Review

CrossMark
click for updates

Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective

Rebecca L. Gómez

Department of Psychology, University of Arizona, Tucson, AZ 85721-0068, USA

RLG, 0000-0002-8347-314X

Statistical structure abounds in language. Human infants show a striking capacity for using statistical learning (SL) to extract regularities in their linguistic environments, a process thought to bootstrap their knowledge of language. Critically, studies of SL test infants in the minutes immediately following familiarization, but long-term retention unfolds over hours and days, with almost no work investigating retention of SL. This creates a critical gap in the literature given that we know little about how single or multiple SL experiences translate into permanent knowledge. Furthermore, different memory systems with vastly different encoding and retention profiles emerge at different points in development, with the underlying memory system dictating the fidelity of the memory trace hours later. I describe the scant literature on retention of SL, the learning and retention properties of memory systems as they apply to SL, and the development of these memory systems. I propose that different memory systems support retention of SL in infant and adult learners, suggesting an explanation for the slow pace of natural language acquisition in infancy. I discuss the implications of developing memory systems for SL and suggest that we exercise caution in extrapolating from adult to infant properties of SL.

This article is part of the themed issue 'New frontiers for statistical learning in the cognitive sciences'.

1. Introduction

Statistical learning (SL) is the study of how learners detect and extract reliable patterns. Learners accomplish this feat by tracking statistical regularities in language, the visual world and everyday events. The occurrence of SL in different modalities suggests that we do not need special inborn mechanisms for learning language, *per se*. Extraction of statistical patterns is also rapid, emerging within minutes of exposure in infant behaviour [1,2].

Rapid learning was a contentious topic in 1996 when there were two talks on SL at the Boston University Conference on Language Development (BUCLD), the preeminent annual meeting on language acquisition. The data presented at that time were published soon after by Saffran *et al.* [1] and Gómez & Gerken [2]. Although the two laboratories investigated different aspects of SL, sequential probabilities as a source of information for extracting words from running speech in the first case, and extraction and generalization of frequent sequential patterns in an artificial grammar or artificial grammar learning, in the second, the findings suggested that infants could rapidly bootstrap knowledge of language with the aid of SL. These findings countered the prevailing thought of the time that infant learning was too rudimentary to explain how children learn language in the span of a few short years, that instead language must develop according to a special language programme. Thus, despite continued debate about the nature of learning in language acquisition [3–5], there is now widespread appreciation for the fact that children rely

on a variety of mechanisms to acquire language, not all innately determined or specific to language. There are also now many such presentations at the BUCLD and many important publications of SL in children and adults.

Yet, the rapid SL observed in laboratory studies stands in stark contrast to the slow unfolding of real-world language acquisition where statistical knowledge emerges slowly in children's language. Robust perception of statistically reliable phonological patterns in natural language does not emerge until eight to 10 months of age, after children have had many months of exposure to language [6–8]. For instance, the perceptual reorganization documented by Werker & Tees [6] is thought to reflect the influence of statistical knowledge on phonetic categories in natural language [9]. Such categories are thought to constrain infants' ability to detect non-native contrasts by eight to 10 months despite the fact that infants show rapid SL of phonetic categories as early as six months in a laboratory study [9].

How do we resolve the seeming inconsistency of rapid extraction of statistics in the laboratory with the slow pace of real-world language learning? For one, laboratory stimuli are often highly focused on the pattern to be learned with high density input, whereas real-life input may occur in spaced intervals with exceptions to the most reliable patterns depending on the language. For another, learners must retain SL over time. I focus on this latter issue, the retention of SL as the major topic of this paper. To address the puzzle of rapid extraction of statistics and slow acquisition of language, I first review the small literature on retention of SL in adults and infants. Although retention of SL over a 24 h period is robust in adults, it is fragile in infants. I go on to discuss the kinds of memory systems that may underlie extraction of SL and its retention. I then discuss the development of memory systems and propose that we can reconcile the contrast between fast extraction of statistics in the laboratory and slow learning in natural language with developmental changes in memory systems. I end by discussing the implications of memory-systems development for SL in language acquisition.

2. Retention of statistical learning

Despite hundreds of studies of SL, research investigating *retention* of SL is rare. Thus, we know little about how the learning we measure in a single laboratory experiment translates into permanent experience in infants or adults. This is a critical gap in knowledge given that memories go through successive stages of generation of a trace during encoding of new information, stabilization in the minutes after encoding, consolidation over hours, and maintenance over days and weeks [10]. Critically, progression from one phase to the next depends on the success of the previous stage, which may differ in younger and older learners. Although it is tempting to address questions of retention based on prior studies of learning, e.g. Ebbinghaus' studies of memory for non-words [11], it is doubtful that Ebbinhaus' methods recruit the same processing as SL. Ebbinghaus memorized the lists of nonsense words to a criterion level, a requirement rarely imposed in SL, he learned lists of items intentionally, whereas SL occurs implicitly and incidentally [12–15], and he purposefully included no statistical structure in his lists, whereas statistical structure drives SL. These are all factors that may differentially affect different stages of memory formation.

We do know that adults exhibit equal retention of visually presented shape triples immediately after SL and 24 h later as measured on implicit [12] and explicit tests [16]. In a separate study, discrimination of statistically predictable versus unpredictable tone sequences improved after a 24 h delay [17]. As we will see, this is a different pattern than that observed for studies of retention of SL in infants. In contrast to adults, infants show weak retention (reflected in savings in new learning) or loss of the fidelity of a memory.

Before addressing retention of SL in infants, it is important to ask whether we can draw from memory work outside of SL. Despite an extensive literature on infant memory using the conjugate mobile paradigm [18], it is unlikely that this procedure recruits the same memory systems infants use for SL. As an example, although six month olds retain high-fidelity memories for two weeks after two 6 min exposures to a visual stimulus over 2 days [19], the conjugate mobile paradigm used in such studies entails reinforcement at encoding in the form of a moving mobile (the infant's foot connects to a mobile that moves when the infant kicks). This sets up a feedback loop that rewards the infant for kicking, culminating in a robust and long-lived association between the visual stimulus (the mobile) and the kick response. This experimental procedure likely recruits a network with very different characteristics than the ones involved in SL where infants learn associations between stimuli (not between a stimulus and response) without overt reinforcement. Brain imaging in adults shows that category learning with and without reinforcement recruits distinct corticostriatal loops [20,21]. Nor do other memory tasks used with infants apply to SL. Although deferred imitation requires infants to retain sequences of events, such learning is thought to rely on explicit as opposed to implicit memory [22]. The fact that infants do not reliably learn transitions between arbitrarily related events in deferred imitation until 22 months of age [22], long after they show evidence of SL, adds further concern that deferred imitation recruits a different memory system than SL. To address the question of retention of SL in infants, I turn to work on retention of the kind of information that is likely to involve SL in natural language and the handful of studies of long-term retention of SL.

Evidence suggests that perceptual prerequisites of SL begin prenatally. Foetuses can detect low-frequency components of maternal speech, primarily changes in pitch that cue major linguistic units such as phrases and clauses [23–25]. Evidence of learning is that neonates express a preference for their mother's voice over another female voice [26]. Neonates also engage in behaviours that prolong passages of speech in their native tongue compared with passages from another language [27,28]. Ruling out the possibility that neonates simply tune to characteristics of their native language within hours of birth, foetuses heard story passages with unique prosodic properties (such as *The cat in the hat*) read aloud by their mother twice each day for six weeks before birth. After birth, the neonates discriminated the familiar story passages from a different story passage altogether [25]. In a replication, expecting mothers read one of two nursery rhymes aloud three times a day for four weeks daily starting at 33 weeks gestation. At 37 weeks, foetuses heard the familiar nursery rhyme and the unfamiliar rhyme read aloud by another talker. Foetal heartrate decreased more to the familiar than to the unfamiliar rhyme suggesting retention of learning [29]. A drawback of these studies is that test occurs a day after the training ends, making it difficult to

know how much exposure neonates need to form a memory and how long they can retain the memory after training ends.

The examples above summarize learning of information that infants receive through massive exposure (a minimum of 84 learning exposures in [25]) compared with a single SL experience. What about retention of linguistic information after brief exposure of the kind used in laboratory studies? Although neonates show brain correlates of learning after a few minutes of exposure to sequential probabilities in a statistical language [30], retention is susceptible to interference. Neonates who heard many repetitions of a novel word over a 2 min interval showed more forgetting 2 min later after exposure to repetitions of new linguistic material compared with neonates who listened to music during the interference interval [31]. In another study, neonates habituated to many repetitions of the same novel word over 30 trials, then dishabituated to the word after a 145 s break suggesting fragile retention [32]. The neonates did show a mild form of retention 24 h later where they heard the same word as the day before across 30 trials or a novel word. Infants hearing the same word across trials turned away from the stimulus more than infants hearing the novel word, showing a savings in learning [32]. Given the findings in Swain *et al.* [32], it is unclear how much savings neonates would exhibit after spaced exposure to a novel word separated by intervening vocabulary.

Retention of SL is still fragile at 6.5 months of age [33]. Given findings in the literature showing a critical role for sleep in retention of learning in infancy [34–38], my students and I introduced a sleep manipulation after familiarizing infants to a continuous stream of four bisyllabic words strung together in random order. Infants listened to the language for 7 min while sitting quietly on their caregiver's lap then slept in the laboratory or stayed awake with an experimenter quietly entertaining them. After the delay, we used a computer-automated head-turn preference procedure [39] to measure infants' listening times to words versus part words of the language (part words consisted of two syllables spanning words of the language). Infants in the nap group showed a significant interaction across two test blocks reflecting numerically longer listening times to part words on the first test block and a reversal of this pattern on the second block. Infants in the wakefulness group showed no discrimination of words and part words. Although the nap group did not show discrimination at the group level, block 1 discrimination varied with slow wave activity during non-rapid eye movement (NREM) sleep. This is a noteworthy finding given that slow wave sleep, the deepest form of NREM sleep, is a neural marker of retention of SL in adults [17,40,41]. Discrimination was short lived as there were no correlations between discrimination and slow wave activity in the second block of testing. We concluded that retention of statistical information is still fragile at 6.5 months with interference from exposure to part words during the test disrupting infants' retention. What we do not know from this study is how much savings infants this age may show on additional exposure.

We do know that by eight months of age, infants can extract and remember frequent words from a storybook text after many days of exposure [42]. An experimenter played pre-recorded versions of three stories to infants at each of 10 home visits. Infants came to the laboratory two weeks after the last home visit for a test of their retention of words from the stories. The test consisted of lists of highly frequent words from the stories versus words not in the stories. The experimental group listened longer, on average, to the lists containing story words compared with novel words demonstrating that eight month olds learned story words with intensive daily exposure that they retained over a two week delay. The procedure of embedding the words in stories also required infants to extract the words from running speech under conditions with much lower transitional probabilities than those in SL corpora, which typically have much higher transitions (e.g. 1.0 versus 0.33 in [1]). A drawback of this study is that infants heard each of the words about 13 times on each of the 10 visits making it difficult to know how much exposure infants this age need to extract and remember new words.

By 15 months of age, infants retain some memory of a single SL experience aided by sleep [34]. Earlier work established that infants detect non-adjacent dependencies in auditory strings such as vot-kicey-jic and pel-wadim-rud if the middle word comes from a set of 18 or more items [43,44], a manipulation that decreases transitional probabilities between adjacent words relative to non-adjacent ones. Infants do not detect non-adjacent dependencies if the middle word is drawn from a set of three words, presumably because they still track the adjacent word dependencies which are non-informative in the language.

My co-workers and I first tested retention across a 4 h delay [34]. Infants in nap and wakefulness groups heard the non-adjacent dependency language with high variability over an exposure period of 15 min. Infants in the nap-control group heard the language with the middle word drawn from a set of three for an equivalent duration to determine whether sleep alone would lead to improvement. After the delay, we tested infants on strings of the form vot-kicey-jic and pel-wadim-rud or vot-kicey-rud and pel-wadim-jic where the second two strings violated the specific non-adjacent dependencies from familiarization. We considered two different dependent measures. One measure, the difference in infant listening time to strings with legal versus illegal non-adjacent dependencies, reflects specific memory. This is the effect we obtain in 15 month olds when we test them immediately after familiarization [44]. Another measure reflects generalization of the non-adjacent pattern, regardless of the specific words entailed. In this case, infants should listen longer, on average, to the first trial type they encounter at test compared with the other type, regardless of whether that type conforms or not to the specific non-adjacent dependencies heard during familiarization. Although infants in the wakefulness group showed a specific memory effect after a 4 h delay, infants in the high-variability nap group showed generalization. This is consistent with loss in the fidelity of the memory over sleep. Otherwise, infants in the nap condition should have exhibited specific memory. Infants in the low-variability nap control condition showed neither effect.

We next asked whether 15 month olds would show retention of non-adjacencies over a longer delay [35]. As in our first study, the nap group slept in the 4 h interval after familiarization, whereas the wakefulness group stayed awake. We tested all infants 24 h later. Nap infants showed generalization compared with the wakefulness group who showed no discrimination on either dependent measure, findings consistent with the idea that naps preserved some aspect of learning in 15 month old infants. The findings for the nap group are consistent with preservation of the memory in a

less specific form that permits generalization. In contrast, the no nap group appears to lose all memory of the SL exposure.

In sum, differing experimental protocols make it challenging to ascertain a developmental profile of SL retention with some studies investigating retention after a single learning experience and others after many exposures. Although findings suggest retention over a two week period by eight months, after many days of highly focused exposure [42], findings after a single exposure suggest fragile retention up through 15 months. Infants this age retain learning over a 24 h delay with loss in the fidelity of their memories that permits generalization [35]. In contrast, adults consolidate SL rapidly, evidenced by equivalent or better performance a day after a single learning exposure [12,16,17]. How do these outcomes square with what we know about memory systems recruited during SL in early development? I start the next section by characterizing memory systems in the adult brain. Following that, I report evidence for the memory systems involved in SL based on imaging adult subjects. In a subsequent section, I discuss a theory of memory consolidation and the mechanisms involved as applied to retention of adult SL. Finally, I describe the development of key memory systems and their implications for retention of SL in early development.

3. Candidate memory systems for statistical learning

Evidence supports the idea that the adult brain contains partly distinct memory systems that interact via larger-scale networks [45–49]. Properties of learning are thought to differ by memory system based on the learning signals they require and the different dynamics their neural architectures support [46]. Networks instrumental in SL involve the neocortex, hippocampus and corticostriatal system (see below). For instance, cortex is thought to be a self-organizing memory system that builds up many overlapping associations supporting integration of information [45]. Although overlapping associations are at an advantage for supporting generalization [45,46,50], this architecture does not encode highly similar inputs in a form that keeps them distinct, leading to interference [51]. In contrast, the hippocampus has a sparse architecture that represents highly similar inputs in distinct neural patterns [45]. Sparse coding in the dentate gyrus of the hippocampus is thought to support the formation of specific episodic memories, with the highly recurrent architecture of CA3 thought to be a powerful pattern associator for retrieving memories with partial input [46]. The corticostriatal memory system is highly dependent on reward, using its presence or absence to select whether to perform an action [21]. Hippocampus, which does not appear to require reward, automatically forms highly specific memories [46]. These learning signals and network dynamics also result in different learning rates [45,50]. The hippocampus supports rapid encoding of associations and robust retention of highly specific details, sometimes after a single exposure. In contrast, cortex is thought to require many more exposures to achieve the same level of encoding [45,46,50].

Evidence for involvement of different memory systems in SL comes from studies using functional magnetic resonance imaging (fMRI) with adults who reveal extensive activation of cortex, hippocampus and striatum during learning. Cortical regions activate according to the mode of learning. During auditory SL, the superior temporal gyrus—a region implicated in auditory perception, speech processing and forward speech prediction [52–54]—activates differentially to statistically predictable versus unpredictable transitions [55–58]. Auditory SL also activates the inferior frontal gyrus during test performance [55] and online learning of novel sequential structure [59] with some indication that this region may reflect recognition of accumulated statistical information [55,59]. There is also evidence that cortical regions activate to speech in three month olds [60]. Visual SL, in contrast, activates visual lateral occipital cortex associated with object selection and ventral occipitotemporal cortex associated with word selection [14] demonstrating that these cortical regions detect information predictable in time (noted in [14]).

The hippocampus also contributes to SL exhibiting greater activation on trials that predict an associated upcoming stimulus versus trials that do not [14,15,41,61]. Schapiro et al. [61] investigated representational changes supporting SL. Adults watched a continuous, centrally presented stream of sequential pairs of visual stimuli presented in random order. Multivoxel pattern analysis of fMRI images before and immediately after training implicated the trisynaptic pathway of the hippocampus in representing temporal order. Regions CA2/CA3/DG of this pathway reflected the encoding of temporal statistics through an increase in the representational similarity of the first item in a pair to the second item. Perirhinal cortex and CA1 from the monosynaptic circuit of the hippocampus showed no shaping as a function of temporal statistics, instead exhibiting bold activity consistent with bidirectional associations (but see [50] for in-depth discussion of this issue). Although this study imaged participants in visual SL, research also implicates the hippocampus in auditory SL [41,62].

Studies also report striatal activation in SL [56,57], with activity in the caudate body predicting discrimination of statistically predictable versus unpredictable transitions in visual and auditory SL [14,57].

So far, I have addressed the memory systems activated during learning, but we must also consider the role of memory systems in long-term retention. Furthermore, because any form of long-term retention necessarily involves sleep, we must address theories of sleep consolidation. As we see, the properties of sleep consolidation differ by memory system.

4. Processes contributing to memory consolidation and retention

A rapidly growing literature implicates sleep in memory consolidation. Active systems consolidation (ASC), a leading theory of sleep consolidation [63], builds upon complementary learning systems theory and the idea that learners encode new information simultaneously in cortical and hippocampal stores with hippocampus forming indices to cortex that provide a unique spatial and temporal context for later retrieval [45]. Up to now, we have discussed only the associative properties of the hippocampus that bind elements, but the hippocampus also plays a crucial role in consolidation of new learning. Specifically, hippocampally generated sleep neural replay is thought to synchronize with other brain oscillations to consolidate memories in cortex [63,64]. Key oscillations include sharp-wave ripples, high-frequency, highly synchronous neural oscillations arising in CA3 and CA1 pyramidal cells [65] reflecting replay of activity experienced during wakefulness [66]. Sleep spindles, short, high-frequency (9–15 Hz), thalamocortical oscillations thought

to reflect communication between brain regions [67], are a second form of oscillation. Finally, slow wave activity, high amplitude, low-frequency oscillations (1–4.5 Hz) originating in neocortex are thought to coordinate sharp-wave ripples and spindles [64]. It is through synchronized activity of these oscillations that the hippocampus appears to support rapid consolidation of one-trial learning [63]. Sleep thus preserves the details of a memory through neural replay with the hippocampus reactivating and replaying memories from the day, strengthening them in cortical structures [63]. In support of this theory, sleep spindles correlate with retention in language-learning tasks [38,68,69] and percentage of time in slow wave sleep [17,41] and slow wave activity [68] correlate with retention after SL and artificial language learning, respectively. Sharp-wave ripples can only be observed by recording directly from implanted electrodes in animals [66] or with direct intracranial electroencephalogram recordings in human epilepsy patients [70].

Consistent with ASC, Durrant *et al.* [41] reported initial parahippocampal involvement after SL of temporal tone sequences that diminished 24 h later accompanied by increased activation of the striatal memory system involving the putamen and caudate, as well as the planum temporale. Also consistent with ASC, initially strong connectivity between the parahippocampus, ventromedial prefrontal cortex and caudate shifted to greater connectivity between the putamen and planum temporale after 24 h. Correlations between the amount of time in slow wave sleep and the increase in discrimination of statistically predictable versus unpredictable tone sequences, as well as the changes in the activation patterns after the delay, suggest that sleep aided in consolidation [41]. Durrant *et al.* argued that initial activation in parahippocampal gyrus made sense for their task involving temporal associations in tone sequences as an area that acts as a gateway between the hippocampus and the superior temporal gyrus [41].

Given robust sleep consolidation observed in adults, how do we explain poor retention observed in infants? Here, we must consider development of the underlying memory systems.

5. Development of memory systems

Although brain connectivity undergoes significant development in the first year of life [71], key structures in the hippocampus have a relatively extended developmental trajectory. Parts of CA1 making up the monosynaptic circuit of the hippocampus develop rapidly over the first 2 years of life, particularly those parts receiving input from cortical regions [72]. However, the portions of CA1 receiving projections from CA3 of the trisynaptic circuit of the hippocampus are relatively less developed [72]. This fact is consistent with the slow development of the trisynaptic circuit, which does not begin to support fairly robust pattern separation until about 4 years of age in human behaviour [73]. Development in non-human primates follows similar behavioural and neuroanatomical trajectories adjusted for human age [72]. Although sharp-wave ripples are one of the earliest oscillations to manifest in development [74], Gómez & Edgin [75] argued that there cannot be mature ASC until the projections from CA3 to CA1 are sufficiently developed for sharp-wave ripples originating in CA3 to propagate to CA1. Although CA1 may propagate some ripples to cortex during infancy, we would not expect

oscillations between the hippocampus and prefrontal cortex to be mature enough to support robust consolidation of the type seen in adults. Consistent with this proposal, and although connectivity of the hippocampus with other memory systems appears to be developing, it is not until about 2 years of age that this connectivity begins to exhibit mature default network activity [71]. Development of long-range circuitry [71] combined with development in hippocampal trisynaptic circuitry [72] led Gómez & Edgin [75] to propose that the hippocampus may not begin to support functional neural replay with sharp-wave ripple propagation to cortex until after 2 years of age. Behavioural outcomes indicating robust consolidation of learning with sleep in children 2 years of age and older [76,77] compared with infants who show more modest behavioural effects [34,35] are consistent with this proposal. Gómez and Edgin proposed that ACS emerges after 24 months of age in humans as the trisynaptic circuit of hippocampus, prefrontal cortex and connections between these brain regions develop more fully [75]. I propose here that before this time, SL is likely dominated by cortex, the corticostriatal loop and CA1 in the monosynaptic circuit of the hippocampus as it becomes increasingly mature over the first 2 years of life [72] and which also has a slow-learning profile relative to the trisynaptic circuit [50,78]. Regarding auditory SL, cortical networks associated with adult language respond to speech in early infancy and thus are functional [60]. Aspects of the corticostriatal network also appear to develop early as indicated in neural [79] and behavioural measures [80,81]. What are the implications for long-term retention of SL before the hippocampus and connections to prefrontal cortex exhibit sufficient maturation to support sleep neural replay?

Two theories of cortical consolidation apply. According to the synaptic homeostasis hypothesis [82], sleep acts to downscale synaptic energy that builds up from behavioural experience during the day. Downscaling is thought to occur monotonically according to the strength of the association so that stronger associations (with greater buildup of synaptic energy) should survive downscaling to a greater extent than weaker ones. Sleep consolidation through cortical long-term potentiation is a competing possibility [83]. Although we would expect stronger associations to consolidate to a greater degree than weaker ones, by this view, we would not expect robust consolidation given the immaturity of cortical networks in early development. Thus, both theories are consistent with the outcomes observed in studies of retention of SL where sleep in infants appears to preserve the more reliable connections from learning to a greater degree than the less reliable ones but not the fidelity of a memory [34,35]. In particular, given that all of the exemplars in the language used by Gómez *et al.* [34,35] have the same characteristic prosodic structure, reflecting primary stress and a rise in pitch across the words of the sentence, the prosodic envelope of the strings and the presence of a non-adjacent pattern are the most reliable characteristics from learning. In contrast, the specific non-adjacent dependencies occur half as frequently and thus are less reliable. If we assume retention in proportion to the reliability of the information encountered during learning, then we might expect that infants would retain some information about the non-adjacent dependency but not the details of the exact words making up the non-adjacent associations. If so, exposure to test strings with the illegal adjacent dependency might reactivate a non-specific memory of the lawful relationship between non-adjacent elements in strings that infants can use

to track new non-adjacencies, consistent with the observed effect in our studies [34,35]. Given the developmental trajectory of the trisynaptic circuit [72], we would expect to see more rapid consolidation and increased fidelity of memory for specific statistically reliable transitions after 24 months of age [75]. Before this time, infants should show modest retention, perhaps first expressed in savings in learning and in a slow buildup of retention of the more reliable properties of the stimulus after repeated exposures. The poor retention of SL exhibited at 6.5 and 15 months of age [33–35] accords with the slow rate of real-world language learning where it takes many months of exposure for infants to express fine tuning of the phonological characteristics of their native language in behaviour [6–8].

6. Implications of multiple memory systems for statistical learning and language acquisition

Here, I make a case for the importance of considering underlying memory systems in SL. Given the time-dependent nature of memory formation, which involves a cascade of cellular and molecular processes and phases, we have little reason to assume that what we measure immediately after familiarization is an accurate depiction of SL as it plays out in real-world learning—the same process, just more of it. This is true for adults and for infants. In adults, memory systems interact differently immediately after learning and a day later, after sleep consolidation [41]. Infants, with their dependence on cortical memory systems may learn and remember altogether differently than adults [75], with implications not necessarily specific to SL. Thus, considering the different learning and retention properties of memory systems is a rich source of ideas for understanding SL and for constraining theory in a neurally plausible manner.

An open question is the extent of hippocampal contributions to SL in infancy given the unique contributions of the monosynaptic circuit of the hippocampus, containing CA1, to SL (see [50]). Although CA1 develops much earlier than the trisynaptic circuitry thought necessary for supporting coordinated sharp-wave ripple activity and robust consolidation over sleep, CA1 undergoes substantial development over the first 2 years of life [72]. It may be that CA1 contributes more substantially to SL later in this period than it does earlier in development when very young infants show clear evidence of tuning to the statistical patterns of their language [6–8].

A consideration of multiple memory systems may also help us address a problem plaguing theories of learning having to do with how learners might recover from erroneous generalizations [84]. Properties of cortical learning are well equipped to deal with this problem. If cortex is a slow-learning system that easily forgets, at least in the earliest stages of learning, then infants are unlikely to form a permanent trace, unless information is highly redundant in the signal. Using language as an example, we would expect SL to be guided primarily by the most statistically reliable aspects of language structure, the very characteristics that support slow learning of the sound structure of language. Furthermore, until the hippocampus is sufficiently mature, we would not expect infants to engage rapid learning and retention of idiosyncratic sequential structure. Evidence

consistent with this view comes from studies of deferred imitation. Although infants can learn and express individual actions after exposure to sequences of arbitrarily related actions, they show little sequencing of such actions before 22 months of age [21]. It is not until 28 months, an age that maps onto increasing hippocampal development, that infants retain novel, arbitrarily related sequences of actions over a delay [22].

The implications of developing memory systems may also suggest limitations in drawing conclusions about the cognitive and neural properties of infant SL based on work with adults, especially if infants and adults are relying on different memory systems. Indeed, the developmental profile of the hippocampus and its communication with cortical regions [85] accords with the decline of the critical period for language learning after 7 years of age and extending through puberty [86]. Although adults acquire a second language with sufficient exposure [87], I argue here that adults are likely to rely on different mechanisms than infants for memory formation, contrary to views proposing continuity in SL mechanisms. With that said, Thiessen [88] points to factors that explain differences in infant and adult SL that do not require different memory mechanisms (also see [89]). For one, infants may encode statistical information in a noisier representation than adults, and thus, may have more difficulty updating older representations with new information. Thiessen also suggests that infants may be more prone to weighting irrelevant input more heavily than adults, slowing learning. I have suggested that infants are less likely to retain irrelevant input over a delay because of its lower reliability in the input. For instance, infants may hear many people uttering the same statistically reliable structure, but each talker will exhibit different indexical characteristics, resulting in lower statistical reliability for these aspects of the input. We would fully expect synaptic homeostasis [82] to downscale less statistically reliable input to a greater degree than the more statistically reliable structure uttered by each talker. The issue of infants having noisier representations may be more critical to factor into theories of memory updating given the degree to which updating depends on new exemplars activating existing representations [90]. This may also be a factor when considering individual differences in SL [91].

Finally, we must consider how the massed exposures we employ in laboratory studies maps onto real-world language acquisition where structures are separated in time. This is a consideration given that intervening linguistic materials appear to interfere with retention in very young infants [31].

In short, despite many important discoveries, we have a great deal more to discover about retention of SL, a process that most certainly plays out in multiple learning systems with different trajectories of development and retention. Frost et al. [92] recently proposed an account of SL involving domain-general computational principles constrained by the modality of the computation and the supporting brain region. I propose here that brain development further constrains SL.

6

rstb.royalsocietypublishing.org   Phil. Trans. R. Soc. B 372: 20160054

References

1. Saffran JR, Aslin RN, Newport EL. 1996 Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928. (doi:10.1126/science.274.5294.1926)

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

3. Endress AD, Mehler J. 2009 The surprising power of statistical learning: when fragment knowledge leads to false memories of unheard words. *J. Mem. Lang.* **60**, 351–367. (doi:10.1016/j.jml.2008.10.003)

4. Ferry AL, Fló A, Brusini P, Cattarossi L, Macagno F, Nespor M, Mehler J. 2015 On the edge of language acquisition: inherent constraints on encoding multisyllabic sequences in the neonate brain. *Dev. Sci.* **19**, 488–503. (doi:10.1111/desc.12323)

5. Saksida A, Langus A, Nespor M. 2016 Co-occurrence statistics as a language-dependent cue for speech segmentation. *Dev. Sci.* (doi:10.1111/desc.12390)

6. Werker JF, Tees RC. 1984 Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* **7**, 49–63. (doi:10.1016/S0163-6383(84)80022-3)

7. Jusczyk PW, Luce PA, Charles-Luce J. 1994 Infants' sensitivity to phonotactic patterns in the native language. *J. Mem. Lang.* **33**, 630. (doi:10.1006/jmla.1994.1030)

8. Mattys SL, Jusczyk PW. 2001 Phonotactic cues for segmentation of fluent speech by infants. *Cognition* **78**, 91–121. (doi:10.1016/S0010-0277(00)00109-8)

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

10. Rudy JW. 2008 *The neurobiology of learning and memory*. Sunderland, MA: Sinauer Associates.

11. Ebbinghaus H. 1913 *Memory: a contribution to experimental psychology*. New York, NY: University Microfilms/Columbia University.

12. Kim R, Seitz A, Feenstra H, Shams L. 2009 Testing assumptions of statistical learning: is it long-term and implicit? *Neurosci. Lett.* **461**, 145–149. (doi:10.1016/j.neulet.2009.06.030)

13. Saffran JR, Newport EL, Aslin RN, Tunick RA, Barrueco S. 1997 Incidental language learning: listening (and learning) out of the corner of your ear. *Psychol. Sci.* **8**, 101–105. (doi:10.1111/j.1467-9280.1997.tb00690.x)

14. Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. 2009 Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J. Cogn. Neurosci.* **21**, 1934–1945. (doi:10.1162/jocn.2009.21131)

15. Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. 2010 Implicit perceptual anticipation triggered by statistical learning. *J. Neurosci.* **30**, 11 177–11 187. (doi:10.1523/JNEUROSCI.0858-10.2010)

16. Arciuli J, Simpson I. 2012 Statistical learning is lasting and consistent over time. *Neurosci. Lett.* **517**, 133–135. (doi:10.1016/j.neulet.2012.04.045)

17. Durrant SJ, Taylor C, Cairney S, Lewis PA. 2011 Sleep-dependent consolidation of statistical learning. *Neuropsychologia* **49**, 1322–1331. (doi:10.1016/j.neuropsychologia.2011.02.015)

18. Rovee-Collier C. 1999 The development of infant memory. *Curr. Dir. Psychol. Sci.* **8**, 80–85. (doi:10.1111/1467-8721.00019)

19. Hill WL, Borovsky D, Rovee-Collier C. 1988 Continuities in infant memory development. *Dev. Psychobiol.* **21**, 43–62. (doi:10.1002/dev.420210104)

20. Cincotta CM, Seger CA. 2007 Dissociation between striatal regions while learning to categorize via feedback and via observation. *J. Cogn. Neurosci.* **19**, 249–265. (doi:10.1162/jocn.2007.19.2.249)

21. Seger CA. 2008 How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neurosci. Biobehav. Rev.* **32**, 265–278. (doi:10.1016/j.neubiorev.2007.07.010)

22. Bauer PJ, Hertsgaard LA, Dropik P, Daly BP. 1998 When even arbitrary order becomes important: developments in reliable temporal sequencing of arbitrarily ordered events. *Memory* **6**, 165–198. (doi:10.1080/741942074)

23. Busnel MC, Granier-Deferre C. 1983 And what of fetal audition? In *The behavior of human infants* (eds A Oliverio, M Zappella), pp. 93–126. New York, NY: Springer.

24. Lecanuet JP, Granier-Deferre C, Jacquet AY, Busnel MC. 1992 Decelerative cardiac responsiveness to acoustical stimulation in the near term fetus. *Q. J. Exp. Psychol. B* **44**, 279–303.

25. DeCasper AJ, Spence MJ. 1986 Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behav. Dev.* **9**, 133–150. (doi:10.1016/0163-6383(86)90025-1)

26. DeCasper AJ, Fifer WP. 1980 Of human bonding: newborns prefer their mothers' voices. *Science* **208**, 1174–1176. (doi:10.1126/science.7375928)

27. Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C. 1988 A precursor of language acquisition in young infants. *Cognition* **29**, 143–178. (doi:10.1016/0010-0277(88)90035-2)

28. Moon C, Cooper RP, Fifer WP. 1993 Two-day-olds prefer their native language. *Infant Behav. Dev.* **16**, 495–500. (doi:10.1016/0163-6383(93)80007-U)

29. DeCasper AJ, Lecanuet J-P, Busnel M-C, Granier-Deferre C, Maugeais R. 1994 Fetal reactions to recurrent maternal speech. *Infant Behav. Dev.* **17**, 159–164. (doi:10.1016/0163-6383(94)90051-5)

30. Teinonen T, Fellman V, Näätänen R, Alku P, Huotilainen M. 2009 Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* **10**, 21. (doi:10.1186/1471-2202-10-21)

31. Benavides-Varela S, Gómez DM, Macagno F, Bion RA, Peretz I, Mehler J. 2011 Memory in the neonate brain. *PLoS ONE* **6**, e27497. (doi:10.1371/journal.pone.0027497)

32. Swain IU, Zelazo PR, Clifton RK. 1993 Newborn infants' memory for speech sounds retained over 24 hours. *Dev. Psychol.* **29**, 312–323. (doi:10.1037/0012-1649.29.2.312)

33. Simon KN, Werchan D, Goldstein MR, Sweeney L, Bootzin RR, Nadel L, Gómez RL. 2016 Sleep confers a benefit for retention of statistical language learning in 6.5 month old infants. *Brain Lang.* (doi:10.1016/j.bandl.2016.05.002)

34. Gomez RL, Bootzin RR, Nadel L. 2006 Naps promote abstraction in language-learning infants. *Psychol. Sci.* **17**, 670–674. (doi:10.1111/j.1467-9280.2006.01764.x)

35. Hupbach A, Gomez RL, Bootzin RR, Nadel L. 2009 Nap-dependent learning in infants. *Dev. Sci.* **12**, 1007–1012. (doi:10.1111/j.1467-7687.2009.00837.x)

36. Seehagen S, Konrad C, Herbert JS, Schneider S. 2015 Timely sleep facilitates declarative memory consolidation in infants. *Proc. Natl Acad. Sci. USA* **112**, 1625–1629. (doi:10.1073/pnas.1414000112)

37. Horváth K, Myers K, Foster R, Plunkett K. 2015 Napping facilitates word learning in early lexical development. *J. Sleep Res.* **24**, 503–509. (doi:10.1111/jsr.12306)

38. Friedrich M, Wilhelm I, Born J, Friederici AD. 2015 Generalization of word meanings during infant sleep. *Nat. Commun.* **6**, 6004. (doi:10.1038/ncomms7004)

39. Kemler Nelson DG, Jusczyk PW, Mandel DR, Myers J, Turk A, Gerken L. 1995 The head-turn preference procedure for testing auditory perception. *Infant Behav. Dev.* **18**, 111–116. (doi:10.1016/0163-6383(95)90012-8)

40. Arciuli J, Vakulin A, D'Rozario A, Openshaw H, Stevens D, McEvoy D, Wong K, Rae C, Grunstein R. 2015 Is statistical learning affected by sleep apnoea? In *Proc. the EuroAsianPacific Joint Conference on Cognitive Science, Turin, Italy* (eds G Airenti, B Bara, G Sandini). Retrieved from http://ceur-ws.org/Vol-1419/paper0080.pdf.

41. Durrant SJ, Cairney SA, Lewis PA. 2012 Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cereb. Cortex.* **23**, 2467–2478. (doi:10.1093/cercor/bhs244)

42. Jusczyk PW, Hohne EA. 1997 Infants' memory for spoken words. *Science* **277**, 1984–1986. (doi:10.1126/science.277.5334.1984)

43. Gómez RL. 2002 Variability and detection of invariant structure. *Psychol. Sci.* **13**, 431–436. (doi:10.1111/1467-9280.00476)

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

45. McClelland JL, McNaughton BL, O'Reilly RC. 1995 Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457. (doi:10.1037/0033-295X.102.3.419)

46. O'Reilly RC, Munakata Y, Frank MJ, Hazy TE, Contributors. 2012 *Computational cognitive neuroscience*, 1st edn. Wiki Book. See http://ccnbook.colorado.edu.

47. Poldrack RA, Foerde K. 2008 Category learning and the memory systems debate. *Neurosci. Biobehav. Rev.* **32**, 197–205. (doi:10.1016/j.neubiorev.2007.07.007)

48. Squire LR. 1992 Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* **99**, 195–231. (doi:10.1037/0033-295X.99.2.195)

49. Tulving E. 1985 How many memory systems are there? *Am. Psychol.* **40**, 385–398. (doi:10.1037/0003-066X.40.4.385)

50. Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA. 2017 Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160049. (doi:10.1098/rstb.2016.0049)

51. McCloskey M, Cohen NJ. 1989 Catastrophic interference in connectionist networks: the sequential learning problem. *Psychol. Learn. Motiv.* **24**, 109–165. (doi:10.1016/S0079-7421(08)60536-8)

52. Cunillera T, Càmara E, Toro JM, Marco-Pallares J, Sebastián-Galles N, Ortiz H, Pujol J, Rodríguez-Fornells A. 2009 Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* **48**, 541–553. (doi:10.1016/j.neuroimage.2009.06.069)

53. Hickok G. 2012 The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. *J. Commun. Disord.* **45**, 393–402. (doi:10.1016/j.jcomdis.2012.06.004)

54. Overath T, Cusack R, Kumar S, Von Kriegstein K, Warren JD, Grube M, Carlyon RP, Griffiths TD. 2007 An information theoretic characterisation of auditory encoding. *PLoS Biol.* **5**, e288. (doi:10.1371/journal.pbio.0050288)

55. Karuza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D. 2013 The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain Lang.* **127**, 46–54. (doi:10.1016/j.bandl.2012.11.007)

56. McNealy K, Mazziotta JC, Dapretto M. 2006 Cracking the language code: neural mechanisms underlying speech parsing. *J. Neurosci.* **26**, 7629–7639. (doi:10.1523/JNEUROSCI.5501-05.2006)

57. Plante E, Patterson D, Gómez R, Almryde KR, White MG, Asbjørnsen AE. 2015 The nature of the language input affects brain activation during learning from a natural language. *J. Neurolinguist* **36**, 17–34. (doi:10.1016/j.jneuroling.2015.04.005)

58. Plante E, Almryde K, Patterson DK, Vance CJ, Asbjørnsen AE. 2015 Language lateralization shifts with learning by adults. *Laterality* **20**, 306–325. (doi:10.1080/1357650X.2014.963597)

59. Plante E, Patterson D, Dailey NS, Kyle RA, Fridriksson J. 2014 Dynamic changes in network activations characterize early learning of a natural language. *Neuropsychologia* **62**, 77–86. (doi:10.1016/j.neuropsychologia.2014.07.007)

60. Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L. 2002 Functional neuroimaging of speech perception in infants. *Science* **298**, 2013–2015. (doi:10.1126/science.1077066)

61. Schapiro AC, Kustner LV, Turk-Browne NB. 2012 Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* **22**, 1622–1627. (doi:10.1016/j.cub.2012.06.056)

62. Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB. 2014 The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci.* **26**, 1736–1747. (doi:10.1162/jocn_a_00578)

63. Diekelmann S, Born J. 2010 The memory function of sleep. *Nat. Rev. Neurosci.* **11**, 114–126. (doi:10.1038/nrn2762)

64. Mölle M, Marshall L, Gais S, Born J. 2002 Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *J. Neurosci.* **22**, 10 941–10 947.

65. Chrobak JJ, Buzsáki G. 1994 Selective activation of deep layer (V–VI) retrohippocampal cortical neurons during hippocampal sharp waves in the behaving rat. *J. Neurosci.* **14**, 6160–6170.

66. Wilson M, McNaughton B. 1994 Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679. (doi:10.1126/science.8036517)

67. Anders TF, Emde RN, Parmelee AH. 1971 *A manual of standardized terminology, techniques and criteria for scoring of states of sleep and wakefulness in newborn infants*. Los Angeles, CA: UCLA Brain Information Service/BRI Publications Office, NINDS Neurological Information Network.

68. Tamminen J, Lambon Ralph MA, Lewis PA. 2013 The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *J. Neurosci.* **33**, 15 376–15 381. (doi:10.1523/JNEUROSCI.5093-12.2013)

69. Tamminen J, Payne JD, Stickgold R, Wamsley EJ, Gaskell MG. 2010 Sleep spindle activity is associated with the integration of new memories and existing knowledge. *J. Neurosci.* **30**, 14 356–14 360. (doi:10.1523/JNEUROSCI.3028-10.2010)

70. Staresina BP, Bergmann TO, Bonnefond M, van der Meij R, Jensen O, Deuker L, Elger CE, Axmacher N, Fell J. 2015 Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nat. Neurosci.* **18**, 1679–1686. (doi:10.1038/nn.4119)

71. Gao W, Zhu H, Giovanello KS, Smith JK, Shen D, Gilmore JH, Lin W. 2009 Evidence on the emergence of the brain's default network from 2-week-old to 2-year-old healthy pediatric subjects. *Proc. Natl Acad. Sci. USA* **106**, 6790–6795. (doi:10.1073/pnas.0811221106)

72. Lavenex P, Banta Lavenex P. 2013 Building hippocampal circuits to learn and remember: insights into the development of human memory. *Behav. Brain Res.* **254**, 8–21. (doi:10.1016/j.bbr.2013.02.007)

73. Ribordy F, Jabès A, Lavenex PB, Lavenex P. 2013 Development of allocentric spatial memory abilities in children from 18 months to 5 years of age. *Cogn. Psychol.* **66**, 1–29. (doi:10.1016/j.cogpsych.2012.08.001)

74. Buzsaki G. 2006 *Rhythms of the brain*. Oxford, UK: Oxford University Press.

75. Gómez RL, Edgin JO. 2016 The extended trajectory of hippocampal development: implications for early memory development and disorder. *Dev. Cogn. Neurosci.* **18**, 57–69. (doi:10.1016/j.dcn.2015.08.009)

76. Kurdziel L, Duclos K, Spencer RM. 2013 Sleep spindles in midday naps enhance learning in preschool children. *Proc. Natl Acad. Sci. USA* **110**, 17 267–17 272. (doi:10.1073/pnas.1306418110)

77. Williams SE, Horst JS. 2014 Goodnight book: sleep consolidation improves word learning via storybooks. *Front. Psychol.* **5**, 184. (doi:10.3389/fpsyg.2014.00184)

78. Nakashiba T, Young JZ, McHugh TJ, Buhl DL, Tonegawa S. 2008 Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science* **319**, 1260–1264. (doi:10.1126/science.1151120)

79. Chugani HT, Phelps ME. 1986 Maturational changes in cerebral function in infants determined by 18FDG positron emission tomography. *Science* **231**, 840–843. (doi:10.1126/science.3945811)

80. Amso D, Davidow J. 2012 The development of implicit learning from infancy to adulthood: item frequencies, relations, and cognitive flexibility. *Dev. Psychobiol.* **54**, 664–673. (doi:10.1002/dev.20587)

81. Werchan DM, Collins AGE, Frank MJ, Amso D. 2015 8-Month-old infants spontaneously learn and generalize hierarchical rules. *Psychol. Sci.* **26**, 805–815. (doi:10.1177/0956797615571442)

82. Tononi G, Cirelli C. 2014 Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. *Neuron* **81**, 12–34. (doi:10.1016/j.neuron.2013.12.025)

83. Aton SJ, Broussard C, Dumoulin M, Seibt J, Watson A, Coleman T, Frank MG. 2013 Visual experience and subsequent sleep induce sequential plastic changes in putative inhibitory and excitatory cortical neurons. *Proc. Natl Acad. Sci. USA* **110**, 3101–3106. (doi:10.1073/pnas.1208093110)

84. Pinker S. 1995 Language acquisition. In *Language: an invitation to cognitive science*, vol. 1 (ed. DN Osherson), pp. 135–182. Cambridge, MA: MIT Press.

85. Ghetti S, Bunge SA. 2012 Neural changes underlying the development of episodic memory during middle childhood. *Dev. Cogn. Neurosci.* **2**, 381–395. (doi:10.1016/j.dcn.2012.05.002)

86. Johnson JS, Newport EL. 1989 Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognit. Psychol.* **21**, 60–99. (doi:10.1016/0010-0285(89)90003-0)

87. Birdsong D, Molis M. 2001 On the evidence for maturational constraints in second-language acquisition. *J. Mem. Lang.* **44**, 235–249. (doi:10.1006/jmla.2000.2750)

88. Thiessen ED. 2017 What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Phil. Trans. R. Soc. B* **372**, 20160056. (doi:10.1098/rstb.2016.0056)

89. Mareschal D, French RM. 2017 TRACX2: a connectionist autoencoder using graded chunks to model infant visual statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160057. (doi:10.1098/rstb.2016.0057)

90. Hupbach A, Gomez R, Hardt O, Nadel L. 2007 Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learn. Mem.* **14**, 47–53. (doi:10.1101/lm.365707)

91. Arciuli J. 2017 The multi-component nature of statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160058. (doi:10.1098/rstb.2016.0058)

92. Frost R, Armstrong BC, Siegelman N, Christiansen MH. 2015 Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn. Sci.* **19**, 117–125. (doi:10.1016/j.tics.2014.12.010)

9

rstb.royalsocietypublishing.org *Phil. Trans. R. Soc. B* **372**: 20160054