determine whether high variability in the middle element would lead to better perception of nonadjacent dependencies even though these were equally frequent in all three set-size conditions. The motivation for the variability manipulation was the observation that long-distance dependencies in natural language between frequently occurring words such as ‘is’ and ‘-ing’ occur with a large number of verbs as opposed to a small number (e.g., ‘is running’, ‘is playing’, ‘is sleeping’). Perhaps the variability of verbs contributes to detection of the long-distance dependencies. Infants as young as 15 months of age acquired the nonadjacent dependency when the intervening element came from a set of 24 possible words, but not when intervening set size was smaller (3 or 12). One might have expected the added noise to impede learning; however, high variability appeared to increase the perceptual salience of the nonadjacent words compared to the middle word, resulting in learning.

Generalization in Learning

Rebecca Gómez and LouAnn Gerken in a study published in 1999 familiarized 12-month-olds with strings from an artificial grammar in one vocabulary and tested them on strings in entirely new vocabulary (e.g., infants heard FIM-SOG-FIM-FIM-TUP and were tested on VOT-PEL-PEL-JIC). Although the constraints on word-ordering remained the same between training and test, vocabulary did not. Infants could not distinguish the two grammars based on the variability manipulation because of the change in vocabulary. The infants made transitional probabilities between remembered word pairs.

Infants discriminated strings with the training pattern from those with a different pattern despite a change in vocabulary (e.g., ba-po-ba vs. ba-po-po).

These are important findings but it is crucial to ask to what degree such learning extends to real-world problems such as those faced by children learning language? The infant abstraction abilities documented in these studies are dependent on learning patterns of repeating and alternating elements (e.g., ABB, ABA, ABC), a form of generalization that is fairly limited in language. Whereas recognizing ba-po-ba and ko-ga-ko as instances of the pattern ABA entails noting that the first and last syllables in sequence are physically identical, most linguistic generalizations involve operations over variables that are not perceptually bound. If we compare the pattern-based representation ABA to the category-based representation Noun Verb Noun, abstracting ABA from ba-po-ba involves noting that the first and third elements in a sequence are physically identical, and thus recognition is perceptually bound. In contrast, the Noun Verb Noun relation holds over abstract categories that do not rely on perceptual identity. “Muffy drinks milk” and “John loves books” share the same category-based Noun Verb Noun structure, despite the obvious physical dissimilarities between category members such as milk and books. Given this observation, researchers have begun to examine learning involving abstract variables.

The ability to perceive category relationships among words in strings is essential to linguistic productivity. For instance, an English speaker must be able to generalize from a novel string like “The pleg mooped” to “Is the pleg mooping?” Generalization is extremely powerful – once a novel word is categorized children can automatically apply all of the syntactic constraints associated with other words in its category. How do children achieve such generalization? Although semantic information most
certainly factors into such learning, infants must parse syntactic categories in the speech they hear in order to link them with semantic referents. This involves learning phonological regularities within words of a category (e.g., noun or verb), and co-occurrence relations between categories (e.g., determiner and noun). An infant who is able to parse the relevant categories in speech has a leg up on the ultimate task of mapping meaning to syntactic phrases.

One way to investigate this kind of learning with artificial languages is to give categories arbitrary labels such as $a$, $X$, $b$, and $Y$. Words from these categories are then combined to form legal phrases. For instance, $aX$ and $bY$ might be legal in a language whereas $aY$ and $bX$ are not.

To give an example, imagine that $a$-elements correspond to ‘a’ and ‘the’ and $b$-elements to ‘will’ and ‘can’ (see Table 1). Infants will only be successful at discriminating a new legal phrase (e.g., ‘a cat’) from an illegal one (‘a eat’) if they have learned that $a$-elements go with nouns (the $X$s), but not with verbs (the $Y$s). As in natural language, where the set of determiners has very few members and the set of nouns is large, $a$- and $b$-categories have fewer members than $X$s and $Y$s. Also, in natural language nouns and verbs tend to have distinguishing phonological features. For instance, in English nouns tend to have more syllables than verbs. Therefore, it is important to incorporate such phonological features into the $aX$ $bY$ language.

Rebecca Gómez and Laura Lakusta in 2004 reported research asking whether 12-month-olds could learn the relationship between specific $a$- and $b$-words and features defining $X$- and $Y$-categories. During training infants heard one of two training languages. One language consisted of $aX$ and $bY$ pairings, the other of $aY$ and $bX$ pairs. $X$s were two-syllable words and $Y$s were one syllable so that infants could use syllable number as a feature for distinguishing $X$- and $Y$-categories. At test, for example, infants trained on $aX$ and $bY$ pairings had to discriminate these from $aY$ and $bX$ pairs. However, in order to assess generalization, all $X$- and $Y$-words were novel. The infants successfully discriminated the legal from illegal pairs, suggesting that they had learned the relationships between the $a$- and $b$-elements and the abstract feature characterizing $X$- and $Y$-words (syllable number). Similar learning may occur in natural language, where children exposed to English may pick up on distributional regularities distinguishing nouns and verbs and link these to specific function words.

Although this study was important for determining whether 12-month-olds could learn to associate $a$- and $b$-words with different category features, the next step is to determine whether infants can form categories of $a$- and $b$-words by themselves. This is an important ability in natural language. For instance, once children form the category of determiner, if they hear a novel word predicted by ‘the’ they will know to use ‘a’ with that word also. At a more general level, this kind of learning feeds into the ability to use the occurrence of words of one category to label the syntactic category of a following word (e.g., using the presence of a determiner in the phrase ‘a dax’ to label the novel word ‘dax’ as a member of the noun category).

How might this kind of learning be realized in our artificial language paradigm? After learners have associated $a$- and $b$-words with $X$/$Y$-cues, they might then go on to categorize individual $a$- or $b$-elements based on their joint association with particular $X$- and $Y$-cues (for instance, in natural language, children would form a category containing ‘a’ and ‘the’ based on features that tend to occur with nouns and not verbs; see Table 2). Once function-word categories are formed, children can rely on memory for a phrase they have heard (e.g., ‘the dax’) and the fact that ‘the’ and ‘a’ are in the same category to make an inference about a phrase they have not heard (e.g., ‘a dax’), regardless of whether the novel word has a defining feature.

LouAnn Gerken and colleagues in 2005 investigated such learning with 17-month-old American infants by exposing them to Russian words in which feminine word stems appeared with the case endings –o(j and –u and masculine word stems appeared with the case endings –ya

**Table 1** A paradigm for investigating category abstraction

<table>
<thead>
<tr>
<th>Natural language example</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$ = the</td>
<td>boy</td>
<td>girl</td>
<td>ball</td>
<td>dog</td>
<td>cat</td>
<td>car</td>
<td>car</td>
</tr>
<tr>
<td>$a_2$ = a</td>
<td>boy</td>
<td>girl</td>
<td>ball</td>
<td>dog</td>
<td>cat</td>
<td>car</td>
<td>car</td>
</tr>
<tr>
<td>$b_1$ = will</td>
<td>jump</td>
<td>run</td>
<td>play</td>
<td>sleep</td>
<td>eat</td>
<td>wait</td>
<td>wait</td>
</tr>
<tr>
<td>$b_2$ = can</td>
<td>jump</td>
<td>run</td>
<td>play</td>
<td>sleep</td>
<td>eat</td>
<td>wait</td>
<td>wait</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Artificial language example</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$ = alt</td>
<td>coomo</td>
<td>fengle</td>
<td>kicey</td>
<td>loga</td>
<td>paylig</td>
<td>wazil</td>
</tr>
<tr>
<td>$a_2$ = ush</td>
<td>coomo</td>
<td>fengle</td>
<td>kicey</td>
<td>loga</td>
<td>paylig</td>
<td>wazil</td>
</tr>
<tr>
<td>$a_1$ = ong</td>
<td>deech</td>
<td>ghope</td>
<td>jic</td>
<td>skige</td>
<td>vabe</td>
<td>tam</td>
</tr>
<tr>
<td>$a_2$ = erd</td>
<td>deech</td>
<td>ghope</td>
<td>jic</td>
<td>skige</td>
<td>vabe</td>
<td>tam</td>
</tr>
</tbody>
</table>

The top half of the table demonstrates the paradigm with natural stimuli with phrases like ‘the boy’ and ‘will jump’. The bottom half of the table demonstrates the paradigm with artificial language stimuli. Infants are familiarized with phrases like ‘alt fengle’ and ‘erd ghope’ then are tested to see if they will generalize correctly to strings with new $X$ and $Y$ words (an example of a new $aX$ string is ‘alt roosa’ and a new $bY$ string is ‘erd pel’).
2. They can then form a functor-like word category of ‘a-elements’ based on the joint association of individual a-words with particular X-features. For instance, in English ‘the’ and ‘a’ belong to the category of determiners.

3. Once function-word categories are formed, children can rely on memory for a phrase they have heard (e.g., ‘the dax’) and the fact that ‘a’ is in the same category as ‘the’ to make an inference about a phrase they have not heard (e.g., ‘a dax’), regardless of whether the novel word has a defining feature.

Table 2  Steps in syntactic category-based abstraction

1. Learners associate a-words with X-features and b-words with Y-features.
2. They can then form a functor-like word category of ‘a-elements’ based on the joint association of individual a-words with particular X-features. For instance, in English ‘the’ and ‘a’ belong to the category of determiners.
3. Once function-word categories are formed, children can rely on memory for a phrase they have heard (e.g., ‘the dax’) and the fact that ‘a’ is in the same category as ‘the’ to make an inference about a phrase they have not heard (e.g., ‘a dax’), regardless of whether the novel word has a defining feature.

Current Issues

The results of the studies in the growing literature on infant statistical learning reveal precocious learning abilities in young infants. The findings are intriguing and raise more general questions. For instance, can infants learn any kind of predictable structure or are there limits on their ability to detect statistical structure? Are the changes lasting or are we only measuring short-term discrimination? If learners are changed by their experiences, do prior learning experiences shape subsequent ones? And finally, does statistical learning observed in the lab have any connection to learning in the real world?

The first question has to do with the robustness of infant learning, especially when the information in perceptual input is noisy. In language all children are exposed to inconsistencies of one type or another during acquisition, in adults’ informal speech, in children’s own ungrammatical utterances, and in the ungrammatical utterances of other learners (such as playmates and siblings). Inconsistencies also occur naturally in language, for instance, in English the degree to which verbs take the regular -ed ending for the past tense, or in Spanish the extent to which feminine nouns end in -a. Other instances of noise in linguistic input are less widespread, such as when children are exposed to nonnative language input (a deaf child who is exposed to American Sign Language through a hearing parent who has not achieved proficiency in this language). In all of these instances, children must distinguish grammatical from ungrammatical instances, and they must generalize beyond the data to which they are exposed, making it important to ask how well infants learn on exposure to probabilistic structure.

Rebecca Gómez and Laura Lakusta investigated this question by familiarizing 12-month-olds with artificial languages with three levels of probabilistic structure. In the 100/0-condition all of the training strings were from the infants’ ‘predominant’ training language. In the 83/17-condition, approximately 83% of the training strings were from the predominant language (the remaining 17% of the strings were from the other language and thus constituted noise). In the 67/33-condition, the split between the predominant and nonpredominant training languages was 67% and 33%. Infants in the 100/0 and 83/17 conditions learned equally well, whereas learning was diminished in the 67/33-condition. The findings suggested that infants are able to track regularities in probabilistic input even when the regularities do not occur with perfect probability (as was the case with the 83/17 ratio) and so infant-learning is robust to some noise. However, learning does need to be based on some minimum degree of regularity, as demonstrated by the fact that infants in the 67/33-condition failed to learn.

Another question has to do with determining whether the learning observed reflects a permanent change. Otherwise, statistical learning studies may simply be registering acclimation to a particular stimulus. Familiarization in statistical learning studies is typically brief (3 min or less) and testing is immediate. Given that very young infants show forgetting of short-term memory after a 15 s delay it is important to determine whether discrimination extends past this window. Very little work has explored long-term memory of a brief learning experience, but recent studies by Rebecca Gómez, Richard Bootzin, Almut Hupbach, and Lynn Nadel show retention of an artificial language after delays of 4 and 24 h. These findings are important for ruling out the possibility that statistical learning studies only measure short-term effects; however, more information is needed regarding how memories for statistical patterns persist and how they affect later learning experiences over the long term. Should parents worry that exposing their children
to artificial languages might affect their natural language learning? They should not. Exposure in the lab is brief in comparison to the overwhelming experience children have with their natural language, and although infants may remember their learning experiences, their learning of natural language will take precedence over information they have acquired in the laboratory.

A third question has to do with whether and how learning at one point in time impacts learning at another point. Jill Lany and Rebecca Gómez recently investigated this question in the context of learning of categories in sequential structure. They familiarized 12-month-olds with aX and bY strings where X- and Y-elements were distinguished by different morphological endings (e.g., -ee or -oo). Infants had to learn that a-elements went with Xs and not Ys (and vice versa for b-elements). After familiarization with the aX/bY structure the infants were able to detect the a-X and b-Y relationships in a more complex language involving long-distance dependencies (e.g., in acX and bcY sentences). This language was particularly challenging because the intervening c-element required the infants to track nonadjacent dependencies between a- and X- and b- and Y-elements. Infants with prior experience with consistent aX and bY pairings were able to generalize to the nonadjacent acX/bcY structure over infants in a control group. This finding is important for showing how infants might scaffold learning of complex structure from learning of more simple forms and is particularly significant because of previous work showing that infants this age are unable to track nonadjacent dependencies.

A fourth question has to do with whether the learning observed in the laboratory scales up to real-life learning. One way to determine this is to ask whether the output of learning can be used as input to real-life learning processes. As discussed previously, Jenny Saffran and her colleagues have addressed this issue in the context of language acquisition in several studies by showing that infants prefer to listen to newly segmented artificial words in the context of natural language sentences frames (as opposed to nonsense frames), and also more readily learn a mapping between these newly segmented words and novel objects (as opposed to a mapping of a nonword and an object). Another approach is to determine whether statistical learning and real-life learning show similar developmental trajectories. Although learning occurring in the real world is far more complex than that observed in the lab, similar developmental trajectories would be partial support for a shared process. Initial evidence for this comes from work by Rebecca Gómez and Jessica Maye showing that the ability to detect long-distance dependencies in an artificial language comes online at 15–18 months, roughly the same time infants begin to detect long-distance dependencies in natural language.

Finally, certain kinds of statistical learning (frequency, joint probability, and transitional probability learning) occur across a broad range of organisms and across different modalities. This suggests that statistical learning may be a very general process and raises questions about how it interacts with other processes and mechanisms known to be involved in memory change. One such mechanism is sleep. New studies with adults show that sleep is instrumental in memory consolidation such that it enhances memory, improves generalization, and also leads to qualitative change. A recent study reported by Rebecca Gómez, Richard Bootzin, and Lynn Nadel tested the effects of sleep on infant learning. Infants who napped in a 4-h interval between familiarization and test were able to generalize a rule in an artificial language. In contrast, infants who did not nap had excellent memory for the strings of the artificial language but they did not generalize. Generalization is a critical form learning that results in greater flexibility. Such learning plays an essential role in cognitive development by sustaining sensitivity to previously encountered information, while enabling learners to generalize to novel cases.

Another important mechanism involved in memory change is memory reconsolidation. According to the literature on this phenomenon, memory is much more dynamic than previously thought. Previous research suggested that memories become crystallized as the result of a consolidation process, varying only in their access. But recent evidence shows that consolidated memories are open to change because when they are accessed, they go into a labile state. Once in this state, memories can be enhanced, altered, or overridden depending on new information encountered. Studies with rats and humans have found that a consolidated memory can be overwritten by new information when exposure to it follows memory reactivation. When the memory is not reactivated, exposure to the new information has no effect. Such a process has important implications for understanding how children recover from erroneous generalizations (if children’s memories can be overwritten by new learning after reactivation of erroneous information) and could also explain how new learning becomes integrated with existing knowledge if it can be shown that new information is merged with old information as part of the reconsolidation process. Almut Hupbach, Rebecca Gómez, Oliver Hardt, and Lynn Nadel have recently shown that such merging does indeed happen, suggesting that reconsolidation can be a constructive process in learning.

Summary

In sum, infants show remarkable learning abilities ranging from the ability to detect statistical patterns of varying complexity to the ability to generalize from these patterns. Learning is rapid, and appears to occur early in development. Infants show learning for a range of different types of
statistical structure including frequency, joint probability, and transitional probability. These particular forms of statistical learning occur across species, across development, and with no explicit intent to learn suggesting that they may be fundamental in learning. Infants also show learning over different types of units including phonetic segments, syllables, words, and visual objects, and they learn at different levels, for instance, at the level of specific syllables or words or at the level of generalization. Although many questions remain with respect to bridging statistical learning and learning in the world in terms of (1) understanding how learning in the lab scales up to real-life learning, and (2) delimiting the robustness of learning in terms of infants’ ability to find signal in noise, their ability to retain their learning experiences, and their ability to build on what they have learned, such learning is sure to play a central role in development.

To go full circle in terms of the history of learning, it is important to note that while the field of learning in the 1960s and 1970s focused primarily on reinforcement issues and paradigms, it has since moved to a greater appreciation of how much infants seem to learn from simple observation of regularities in their environment, both in the language and in other domains. Even so, there are still critics of learning points of view. One classic argument against learning is that simple learning mechanisms are not sufficiently powerful to explain learning of complex information found in the world (such as in language). However, in contrast to the simple learning mechanisms documented by the behaviorists, statistical learning appears to be quite powerful, with mechanisms capable of tracking vast amounts of information as well as engaging in generalization. Such sophistication raises the possibility that infant-learning may contribute substantially to the acquisition of complex skill. Yet, no matter how powerful infant-learning turns out to be, it must certainly be constrained by the biological dispositions learners bring into the world. Just how much is contributed independently by learning, how much by the child’s biological preparedness, and how much arises in the interaction of the two, has yet to be determined. Ongoing and future studies will be important for specifying the kinds of learning mechanisms children are born with and how they develop.

See also: Auditory Perception; Categorization Skills and Concepts; Cognitive Development; Imitation and Modeling; Language Acquisition Theories; Learning Disabilities; Memory; Perceptual Development; Preverbal Development and Speech Perception; Semantic Development; Speech Perception.

Suggested Readings


---

Learning Disabilities

H Liang and E Simonoff, King’s College London, London, UK

2008 Elsevier Inc. All rights reserved.

Glossary

- **Autosomal dominant** – A form of genetic (Mendelian) inheritance in which a single copy of a mutant gene will cause the disorder. Autosomal dominant disorders affect both sexes equally and are passed from parent to child, on average in 50% of cases.

- **Autosomal recessive** – A form of genetic (Mendelian) inheritance in which two copies of a mutant gene, one from each parent, are required.