Advanced Review

Statistical learning and the critical period: how a continuous learning mechanism can give rise to discontinuous learning

Erik D. Thiessen,* Sandrine Girard and Lucy C. Erickson

Infants and children are generally more successful than adults in learning novel languages, a phenomenon referred to as a *critical* or *sensitive* period for language acquisition. One explanation for this critical period is the idea that children have access to a set of language learning processes or mechanisms unavailable to adults. From this perspective, developmental change is explained in terms of a discontinuity of learning processes. We suggest that this is not the only possible explanation for developmental change in language learning outcomes. Instead, we propose that the mechanisms underlying language acquisition (in particular, we highlight statistical learning) are largely continuous across the lifespan. From this perspective, developmental change is explained in terms of experience, differences in the input with age, and maturational changes in the cognitive architecture supporting learning, even while the learning process itself operates continuously across developmental time. © 2016 Wiley Periodicals, Inc.

How to cite this article:
WIREs Cogn Sci 2016. doi: 10.1002/wcs.1394

INTRODUCTION

Perhaps no single phenomenon in the field of psycholinguistics is more readily observed, or has greater practical import, than the observation that language acquisition becomes more difficult with age. Though adults exceed infants on almost every other measure of cognitive ability, infants are more successful language learners than adults. The evidence for this claim is widespread. Adults acquiring a second language do so less easily than children.[1–3] Children who have suffered brain lesions recover linguistic function more easily than adults suffering similar lesions.[4,5] Perhaps most compellingly, even in cases of first language acquisition there appears to be an advantage for younger learners.[6–8] This phenomenon has often been described as a critical period: a maturational window in which younger learners are maximally prepared to acquire language.

Despite the strength of the evidence in favor of the existence of a critical period for language acquisition, its explanation is less clear. One popular form of explanation suggests that it occurs because younger learners have access to a different set of learning mechanisms than older learners.[9–11] From this perspective, older individuals learn language in a qualitatively different way than younger individuals. In this article, we will review evidence related to the critical period for language acquisition. Our argument is that the supposition that older learners are learning language using a qualitatively different set of processes or mechanisms than younger learners is not necessarily correct. Instead, we will explore the idea that the decline in language acquisition ability with age can be explained by theoretical accounts that posit consistency between the learning mechanisms in childhood and adulthood.

*Correspondence to: thiessen@andrew.cmu.edu

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

Conflict of interest: The authors have declared no conflicts of interest for this article.

EVIDENCE FOR DIFFERENT LEARNING MECHANISMS

The argument that older learners acquire language using a qualitatively different set of processes than younger learners was popularized by Chomsky.[12] Chomsky argued that the evidence from the input is too sparse to allow children to converge on the correct grammatical structure of their native language (cf. Ref 13). Therefore, given that children almost unfailingly converge on the correct structure, learning from the input must be supplemented by innate knowledge (termed *Universal Grammar*). Eventually, this knowledge decays, leaving adults unable to take advantage of innate knowledge of linguistic structure and forcing them to rely on alternative learning mechanisms to learn language. An alternative formulation of this type of explanation for the critical period suggests that changes in language learning are driven by the developmental onset of new learning mechanisms. For example, Nazzi and Bertoncini[14] argue that the vocabulary explosion around 18 months is best explained by a shift from learning words in an associationist manner to a referential manner. As these examples indicate, accounts of the critical period vary widely, and encompass many different aspects of the linguistic system from phonology to semantics to syntax, but a central feature of many such accounts is the claim that developmental changes in the facility for language learning are driven by qualitative changes in underlying learning processes.[5,9–11,15]

This type of theoretical account of the critical period makes a more specific claim than the observation that language acquisition is worse in adulthood. Additionally, this kind of account suggests that language acquisition in adulthood is accomplished via *different* kinds of underlying processes or mechanisms than earlier language acquisition. Two empirical phenomena have been seen as particularly compelling support for the claim that the critical period is best explained by a developmental shift in underlying learning mechanisms. The first is that the maturational state of the learner predicts the course of acquisition. For example, a large scale study of immigrants to the United States found that the relation between time spent learning English and ultimate proficiency differed for young and old learners.[3] For those who had immigrated at a young age (before about age 13), the amount of time spent learning English was linearly related to proficiency: those who immigrated earlier were better at using English than those who immigrated later. By contrast, among those who immigrated at an older age (after about age 14), age at immigration had no predictive relation to ultimate attainment. This inflection point around puberty is suggestive of a different path to proficiency—a different underlying mechanism of attainment—for young learners and older learners. However, it should be noted that while Johnson and Newport went to great lengths to control for the effects of potential third variables, it is very difficult to separate out the effects of age of immigration, length of residence, and social and linguistic backgrounds of the participants.[16] For example, some subsequent research indicates that age of learning effects disappear when education is controlled.[17]

A second piece of evidence consistent with a shift in underlying mechanisms is that young learners appear to generalize from linguistic input in a different manner than older learners, and are better at identifying structure from impoverished input. In Nicaragua, e.g., children acquiring sign language for the first time transformed a rudimentary gestural system created by the adult community into a full-fledged linguistic system, consistent with the claim that children naturally possess learning abilities capable of giving language its fundamental structure.[7] Bickerton[18] documented a similar phenomenon with Hawaiian Pidgin and Creole. The term Pidgin refers to an auxiliary language that develops from a need to have a common communication method between closely situated speakers of mutually unintelligible languages. Consequently, Pidgin languages have no native speakers and are constructed from multiple languages and cultures. They are characterized by impoverished and inconsistent syntactic patterns. A creole, on the other hand, is derived from Pidgin input and involves much more complex and consistent vocabulary, syntax, and grammar. The creolization of a Pidgin only happens when children (rather than adults) learn a Pidgin as their native language, which has been taken as evidence consistent with the idea that young children acquire language using a different set of mechanisms than do adults.

A similar set of conclusions about the special status of young learners has resulted from studies of deaf children acquiring a first language. One such example is Simon, a deaf child who learned American Sign Language (ASL) from his parents. Both of his parents learned ASL after the age of 15 and were not proficient according to native signing criteria. Consequently, Simon received highly inconsistent input about the rules of ASL compared to other deaf children with natively signing parents. Despite the inconsistent input, Simon was able to perform quite similarly to his peers on tasks of ASL morphology, and greatly surpassed his parents.[8] These results, consistent with research on the transition between Pidgin and Creole languages, suggest that children regularize and enrich simple linguistic input, effectively creating language in a way that adult learners do not. This has been taken as evidence consistent with the argument that the critical period is best explained by children learning language via different processes or mechanisms than adults use to learn language.

CRITICISMS OF THE DIFFERENTIAL MECHANISMS APPROACH

The argument that developmental changes in language acquisition can be explained by qualitative change(s) in underlying mechanism(s) is attractive both for its apparent fit to the data described above, and its straightforward nature. However, subsequent investigations of the neurobiology of language development do not lend themselves well to such a straightforward explanation. To the extent that members of a species share a biologically determined shift from one mechanism to another, this shift should be detectable in terms of a consistent change in neural organization, function, or response. By contrast, the biology underlying language acquisition is quite variable across individuals, both in terms of localization of function and in terms of its maturational timeline.[19–22] Individual variability in the timing of learning is not inconsistent with the existence of a critical period; research exploring both human and animal learning has demonstrated that a variety of experiential and pharmacological factors can shift the opening and closing of the critical period.[23,24] Recall, however, that one of the foundational arguments for a qualitative shift in learning mechanism comes from the observation of an inflection point after which age of acquisition is no longer predictive of ultimate attainment.[3]

We suggest that this degree of variability is more consistent with an account in which the critical period is best explained by a shift in the effectiveness of a continuous set of underlying mechanisms than it is with accounts in which the critical period is explained by a shift from one set of processes to another.

An additional difficulty for mechanism-change accounts of the critical period is that learning outcomes and neurological changes appear to be more strongly influenced by proficiency than by age of acquisition. Specifically, native speakers show differential activation compared to low-proficient learners but similar activation compared to both high-proficient early and late learners.[25–34] Similarly, many recent explorations of age of acquisition have shown a gradual decline, rather than the sharp inflection predicted by mechanism-change accounts. For instance, a conceptual replication of Johnson and Newport's[3] study of age of acquisition effects on second language proficiency with Spanish natives failed to show a specific age at which second language acquisition became impaired.[35] Instead, second-language performance was negatively correlated with age at the onset of learning, regardless of when learning began. Furthermore, an analysis of the 1990 U.S. Census failed to indicate discontinuities in learning. Responses from 2.3 million immigrants with Spanish or Chinese language backgrounds were evaluated and regression of second-language attainment on age of immigration failed to produce the hallmark pattern of discontinuity seen in a critical period.[16] A similar study by Bialystok and Miller[36] also failed to reproduce systematic critical period effects. In their experiment, native Spanish and Chinese speakers were divided into two groups: those that had learned English before the age of 15 and those that had learned English after the age of 15. Late Spanish learners performed significantly worse on the grammaticality judgment task compared to early Spanish learners. However, Chinese learners did not show differences in performance when comparing early and late learners. From a theoretical perspective arguing that developmental change is due to changes in underlying learning mechanisms, it is difficult to explain why one population would undergo this shift, and another would not.

AN ALTERNATIVE PERSPECTIVE ON DEVELOPMENTAL CHANGE

It is indisputable that language learning outcomes become worse the later in life that a learner begins to acquire a language. The discontinuity hypothesis explains this developmental trajectory in terms of a transition from one set of processes—often thought to be specialized in some way to support language acquisition—to a different set of processes—often described as less well suited to language acquisition.[5,9,12,15,37] In the remainder of this article, we will outline an alternative account, one that explains the impairment in language acquisition with age in terms of a single set of language learning processes that are continuously present across the lifespan. From this perspective, developmental change arises, not from a switch from one set of processes to another, but due to changes in the effectiveness of a continuously present set of learning processes. This account is consistent, in its aims, with a wide variety of recent theoretical arguments that developmental discontinuities need not be explained by underlying changes in mechanism (e.g., stage theories such as Piaget's), but instead can be explained in terms of maturational and experiential changes in a continuous set of processes.[38–40]

The case of vocabulary development provides a relevant example from language acquisition. Vocabulary acquisition starts slowly. Between 12 months and 16 months, the average child adds around 30 words to their productive vocabulary. Subsequently, this rate of acquisition triples, and the average child adds over 100 words to their productive vocabulary between 16 and 20 months. This increase in rate of acquisition (often referred to as the *vocabulary explosion*) continues well into the third year.[41] One explanation for the vocabulary explosion is that it reflects the onset of a specialized mechanism for word learning.[42–44] However, modeling work demonstrates that the vocabulary explosion can arise in systems where learning processes are continuous across developmental time. McMurray[45] demonstrated that this pattern will emerge in any learning system where words are learned in parallel, and the set of words to be learned have a Gaussian distribution of difficulty (which should characterize all natural languages). Words that are easy to learn (due to factors such as frequency, phonological form, and semantic transparency) take less time to acquire. These words are learned first, but due to the distribution of difficulty in the input, there are relatively fewer of these words than there are words of more moderate difficulty that take longer to acquire. Word learning will begin to accelerate once the learner has enough time on task to learn the greater population of words of greater difficulty. As such, learning begins slowly, but accelerates over time, without requiring any change in the nature of the underlying learning mechanism.

Connectionist models of learning provide a similar demonstration that discontinuous learning can emerge from a set of learning mechanisms that are continuously present across the course of learning.[46] Perhaps the most relevant of these demonstrations for discussion of the critical period relates to the phenomenon of *entrenchment*, in which early learning causes a stabilization of representation and expectation, making subsequent learning (that contradicts this early learning) more difficult while supporting learning and performance that is consistent with this early learning.[47] These models suggest that age of acquisition effects—in which early-learned words are more easily processed than later-learned words—in reading may be due to greater entrenchment for words that are more practiced.[48] These principles suggest that early learning and entrenchment of a first language may inhibit subsequent learning of a second language, and thus provide a partial explanation for the critical period. More generally, this class of models provides an existence proof that a continuous learning mechanism can give rise to discontinuous learning outcomes.[49,50]

A STATISTICAL LEARNING ACCOUNT OF THE CRITICAL PERIOD

We believe that a continuous learning account can explain the developmental changes in learning outcomes associated with the critical period. Specifically, we suggest that statistical learning is a likely (although perhaps not the only) candidate mechanism that can give rise to this developmental pattern. Statistical learning refers to the ability to detect and adapt to statistical regularities in the input, primarily frequency, variability, distribution, and co-occurrence probability.[115] One reason that statistical learning is a good candidate mechanism to explain developmental outcomes in language acquisition is that language is characterized by a multitude of statistical regularities, such as the distribution of phonemes in the input, the co-occurrence of words and reference, and the predictive relations between words in phrases.[51–53] Another reason is that statistical learning is present from early in life,[54] a necessary prerequisite for any learning mechanism proposed to contribute to language acquisition.

Our conception of statistical learning involves two processes: extraction and integration.[55] Extraction refers to the process of extracting units from the input (such as words from continuous utterances), and is related to sensitivity to conditional statistical information such as transitional probabilities.[56,57] Integration refers to the process of combining information across these extracted exemplars to discover the central tendency of a set of items. It is related to sensitivity to distributional statistical information such as frequency and variability.[51,58] In this conception, statistical learning is not a monolithic process; it involves the action (and interaction) of two processes. Nor does it involve the explicit computation of statistics; instead, we conceive of statistical learning as a natural outgrowth of domain-general characteristics of human memory such as interference and decay.[59] Further, as we outline in more detail below, while the interrelated processes of extraction and integration are active across the lifespan, their outcome will differ as a function of the learner's prior experience and maturational state.

Infants and adults are sensitive to the conditional relations between elements of the input: information about how one element is likely to predict subsequent elements. The most well-known metric of conditional information is transitional probability: the probability that some event Y will occur given that some other event X has already occurred. It is measured as the number of times that event XY occurs divided by the overall frequency of X. For example, if XY occurs 60 times and X occurs 100 times, the transitional probability of X to Y is 0.6. Thus, transitional probability incorporates raw frequency of co-occurrence but is often a better index of structure than mere co-occurrence because items can occur together frequently simply because they are both high frequency items (e.g., 'the dog'[56]). But in addition to transitional probability, infants and adults are sensitive to conditional relations among simultaneously presented (as opposed to sequentially presented) information.[60,61] Similarly, they are sensitive to conditional relations among elements of the input that occur non-adjacently.[62,63] Sensitivity to conditional statistical structure is likely to be relevant for language acquisition, given that many aspects of language are characterized by statistically predictable relations among elements of the input.

Distributional statistical structure refers to those aspects of the input that characterize the distribution of elements (rather than predictive relations between them), such as frequency and variability. As with conditional information, this kind of statistical structure is related to many aspects of the linguistic input. Consider, e.g., learning the phonemic categories of one's native language. Maye et al.[51] found that when infants were exposed to a distribution of exemplars along a continuum of voice onset time, ranging from /d/ to /t/, the distribution of the exemplars influenced infants' discrimination of tokens of /d/ and /t/. When infants were presented with a bimodal distribution of sounds that included prototypical exemplars of the two phonemes, infants were more likely to exhibit evidence of discriminating exemplars of the two categories. In contrast, infants who experienced a unimodal distribution that frequently included a sound intermediate between the two prototypical phoneme exemplars were less likely to show evidence of discrimination. Distributions of phonetic exemplars in the input are likely to reflect the phonemic categories of the native language.[64] When a language makes a contrast between two phonemes, exemplars between those two categories (which are ambiguous) are disfavored, such that infants are more likely to be presented with a bimodal distribution. When a language does not employ a contrast between two phonemes, the distribution of exemplars ranging between them is more likely to be unimodal. Sensitivity to distribution may play an important role in discovering many aspects of the phonetic structure of language, as well as identifying other kinds of relevant category structure, such as syntactic categories.[65]

While statistical learning may play an important role in language acquisition (for a more extensive discussion, see Ref 66), it is a domain-general mechanism that is not limited to operation over linguistic stimuli.[60,67] Moreover, we believe that it is linked to—and potentially even arises from—fundamental cognitive processes such as attention and memory.[55] Conditional statistical learning is critically dependent on attention: learners only discover that two elements predict each other when they are simultaneously held in attention and can be bound together into a discrete representation.[68,69] Similarly, distributional statistical learning is fundamentally about discovering the central tendency of a set of exemplars, suggesting a commonality with more general processes of memory such as prototype formation. Indeed, it is possible to model the results of many distributional statistical learning tasks using exemplar memory models invoking processes such as activation, decay, and integration of information into a prototype.[58] As such, a focus on the role of statistical learning in language acquisition suggests that developmental changes in language learning outcomes may be linked with more general cognitive changes that occur over the course of development, changes that alter the outcome of a continuously present statistical learning process.

Given the characteristics of the statistical learning processes described above, there are two major factors that may plausibly be linked to changes in language learning outcomes associated with increasing age. The first is increasing familiarity of language, which makes a learner better adapted to the languages with which they are familiar, but less able to adapt to novel languages. The second factor is maturational changes, which alter both the cognitive architecture supporting statistical learning, and the degree of plasticity with which the learner's neurobiological organization can adapt to novel input. Together, we suggest that these two factors can explain the degradation of language learning abilities with age in a way that does not necessitate the invocation of a qualitative change from one learning mechanism to another learning mechanism. In fact, this account may be a better fit to empirical observations of 'critical period' phenomena, which are somewhat more variable and inconsistent across learners than 'mechanism-change' critical period explanation predicts.

CHANGE IN LEARNING OUTCOMES DUE TO EXPERIENCE

From this perspective, the advantage of younger language learners relates, in part, to the fact that infants and young children are highly adaptable learners, but not yet strongly adapted to their particular linguistic environment. That is, infants are flexible learners able to acquire any of the world's languages. As an example, they are born with the ability to perceive a wide range of phonemic contrasts.[70] However, as infants acquire experience with their native language, their representations, learning biases, and expectations become adapted to the language(s) with which they are familiar.[71] This adaptation is a double-edged sword. The loss of sensitivity to contrasts not used in the native language is associated with an increase in sensitivity to the phonemic contrasts of the native language,[72] which is likely to facilitate subsequent learning in the native language.[73] However, while this adaptation facilitates comprehension and production of the native language, it impairs learning in novel linguistic settings characterized by a different set of phonemic contrasts.

This pattern of early, flexible learning being shaped by experience toward more efficient—but specialized—learning can be seen not just in phonemic perception, but across a wide range of linguistic structures, including phonotactics, syntax, and phonology. To provide another example, consider the lexical stress of English, where most content words are stressed on their first syllable.[74] Adult English speakers use this as a cue to word segmentation, and treat stressed syllables as a cue to word onset.[75] In contrast, English-learning infants younger than 7 months do not treat lexical stress as a cue to segmentation.[76] Only when infants have become familiar with enough words to discover the statistical relation between lexical stress and word onsets do they begin to use lexical stress as a cue to word segmentation.[77] Discovering this relation not only makes subsequent word segmentation easier, but also slows learning in novel linguistic environments where lexical stress is uninformative or occurs on a different word position.[78] As these examples indicate, early learning informs subsequent learning in ways that allow language learners to specialize to their native language, but can hinder discovery of regularities in novel linguistic contexts. One prediction that falls out of this perspective is that the degree of degradation of language learning outcomes with age ought to be related, at least in part, to the similarity between the language(s) learned early in development and the language(s) learned later in life.[79] Note that this prediction is not as simple as suggesting that similar languages should be learned more easily; in some cases, similarity can be harmful.[80] Exploring this prediction may help to explain some of the variability observed in previous reports of language learning in adulthood; while many adults learn new languages poorly, some adults (as many as 20%, in some samples) attain native-like levels of competency.[35]

Finally, we suggest that experience alters the manner in which infants encode information. In particular, we believe that infants encode information in a manner that is both noisier than adults—such that representations of identical or near-identical events are more likely to differ[81]—and less well-tuned to the characteristics of the linguistic input—such that infants are more likely to encode irrelevant information such as indexical characteristics of the speaker instead of, or in addition to, linguistically relevant features such as phonemic identity.[82–84] This has several implications that differentiate younger learners from older learners. The first of these relates to generalization: when the features of the current input match a high percentage of the features of the old information stored in memory, the old information is activated and guides processing of the new input. Encoding a greater number of idiosyncratic features reduces the likelihood that two exemplars related to the same central concept will activate each other, which decreases a learner's ability to identify the features that are common across them and generalize that commonality to novel exemplars and contexts.[58] This predicts slower learning early in language acquisition until infants discover the features that are relevant to their native language.[71]

However, noisy representations containing a greater weight (compared to adult representations) on idiosyncratic features may also help infants with the process of discovering features that are relevant to their native language. Across any set of N exemplars characterized by both common features and idiosyncratic features, common features are more likely to 'survive' being encoded in the presence of noise.[58,85] This is because idiosyncratic features are present across fewer members of the set, so are more likely to be erased or altered by noisy, inaccurate encoding. As such, immature encoding is actually likely to accentuate the commonality across a set of exemplars, leading to greater likelihood of detecting the commonality across them. By contrast, encodings that are weighted toward a particular set of features—as is the case for adults who have discovered the regularities in their own language, such as its phonemic inventory—are less likely to discover novel commonalities across a set of exemplars that contravene those weightings. Thus, while infant learning is slower, it is better adapted to discover structure in novel linguistic environments. Conversely, once infants have identified a set of relevant representations from their input, they are less likely to generalize over input that is common across aspects of the input that are less relevant in their prior experience.[86,87]

The second implication of a noisier encoding system relates to the response to probabilistic structure in the input. Because adults are encoding the input more precisely, they can represent the probability structure of the input more accurately. By contrast, because infants are encoding the input less precisely, they should be biased toward both representing and producing high probability occurrences over low probability occurrences (because high probability occurrences happen more often, and so are more likely to be preserved or accessed in the face of a noisy encoding system). This facet of memory development may well be responsible for features of child language learning such as overgeneralization.[88,89] Artificial grammar learning experiments are consistent with this perspective. Presented with input in which an element X is preceded by A 67% of the time and by B 33% of the time, adults match that probability when they are asked to produce or verify examples from the language. By contrast, infants and young children appear to learn the more probable mapping (AX), but not the less probable mapping (BX).[90] This tendency toward overregularization may explain why infants and young children, but not adults, develop a productive language system when they are exposed to inconsistent input (as in Pidgin to Creole transitions, or learning a native language from foreign speakers): they represent the central tendency of the input, and effectively ignore the noise.[8] By contrast, adult learners mimic the input more faithfully, which can be a disadvantage when learning from an inconsistent model.

CHANGES IN LEARNING OUTCOMES DUE TO MATURATION

Prior experience alone, however, cannot explain impaired language learning in adulthood. If so, late L1 learners (such as adults exposed to language for the first time) should learn language as well as children. This is clearly not the case, as can be seen both from anecdotal examples of children raised in isolation, and from empirical research exploring outcomes of cochlear implantation as a function of age.[91] In addition to the effects of prior experience, we suggest that maturational changes also alter language learning outcomes with age. By maturation, we refer to biological changes that are expected to occur with age given typical experience, while acknowledging that a certain degree of experience is often required for, or inextricable from, maturational changes. Our statistical learning framework highlights two kinds of maturational changes that are likely to have explanatory power for changes in learning outcomes. The first is age-related changes in the cognitive architecture supporting statistical learning, particularly memory and attention. In this regard, our approach is similar to Newport's[92] Less is More hypothesis, which suggested that younger learners may benefit from an immature memory system. Note, however, that our approach highlights different development changes to the memory system, in part due to recent work suggesting that Less is More is in some ways incomplete or inaccurate.[93,94] The second maturational change that ought to matter, from this perspective, is the degree of plasticity (or entrenchment) of neural organization across age.

Our framework suggests that age-related changes to memory and attention are particularly likely to influence statistical learning outcomes. For example, conditional statistical learning is dependent on attention, such that only stimuli that are simultaneously held in attention can be associated into a single unit.[68,95] This explains why organisms exposed to the same input can learn different associations; only those associations that are attended to can be learned.[96] It may similarly explain why different statistical learning tasks share little variance (e.g., Ref 97; cf. Ref 66); to the extent that these tasks focus attention on different aspects of the input, they may tap into different kinds of representations and give rise to different outcomes.[98] Relevant to the critical period, attention changes dramatically as a function of age, as adults are much better than children at effortfully controlling their focus of attention.[99] In learning a novel language, however, this ability to control attention may be disadvantageous.[100] This is particularly clear in situations where the structure of the input does not match adults' expectations. For example, when adults are focused on parsing words, they struggle to discover phrase level regularities such as gender agreement.[93] More generally, the ability to control attention and search for evidence consistent with pre-existing or rapidly emerging hypotheses may make adults less sensitive to the statistical structure of the input.[101]

In addition to providing the ability to focus on hypothesis-consistent aspects of the input, mature attentional systems may impair language learning via the ability to inhibit (seemingly) irrelevant information. The existence of competitors can actually promote learning. Learning is often modeled as a process in which the ability to discriminate and predict the environment is successively refined via error and competition. The ability to inhibit competitors (such as potential alternative syntactic and morphological constructions) may impede learning.[102] As such, one reason why child learning may be more successful is that a relatively underdeveloped prefrontal cortex may make inhibition of potential alternatives more difficult, with the resulting competition strengthening learning as the dominant pattern must be continually reactivated in the face of competitors. This discussion suggests that distraction may actually promote adult learning. While this prediction has not been studied extensively, at least some results suggest that adults learning language or language-like systems learn more slowly under distraction, but eventually generalize more successfully.[103]

Finally, we suggest that maturation influences the extent to which experience can alter the neural circuitry associated with language. Learning depends on updating the connections between neurons. Owing to factors such as the overpopulation of synapses in infancy, infants may be especially flexible in their updating of these connections, a state referred to as *plasticity*. Studies of congenitally deaf children who receive cochlear implants, e.g., suggest that if the neural systems fail to receive input for several years, this period of plasticity ends by adolescence.[104] Similarly, in typically developing children, once the auditory cortex has adapted to the input of a native language, this organization may become entrenched and render learning of new languages more difficult.[105] This is consistent with evidence that infants are better able to recover from damage to left hemisphere regions associated with language comprehension and processing than are adults.[106]

Undoubtedly, some of these changes in plasticity are driven by experience. The degree to which an infant has discovered one regularity may cause the neural circuitry supporting this regularity to become entrenched and resistant to change.[107] However, recent work indicates that it is also driven, in part, by biological state.[108] Updating neuronal connections depends on a complex interplay of synaptic state and neurotransmitter balance.[114] Age-related changes in the balance of factors that trigger plasticity, and that provide a brake on plasticity, change the extent to which the neural architecture supporting language processing and comprehension can be altered.[109] Regardless of the precise balance of experiential and maturational factors in causing age-related changes in plasticity, the end effect is that with age, the neural architecture supporting language becomes more resistant to change. We suggest that this means that even if the same underlying set of learning mechanisms are operating across age, those learning mechanisms will be less effective in allowing infants to adapt to novel linguistic input.

SUMMARY AND FUTURE DIRECTIONS

Explanations of the critical period based on a change in underlying learning process are appealing due to their straightforward nature. However, there are serious problems with these kinds of accounts. One set relates to their descriptive adequacy. Research on the developmental trajectory of language acquisition has revealed inconsistency in the timing of a purported critical period,[16,17,35] which aspects of language acquisition are impaired,[36] and the degree to which native-like adult language acquisition is possible.[27] Another set of problems relates to their theoretical content. The underlying learning mechanisms of discontinuous learning accounts are, in many cases, not well tied to the biology of language acquisition.[27] In other cases, the learning mechanisms themselves are poorly defined, so that these arguments are marked by circular logic: early learners learn language better because they have access to a language learning mechanism; we know they have access to a language learning mechanism because they learn language better!

We argue that it is instead logically possible to explain critical periods in terms of the same set of continuously present learning mechanisms operating with different degrees of efficiency across the lifespan due to experience and maturation. Indeed, cross-sectional investigations of statistical learning have begun to demonstrate that young learners detect and generalize statistical regularities differently than older learners.[90,110] A continuous learning mechanism approach may, in fact, provide a better account of the critical period than mechanism-change accounts for two main reasons. First, because it emphasizes learning, it has the potential to provide a better fit to the inconsistency and individual variability we have highlighted in this review. Early learning has a cascading effect on subsequent learning. On this perspective, our account is consistent with the emergentist perspective outlined by Golinkoff and others,[38] suggesting that learning shifts as the infant learns to make use of different aspects of the environment. As a function of what the child learns early, and the fit between that learning and the linguistic environments in which the learner finds themselves (e.g., the similarity between L1 and L2; for discussion, see Ref 79), outcomes for later learning can be quite different. Second, because this type of account is focused on the mechanisms underlying learning, and the way those mechanisms change over the course of experience and maturation, it yields a set of falsifiable predictions about the causes of individual differences in language learning outcomes, as well as the kinds of manipulations (to both the learner, and the input) that might facilitate language learning.

Many of these testable predictions relate to the specific set of statistical learning processes we have invoked in this account. For example, although it is clear that both infants and adults are capable of statistical learning, we have suggested that the learning outcomes of these processes change over the course of development both as a function of experience, and as a function of the maturational state of the learner. Some evidence already exists that the outcome of statistical learning changes as the learner acquires experience,[77,111] but much of this evidence comes from artificial grammar learning. As such, it serves only as an existence proof that this alteration is possible, not that these kinds of changes actually occur as infants are exposed to real languages (though see Ref 112). Similarly, there is good evidence that statistical learning is tied to more general cognitive processes such as memory and attention, yet the way in which maturational changes in these more general processes might alter statistical learning is still largely uninvestigated (although see Ref 113 for one counterexample). Investigating these predictions will be necessary for both an assessment of the validity of our continuous-learning account of developmental change, and for building a fully specified account of this developmental change. Such an account remains a promissory note at this point, but we believe that a focus on developmental change in the context of a continuous set of underlying processes is both a plausible and a fruitful approach to thinking about developmental changes in language acquisition.

REFERENCES

1. Best CT. The emergence of native-language phonological influences in infants: a perceptual assimilation model. In: *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge, MA: MIT Press; 1994, 167–224.
2. Flege JE. Second language speech learning: theory, findings, and problems. In: *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore, MD: York Press; 1995, 233–277.
3. Johnson JS, Newport EL. Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn Psychol* 1989, 21:60–99.
4. Baser LS. Hemiplegia of early onset and the faculty of speech with special reference to the effects of hemispherectomy. *Brain* 1962, 85:427–460.
5. Lenneberg E. *Biological Foundations of Language*. New York: John Wiley & Sons; 1967.
6. Senghas A, Coppola M. Children creating language: how Nicaraguan sign language acquired a spatial grammar. *Psychol Sci* 2001, 12:323–328.
7. Senghas A, Kita S, Özyürek A. Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science* 2004, 305:1779–1782.
8. Singleton JL, Newport EL. When learners surpass their models: the acquisition of American Sign Language from inconsistent input. *Cogn Psychol* 2004, 49:370–407.
9. Bley-Vroman RW, Felix SW. The accessibility of Universal Grammar in adult language learning. *Second Lang Res* 1988, 4:1–32.
10. Komarova NL, Nowak MA. Natural selection of the critical period for language acquisition. *Proc R Soc Lond B Biol Sci* 2001, 268:1189–1196.
11. Singleton D. The critical period hypothesis: A coat of many colours. *IRAL Int Rev Appl Linguist Lang Teach* 2005, 43:269–285.
12. Chomsky N. *Aspects of the Theory of Syntax*, vol. 11. Cambridge, MA: MIT Press; 1965.
13. Gold EM. Language identification in the limit. *Inform Control* 1967, 10:447–474.
14. Nazzi T, Bertoncini J. Before and after the vocabulary spurt: two modes of acquisition? *Dev Sci* 2003, 6:136–142.
15. Seliger HW. Implications of a multiple critical periods hypothesis for second language learning. In:

Richie WC, ed. *Second Language Acquisition Research: Issues and Implications*. New York: Academic Press; 1978, 11–19.

16. Hakuta K, Bialystok E, Wiley E. Critical evidence: a test of the critical-period hypothesis for second-language acquisition. *Psychol Sci* 2003, 14:31–38.
17. Flege JE, Yeni-Komshian GH, Liu S. Age constraints on second-language acquisition. *J Mem Lang* 1999, 41:78–104.
18. Bickerton D. The language bioprogram hypothesis. *Behav Brain Sci* 1984, 7:173–188.
19. Knecht S, Dräger B, Deppe M, Bobe L, Lohmann H, Flöel A, Henningsen H. Handedness and hemispheric language dominance in healthy humans. *Brain* 2000, 123:2512–2518.
20. Newman AJ, Bavelier D, Corina D, Jezzard P, Neville HJ. A critical period for right hemisphere recruitment in American Sign Language processing. *Nat Neurosci* 2002, 5:76–80.
21. Ojemann GA, Whitaker HA. Language localization and variability. *Brain Lang* 1978, 6:239–260.
22. Sommer IEC, Ramsey NF, Mandl RCW, Kahn RS. Language lateralization in monozygotic twin pairs concordant and discordant for handedness. *Brain* 2002, 125:2710–2718.
23. Livingston FS, Mooney R. Androgens and isolation from adult tutors differentially affect the development of songbird neurons critical to cortical vocal plasticity. *J Neurophysiol* 2001, 85:34–42.
24. Letzkus JJ, Wolff SB, Meyer EM, Tovote P, Courtin J, Herry C, Luthi A. A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature* 2011, 480:331–335.
25. Chee MW, Hon N, Lee HL, Soon CS. Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *Neuroimage* 2001, 13:1155–1163.
26. Dehaene S, Dupoux E, Mehler J, Cohen L, Paulesu E, Perani D, van de Moortele PF, Lehéricy S, Le Bihan D. Anatomical variability in the cortical representation of first and second language. *Neuroreport* 1997, 8:3809–3815.
27. Friederici AD, Steinhauer K, Pfeifer E. Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proc Natl Acad Sci USA* 2002, 99:529–534.
28. Hahne A. What's different in second-language processing? Evidence from event-related brain potentials. *J Psycholinguist Res* 2001, 30:251–266.
29. Hahne A, Friederici AD. Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingual Lang Cogn* 2001, 4:123–141.
30. Ojima S, Nakata H, Kakigi R. An ERP study of second language learning after childhood: effects of proficiency. *J Cogn Neurosci* 2005, 17:1212–1228.
31. Perani D, Paulesu E, Galles NS, Dupoux E, Dehaene S, Bettinardi V, Mehler J. The bilingual brain. *Brain* 1998, 121:1841–1852.
32. Proverbio AM, Čok B, Zani A. Electrophysiological measures of language processing in bilinguals. *J Cogn Neurosci* 2002, 14:994–1017.
33. Stowe LA, Sabourin L. Imaging the processing of a second language: effects of maturation and proficiency on the neural processes involved. *Int Rev Appl Linguist Lang Teach* 2005, 43:329–353.
34. Wartenburger I, Heekeren HR, Abutalebi J, Cappa SF, Villringer A, Perani D. Early setting of grammatical processing in the bilingual brain. *Neuron* 2003, 37:159–170.
35. Birdsong D, Molis M. On the evidence for maturational constraints in second-language acquisition. *J Mem Lang* 2001, 44:235–249.
36. Bialystok E, Miller B. The problem of age in second-language acquisition: influences from language, structure, and task. *Bilingual Lang Cogn* 1999, 2: 127–145.
37. Pearl L, Lidz J. When domain general learning fails and when it succeeds: identifying the contribution of domain specificity. *Lang Learn Dev* 2009, 5:235–265.
38. Golinkoff RM, Hirsh-Pasek K. Baby wordsmith from associationist to social sophisticate. *Curr Dir Psychol Sci* 2006, 15:30–33.
39. MacWhinney B. The competition model: the input, the context, and the brain. In: *Cognition and Second Language Instruction*. Cambridge, UK: Cambridge University Press; 2001, 69–90.
40. van Geert P, Steenbeek H. The dynamics of scaffolding. *New Ideas Psychol* 2005, 23:115–128.
41. Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ, Stiles J. Variability in early communicative development. *Monogr Soc Res Child Dev* 1994, 59:1–185.
42. Bloom P. Précis of How children learn the meanings of words. *Behav Brain Sci* 2001, 24:1095–1103.
43. Waxman SR, Booth AE. Principles that are invoked in the acquisition of words, but not facts. *Cognition* 2000, 77:B33–43.
44. Waxman SR, Booth AE. On the insufficiency of evidence for a domain-general account of word learning. *Cognition* 2001, 78:277–279.
45. McMurray B. Defusing the childhood vocabulary explosion. *Science* 2007, 317:631–631.
46. Plunkett K, Marchman VA. Learning from a connectionist model of the acquisition of the English past tense. *Cognition* 1996, 61:299–308.

47. Zevin JD, Seidenberg MS. Age of acquisition effects in reading and other tasks. *J Mem Lang* 2002, 47:1–29.
48. Zevin JD, Seidenberg MS. Age-of-acquisition effects in reading aloud: tests of cumulative frequency and frequency trajectory. *Mem Cognit* 2004, 32:31–38.
49. Seidenberg MS. Language acquisition and use: learning and applying probabilistic constraints. *Science* 1997, 275:1599–1603.
50. Seidenberg MS, MacDonald MC. A probabilistic constraints approach to language acquisition and processing. *Cognit Sci* 1999, 23:569–588.
51. Maye J, Werker JF, Gerken L. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 2002, 82:B101–111.
52. Vouloumanos A. Fine-grained sensitivity to statistical information in adult word learning. *Cognition* 2008, 107:729–742.
53. Chemla E, Mintz TH, Bernal S, Christophe A. Categorizing words using 'frequent frames': what cross-linguistic analyses reveal about distributional acquisition strategies. *Dev Sci* 2009, 12:396–406.
54. Teinonen T, Fellman V, Näätänen R, Alku P, Huotilainen M. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci* 2009, 10:21.
55. Thiessen ED, Kronstein AT, Hufnagle DG. The extraction and integration framework: a two-process account of statistical learning. *Psychol Bull* 2013, 139:792–814.
56. Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. *Psychol Sci* 1998, 9:321–324.
57. Perruchet P, Vinter A. PARSER: a model for word segmentation. *J Mem Lang* 1998, 39:246263.
58. Thiessen ED, Pavlik PI. iMinerva: a mathematical model of distributional statistical learning. *Cognit Sci* 2012, 37:310–343.
59. Thiessen ED, Erickson LC. Discovering words in fluent speech: the contribution of two kinds of statistical information. *Front Psychol* 2012, 3. doi:10.3389/fpsyg.2012.00590.
60. Fiser J, Aslin RN. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci* 2001, 12:499–504.
61. Fiser J, Aslin RN. Statistical learning of new visual feature combinations by infants. *Proc Natl Acad Sci USA* 2002, 99:15822–15826.
62. Gebhart AL, Newport EL, Aslin RN. Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon Bull Rev* 2009, 16:486–490.
63. Gómez RL. Variability and detection of invariant structure. *Psychol Sci* 2002, 13:431–436.
64. Werker JF, Pons F, Dietrich C, Kajikawa S, Fais L, Amano S. Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition* 2007, 103:147–162.
65. Mintz TH. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 2003, 90:91–117.
66. Erickson LC, Thiessen ED. Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Dev Rev* 2015, 37:66–108.
67. Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 2002, 83: B35–42.
68. Baker CI, Olson CR, Behrmann M. Role of attention and perceptual grouping in visual statistical learning. *Psychol Sci* 2004, 15:460–466.
69. Toro JM, Sinnett S, Soto-Faraco S. Speech segmentation by statistical learning depends on attention. *Cognition* 2005, 97:B25–34.
70. Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Näätänen R. Development of language-specific phoneme representations in the infant brain. *Nat Neurosci* 1998, 1:351–353.
71. Werker JF, Tees RC. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav Dev* 1984, 7:49–63.
72. Kuhl PK, Stevens E, Hayashi A, Deguchi T, Kiritani S, Iverson P. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev Sci* 2006, 9:F13–21.
73. Tsao FM, Liu HM, Kuhl PK. Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Dev* 2004, 75:1067–1084.
74. Cutler A, Carter DM. The predominance of strong initial syllables in the English vocabulary. *Comput Speech Lang* 1987, 2:133–142.
75. Cutler A, Norris D. The role of strong syllables in segmentation for lexical access. *J Exp Psychol Hum Percept Perform* 1988, 14:113.
76. Thiessen ED, Saffran JR. When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Dev Psychol* 2003, 39:706.
77. Thiessen ED, Saffran JR. Learning to learn: infants' acquisition of stress-based strategies for word segmentation. *Lang Learn Dev* 2007, 3:73–100.
78. Thiessen ED, Saffran JR. Spectral tilt as a cue to word segmentation in infancy and adulthood. *Percept Psychophys* 2004, 66:779–791.
79. Kellerman E. Crosslinguistic influence: transfer to nowhere? *Annu Rev Appl Linguist* 1995, 15:125–150.
80. Best CT, McRoberts GW, Sithole NM. Examination of perceptual reorganization for nonnative speech

contrasts: Zulu click discrimination by English-speaking adults and infants. *J Exp Psychol Hum Percept Perform* 1988, 14:345–360.

81. Thiessen ED, Pavlik PI. Modeling the role of distributional information in children's use of phonemic contrasts. *J Mem Lang* 2016, 88:117–132.
82. Houston DM, Jusczyk PW. The role of talker-specific information in word segmentation by infants. *J Exp Psychol Hum Percept Perform* 2000, 26:1570.
83. Newman RS. The level of detail in infants' word learning. *Curr Dir Psychol Sci* 2008, 17:229–232.
84. Singh L. Influences of high and low variability on infant word recognition. *Cognition* 2008, 106: 833–870.
85. Hintzman DL. MINERVA 2: a simulation model of human memory. *Behav Res Methods Instrum Comput* 1984, 16:96–101.
86. Dawson C, Gerken L. From domain-generality to domain-sensitivity: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition* 2009, 111:378–382.
87. Gerken L, Bollt A. Three exemplars allow at least some linguistic generalizations: implications for generalization mechanisms and constraints. *Lang Learn Dev* 2008, 4:228–248.
88. Kuczaj SA. Children's judgments of grammatical and ungrammatical irregular past-tense verbs. *Child Dev* 1978, 49:319–326.
89. Regier T. Emergent constraints on word-learning: a computational perspective. *Trends Cogn Sci* 2003, 7:263–268.
90. Hudson Kam CL, Newport EL. Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Lang Learn Dev* 2005, 1:151–195.
91. Robinson K. Implications of developmental plasticity for the language acquisition of deaf children with cochlear implants. *Int J Pediatr Otorhinolaryngol* 1998, 46:71–80.
92. Newport EL. Constraints on learning and their role in language acquisition: studies of the acquisition of American Sign Language. *Lang Sci* 1988, 10: 147–172.
93. Arnon I, Ramscar M. Granularity and the acquisition of grammatical gender: how order-of-acquisition affects what gets learned. *Cognition* 2012, 122: 292–305.
94. Rohde DLT, Plaut DC. Language acquisition in the absence of explicit negative evidence: wow important is starting small?. *Cognition* 1999, 72:67–109.
95. Thiessen ED, Hill EA, Saffran JR. Infant directed speech facilitates word segmentation. *Infancy* 2005, 7:53–71.
96. Garcia J, Koelling RA. Relation of cue to consequence in avoidance learning. *Psychon Sci* 1966, 4:123–124.
97. Siegelman N, Frost R. Statistical learning as an individual ability. *J Mem Lang* 2015, 81:105–120.
98. Zhao J, Ngo N, McKendrick R, Turk-Browne NB. Mutual interference between statistical summary perception and statistical learning. *Psychol Sci* 2011, 22:1212–1219.
99. Garon N, Bryson SE, Smith IM. Executive function in preschoolers: a review using an integrative framework. *Psychol Bull* 2008, 134:31.
100. Finn AS, Lee T, Kraus A, Kam CLH. When it hurts (and helps) to try: the role of effort in language learning. *PLoS One* 2014, 9:e101806.
101. Fletcher PC, Zafiris O, Frith CD, Honey RAE, Corlett PR, Zilles K, Fink GR. On the benefits of not trying: brain activity and connectivity reflecting the interactions of explicit and implicit sequence learning. *Cereb Cortex* 2005, 15:1002–1015.
102. Thompson-Schill SL, Ramscar M, Chrysikou EG. Cognition without control when a little frontal lobe goes a long way. *Curr Dir Psychol Sci* 2009, 18:259–263.
103. Cochran BP, McDonald JL, Parault SJ. Too smart for their own good: the disadvantage of a superior processing capacity for adult language learners. *J Mem Lang* 1999, 41:30–58.
104. Ponton CW, Eggermon JJ. Of kittens and kids: altered cortical maturation following profound deafness and cochlear implant use. *Audiol Neurotol* 2001, 6:363–380.
105. Bedny M, Pascual-Leone A, Dravida S, Saxe R. A sensitive period for language in the visual cortex: distinct patterns of plasticity in congenitally versus late blind adults. *Brain Lang* 2011, 122:162–170. doi:10.1016/j.bandl.2011.10.005.
106. Bates E. Language and the infant brain. *J Commun Disord* 1999, 32:195–205.
107. Seidenberg MS, Zevin JD. Connectionist models in developmental cognitive neuroscience: critical periods and the paradox of success. In: *Attention & Performance XXI: Processes of Change in Brain and Cognitive Development*. Oxford, England: Oxford University Press; 2006, 585–612.
108. Takesian AE, Hensch TK. Balancing plasticity/stability across brain development. *Prog Brain Res* 2013, 207:3–34.
109. Werker JF, Hensch TK. Critical periods in speech perception: new directions. *Psychology* 2015, 66:173.
110. McNealy K, Mazziotta JC, Depretto M. Age and experience shape developmental changes in the neural basis of language-related learning. *Dev Sci* 2011, 14:1261–1282.
111. Lew-Williams C, Pelucchi B, Saffran JR. Isolated words enhance statistical language learning in infancy. *Dev Sci* 2011, 14:1323–1329.

112. Onnis L, Thiessen E. Language experience changes subsequent learning. *Cognition* 2013, 126:268–284.

113. Cherry KE, Stadler ME. Implicit learning of a nonverbal sequence in younger and older adults. *Psychol Aging* 1995, 10:379.

114. Hensch TK. Critical period plasticity in local cortical circuits. *Nat Rev Neurosci* 2005, 6:877–888.

115. Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science* 1996, 274:1926–1928.