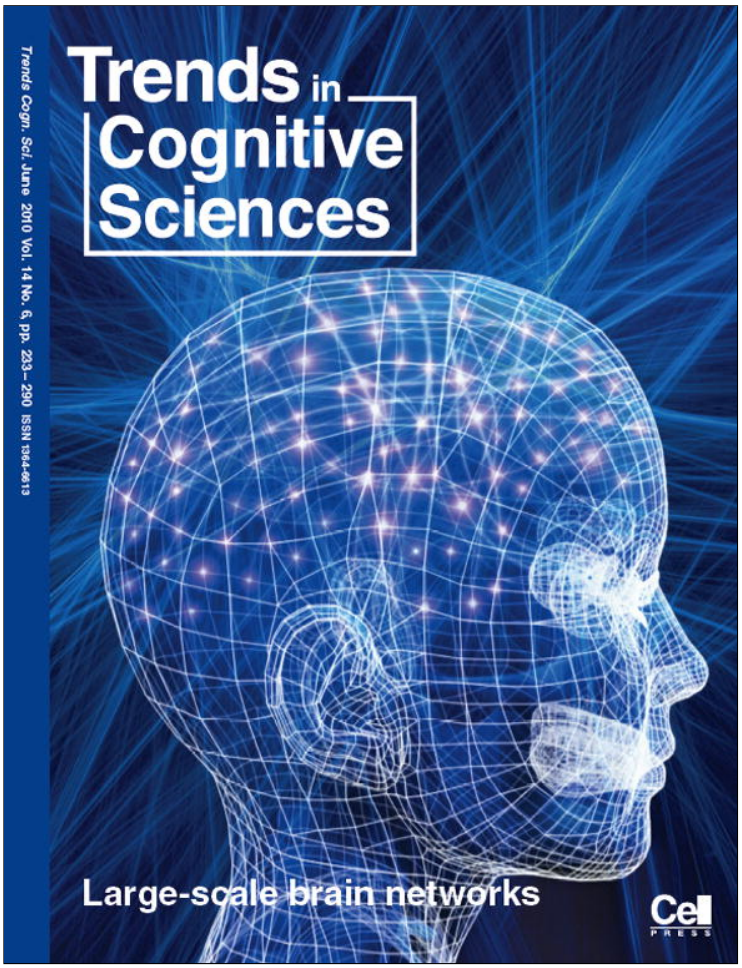

Opinion

Cell PRESS

General cognitive principles for learning structure in time and space

Michael H. Goldstein[1], Heidi R. Waterfall[1,2], Arnon Lotem[3], Joseph Y. Halpern[4], Jennifer A. Schwade[1], Luca Onnis[5] and Shimon Edelman[1]

[1] 211 Uris Hall, Department of Psychology, Cornell University, Ithaca, NY 14853, USA
[2] Department of Psychology, University of Chicago, Chicago, IL 60637, USA
[3] Department of Zoology, Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
[4] Department of Computer Science, 4144 Upson Hall, Cornell University, Ithaca, NY 14853, USA
[5] Department of Second Language Studies, and Center for Second Language Research, University of Hawaii, 1890 East-West Rd., Honolulu, HI 96822 USA

How are hierarchically structured sequences of objects, events or actions learned from experience and represented in the brain? When several streams of regularities present themselves, which will be learned and which ignored? Can statistical regularities take effect on their own, or are additional factors such as behavioral outcomes expected to influence statistical learning? Answers to these questions are starting to emerge through a convergence of findings from naturalistic observations, behavioral experiments, neurobiological studies, and computational analyses and simulations. We propose that a small set of principles are at work in every situation that involves learning of structure from patterns of experience and outline a general framework that accounts for such learning.

The goal of development is to learn structure in time and space

In the *Principles of Psychology*, William James ([1]; v.I, p. 488) (http://psychclassics.yorku.ca/James/Principles/) illustrated the fundamental challenge of sensorimotor development:

> Experience, from the very first, presents us with concreted objects, vaguely continuous with the rest of the world which envelops them in space and time, and potentially divisible into inward elements and parts. . .The baby, assailed by eye, ear, nose, skin and entrails at once, feels it all as one great blooming, buzzing confusion[.]

Developing cognitive systems overcome 'confusion' by discovering ways in which reality can be structured. They extract reliable units and relationships from the input (e.g. co-occurring sequences of phonemes and the regularities in their juxtaposition [2,3]), thereby becoming capable of principled, systematic generalization over those units [4,5] – the epitome of sophisticated cognition. Distilling spatial and temporal patterns in the stream of experience makes prediction of events and actions possible. Thus the primary goal of development – sensory, motor and, arguably, conceptual [6] – is to learn structure in space and time. Here we outline a theoretical framework that integrates the perceptual, cognitive and social mechanisms by which infants find and learn patterns from the stream of experience.

Much of what infants learn is organized serially (over time), and possibly hierarchically [7], including locomotion, social interaction and, ultimately, language [8]. Patterns are learned and used over multiple timescales simultaneously. For example, over short timescales, infants use transitional probabilities to extract co-located sequences of phonemes from continuous input [2] and then associate those with novel objects [9]. Over longer timescales, infants can detect parents' use of common grammatical constructions (e.g. *What's* . . .) and incorporate them in their own speech [10].

How are hierarchically structured sequences of objects, events or words learned from experience? Answers to this question are starting to emerge in several disciplines through methodologies that include naturalistic observations, behavioral experiments, neurobiological studies, and computational analyses and simulations. However, the very breadth of the existing work has hindered its integration into a single, coherent theory. We outline a simple conceptual computational framework that ties together the disparate strands of evidence and propose that a small set of principles operate in every situation that involves learning structure from patterns of experience, both spatial and temporal. Within this framework, we discuss how infants detect patterns in structured input and determine their relevance to interacting with the world.

Learning structure: the fundamental computational problem and a probable solution

Consider the computational problem of finding common structure, such as recurring parts, in a continuous stream of experience – a succession of scenes that might contain some of the same objects, or utterances that might share sound sequences. There are multiple levels to this problem: reusable units have to be discovered, patterns over the units inferred, and the reliability and predictive value of patterns assessed to allow generalization and prediction. Thus the task of learning structure gives rise to three related issues: how to find units in the input stream, how to infer potentially useful patterns and how to distill those patterns into reliable knowledge.

Corresponding author: Goldstein, M.H. (mhg26@cornell.edu).

1364-6613/$ – see front matter doi:10.1016/j.tics.2010.02.004 Trends in Cognitive Sciences 14 (2010) 249–258

The first intimations of a viable solution to the fundamental problem of learning from experience come from David Hume [11] (http://www.gutenberg.org/etext/4705). Concerning pattern discovery, Hume wrote, "All kinds of reasoning consist in nothing but a comparison, and a discovery of those relations, either constant or inconstant, which two or more objects bear to each other" ([11]; Part III, Sect. II). Furthermore, his realization that "all knowledge resolves itself into probability" ([11]; Part IV, Sect. I) points to a resolution of the problem of reliable inference.

In present-day theories of language acquisition, Hume's ideas have been echoed by Zellig Harris [3,12]. On the problem of discovering words in continuous speech, Harris noted that "when only a small percentage of all possible sound sequences actually occurs in utterances, one can identify the boundaries of words, and their relative likelihoods, from their sentential environment; this, even if one was not told (in words) that there exist such things as words" ([12], p. 32). In computational terms, the as-yet undifferentiated stimulus stream, buzzing with potential patterns, needs to be (i) parsed into units and (ii) *aligned* and *compared* to time-shifted versions of itself to reveal commonalities and differences, which then must be (iii) tested for statistical significance ([3]; Box 1). This process is iterated over multiple levels; first the raw input is parsed to reveal candidate units (e.g. phonemes) whose co-occurrences are analyzed to find higher-order structures (e.g. words and eventually syntax). In the first phase of learning structure ('going digital' [13]), continuous sound sequences are segmented into discrete units. In the second phase, ('going recursive' [13]) successive analysis of patterns over these units can yield a hierarchy of constructions [14,15].

This approach to structure learning gives rise to two related challenges. First, the alignment procedure, if invoked indiscriminately, runs into the computationally intractable need to compare every aspect of every piece of incoming data with records of all previous experience. Second, estimating the significance of an outcome against the background of multiple spurious alignments requires a large corpus of experience, which in turn exacerbates the problem of alignment and the bookkeeping it entails.

Previous efforts utilizing statistical ('distributional') learning have come short of resolving these difficulties. Most of the literature on distributional learning in language theorizes about or models highly circumscribed phenomena, usually one at a time [16]. Very few efforts chose to take on the most challenging problem: learning a generative grammar from large-scale, raw, unannotated corpora of child-directed language. Although the performance of the three systems that do address this problem [17–19] is encouraging (as judged by large-scale tests of coverage and generativity), it falls far short of human infant performance, and resorts to psychologically unrealistic techniques, such as multiple passes over the corpus.

We propose that human infants (as well as other young animals that must learn structure) resolve the

Box 1. The structure of variation sets

It is not necessary to be able to read, let alone understand, the languages in Figure I to identify the most prominent structural feature common to these clusters of utterances: each cluster forms a variation set [30], that is, the utterances in it are partially alignable. This feature, and the structures revealed through alignment and comparison, should, therefore, be readily apparent to a prelinguistic baby. Indeed, developmental studies indicate that infants learn structures that appear in variation sets particularly efficiently (H. Waterfall, PhD thesis, University of Chicago, 2006).

Partial repetition in variation sets can occur within one speaker's utterances, as in the Italian example in Figure I, or across speakers. Parents commonly use partial repetition and expansion of child utterances, especially when children's speech is ungrammatical or incorrect [60,61]. Children make immediate corrections to their speech in response to parent expansions, often incorporating parents' corrections [60,62]. Variation sets can facilitate parsing even for elements that occur outside them [32]. In infants, variation sets might support learning structure at multiple levels, from detecting word boundaries to associating labels with objects to acquiring grammatical constructions.

Figure I. Samples of child-directed speech in six languages (all from CHILDES corpora [59]). As an example of the available structure, consider the Italian sample, which would seem to a novice learner to be a relatively short, initially undifferentiated sequence of sounds, *dovesonodovesonoiconiglietti*. The partial repetition across phrases (*dovesono* and *dovesonoiconiglietti*) facilitates comparison across the utterances produced by the parent. When aligned with itself, the partial overlap suggests two candidate units (*dovesono*, and *iconiglietti*), as well as a phrasal pattern (*dovesono [__]*). This happens to be as good an outcome as one could hope for: of the units, the first is a common collocation (''where_are''), the second is a noun phrase (''the bunnies''), and the pattern is the construction commonly used to ask, ''where are X?'' in Italian. The effectiveness of such cues has been demonstrated both across developmental time [25,80] and over shorter timescales [32].

computational conundra, and thereby solve the twin problems of pattern discovery and evaluation, by restricting their search for structure to a small time window. Its duration is constrained by multiple factors, such as the timescale of experience- and reward-based synaptic modification and properties of auditory and visual working memory [20]. In addition, we propose that the time window can be influenced by the dynamics of attentional focus as guided by social interaction. In interactions with caregivers, the structure to be learned is typically presented redundantly, which fits well the working memory constraints of a young learner. Consider the sample of Italian child-directed speech (CDS) in Box 1: it affords the identification of several candidate structures (because it can be aligned with itself in several distinct ways) and at the same time highlights their significance (because the probability that the partial self-matches in that string are due to chance is very small). Thus, the gist of our learning principle can be conveyed by the acronym ACCESS: Align Candidates, Compare, Evaluate Statistical/Social Significance.

ACCESS to structure: temporally constrained, socially embedded learning

According to ACCESS, infants learn by integrating, over a restricted time window, prominent statistical regularities with contextual cues such as social interaction and reward. *Statistical significance* is realized by recognizing patterns of co-occurrences that emerge above background noise. Restricting alignment and comparison to a small window amounts to a powerful test for significance: patterns that are prima facie rare but nevertheless recur within a short time of each other are likely to be meaningful [21,22]. *Behavioral significance* is achieved through contextual cues (e.g. social reinforcement or a food reward) which, if appearing within an appropriately short time window, add further independent support to the statistical evidence derived from the other data stream(s).

Sufficient statistical and behavioral support for a sequence should cause it to be segmented as a unit. The relation between these two types of cues to structure depends on the learning environment. For example, contextual cues can dominate others by causing a sequence to be learned as a unit after it is experienced once, but only under strongly reinforcing or aversive conditions. More generally, embedding learning in social interaction complements the effect of statistical regularities, which is based on item co-occurrence within the primary data stream (such as speech). Socially embedded learning weights such statistical regularities by their co-occurrence with supporting contextual cues.

However, the synergy between statistical and social cues is not limited to simple reinforcement. A much more powerful interactive learning mechanism [23] has the young learner and an adult engage in iterated turn-taking, in which the adult quickly and contingently responds to and expands upon the learner's immature contributions. In the development of complex communication systems such as language and birdsong, in which learning typically involves social interaction [23,24], conspicuous co-occurrence of patterns might indicate a functionally significant exchange of information (Figure 1). Thus the temporal structure of social interactions is crucial. For example, prelinguistic infants will learn to produce new phonological patterns from speech that is contingent on their babbling, but not from identical but non-contingent speech [24].

Social routines between parents and infants, in which each partner plays a predictable role in an interaction, might link statistical patterns at small and large timescales (Figure 2). The role of social interaction in making salient statistically significant structures is crucial to language learning. The possible importance of social routines (e.g. peekaboo) for language learning was identified by Bruner [25], who argued that small variations within predictable interactions could help infants detect patterns not only in language, but in others' social behaviors. Recent experimental studies of word learning showed that moment-to-moment changes in the structure of social interaction alter the likelihood of infants detecting patterns and forming associations between words and objects [26,27].

Although both statistical learning (including unsupervised discovery of patterns in data) and behaviorally motivated learning have a long history in psychology and in computer science, they have traditionally focused on somewhat distinct problems. For instance, distributional methods are best known for the discovery of recurring units and their classification [28]. Importantly, all existing unsupervised algorithms that are capable of working with raw, unannotated data and of scaling to large corpora while 'going recursive' and learning syntax-like rules (ADIOS [17]; U-DOP [19]; ConText [18]) focus exclusively on intrinsic data (transcribed speech or text alone; see Box 2). Moreover, they are all designed to operate in a 'batch mode' by considering the entire corpus of data simultaneously, rather than in the proper temporal order. In comparison, the incremental reinforcement learning algorithms [29] track the learner's long-term performance rather than contingent behavioral clues and do not deal with recursive structure. The ACCESS framework indicates how these families of computational approaches can be made more powerful (and perhaps converge) by making full use of the richness of time-locked multimodal data.

The ACCESS principles predict that infants should learn phonology, vocabulary and syntax most effectively when the relevant structures are highlighted in their caregivers' speech by occurring contingently on the infants' behavior. Indications of the ACCESS significance tests are found throughout early development, from the learning of speech sounds and syntax to the acquisition of complex conceptual structures and sequential skills. A ubiquitous aspect of parent–child interaction, CDS, fits this prediction. CDS uses both temporal proximity and social feedback to highlight significant structures. One of its characteristics is the prevalence of *variation sets* ([30]; Box 1) – utterances with partial repetitions that cluster in time (H. Waterfall, PhD thesis, University of Chicago, 2006; [18]).

The proportion of CDS utterances contained in variation sets is surprisingly constant across languages: 25% in English, 22% in Mandarin and 20% in Turkish. It grows

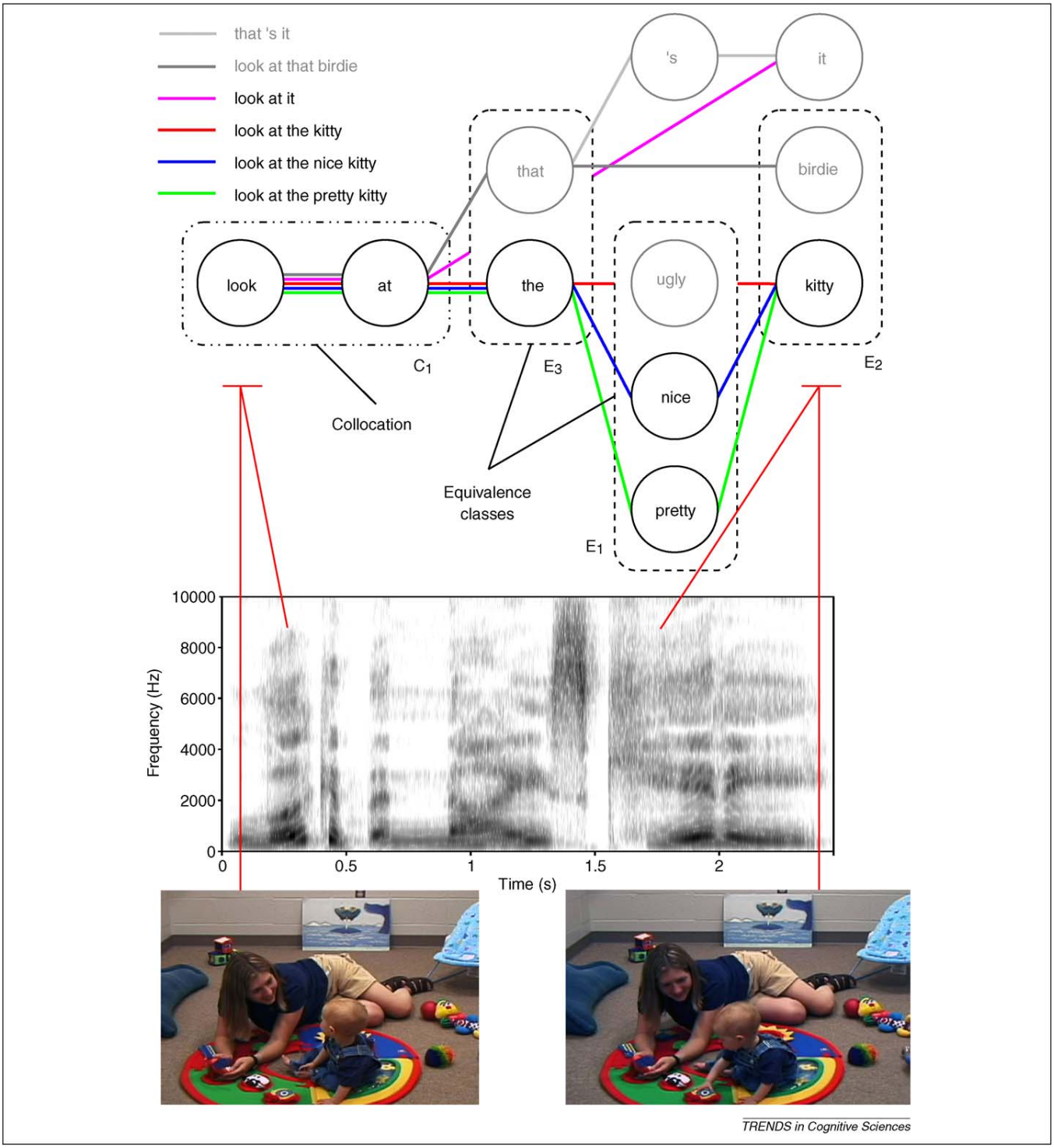

Figure 1. By integrating social and statistical learning, the ACCESS principles provide a framework for understanding how knowledge is acquired from social interaction. *Bottom, left:* A mother and her infant are engaged in a social interaction in which they are sharing attention to the same set of objects (i.e. joint attention). *Bottom, right:* The infant looks at the toy her mother is holding and babbles at it. Her mother labels the toy using a variation set (''Look at the nice kitty. What a nice kitty. What a pretty kitty. See the kitty? Do you see it?''). Variation sets commonly occur in caregivers' CDS. The multiple overlapping phrases in the variation set provide cues to grammatical structure (e.g. that ''the kitty'' and ''it'' can be substituted for each other; that ''nice'' and ''pretty'' can be substituted for each other). The sentences occur within a brief window of time, which facilitates their comparison. Because the mother labels the object contingently on an object-directed babble by the infant (which signals a state of focused attention [26]), the co-occurrence of speech and object acquires behavioral significance and is more likely to be learned. Learning can also be facilitated when one of the conversational partners holds the object [27]. *Middle*: Changes in prosody during the utterance ''Look at the nice kitty''. The exaggerated pitch changes on the word ''kitty'' increase its perceptual salience for infants [78]. The prosodic information co-occurs with mother and infant joint visual attention to the kitty. *Top:* Based on the interaction, the infant could extract a series of probable speech patterns (collocations and equivalences, also see Box 2). Detection of speech patterns is a result of sensitivity to statistically significant structure in auditory input. Pattern recognition could be facilitated when it co-occurs with supporting social information (e.g. the contingency of ''look at the nice kitty'' on the infant's vocalization, the shared visual attention to the kitty).

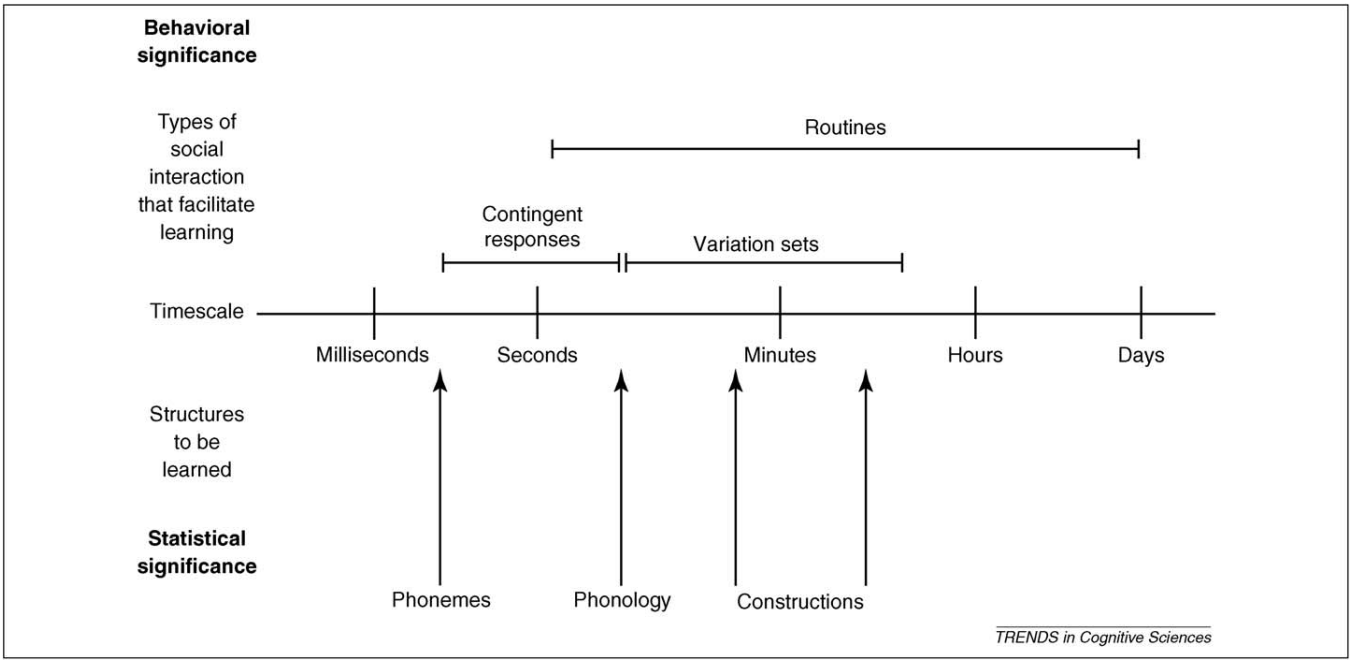

Figure 2. ACCESS uses two types of cues to gate the learning of structure. There exist patterns with temporal proximity at multiple timescales that are statistically significant, compared to a baseline of chance alignments. There also exist behaviorally significant patterns that are important given the learner's perceptual preferences or previous reward history. Infant learning is facilitated when they receive prompt and contingent feedback for their vocalizations. Variation sets also promote language development. A type of social interaction that includes both contingent responses to infant behavior and variation sets is routines (cf. Bruner's *formats* [25]). Caregiver–infant dyads often engage in routines, which are activities in which each partner's contributions are structured and predictable (e.g. peekaboo, tickle games, reading a picture book). For example, in tickling games, a caregiver might use exaggerated prosodic contours and gestures while saying ''*I'm gonna get your nose! I got it! I got your nose! I'm gonna get your belly! I got your belly!*'' Partial repetition in caregiver's speech facilitates comparison across utterances. The exaggerated speech and gestures provide salient information that reinforces correspondences between the varied items in the utterance (e.g. *nose, belly*) and their referents. Repeated pre-tickling gestures allow infants to predict the next step in the routine. The salient outcome of being tickled enables infants to confirm predictions about caregivers' speech and actions. Over developmental time, caregivers adjust their own role and the infant's role in the social routines, encouraging infants to increase their contribution to the game as infants' abilities improve (i.e. scaffolding). Together, the availability of statistically and behaviorally significant cues explains the robust effects of socially guided learning on the development of adaptive skills by facilitating the reliable identification of reliable patterns of information in the environment.
For example, infants' babbling facilitates phonological learning because caregivers' responses to early speech tend to be appropriately structured and temporally coordinated with child utterances [79], and infants learn new patterns of vocal production from caregiver speech that is contingent on infant production [23,24]. As this process is iterated across utterances, new phonological patterns are rapidly learned and entrenched; conversely, in prelinguistic vocal production, infants fail to learn new forms from non-contingent exposure to speech. Thus, infants show greatly increased capacity for learning statistical regularities when social cues work in conjunction with the statistics of the input. These findings imply that studies of production should complement studies of statistical learning in speech perception (e.g. [2]).

to about 50% if a gap of two utterances is allowed between those that partially match [18,30]. The alignability of consecutive or nearby utterances is an effective time-local statistical cue to structure, the significance of which can be estimated by comparing their string edit distance [31] to the cumulative average computed over the entire corpus [18,32]. As indicated by ACCESS, the social dimension of caregiver–child interaction adds another powerful cue: the alignment often spans both sides of the conversation and, even more importantly, constitutes an exchange that is initiated by the learner, for example (H. Waterfall, PhD thesis, University of Chicago, 2006):

child: Disappear.
mother: It disappeared.
child: Yes, yes it did disappear.

Cross-sentential cues facilitate the learning of syntax [33]. Recent studies showed that parents' use of nouns and verbs in variation sets in CDS is related to children's verb and noun use at the same observation, as well as to later child production of verbs, pronouns and subcategorization frames ([34]; H. Waterfall, PhD thesis, University of Chicago, 2006). Evidence for the causal role of variation sets that appear within a small time window emerges from artificial language learning experiments. Adults exposed to input containing variation sets performed better in word segmentation and phrase boundary judgment tasks than controls who heard the same utterances in a scrambled order, without variation sets [33].

The fit between the structure in the environment and the learner's capacity to perceive that structure is crucial for development. When developmental or perceptual problems lead to data with atypical distributions, or learners attend to inappropriate behavioral cues, anomalous structure can be learned (as, perhaps, in autism [22]; Box 3) (http://hdl.handle.net/1813/10178). Experimental work indicates that restructuring input can affect learning. For example, presenting words in variation sets that overlap in orthography can quickly improve the reading skills of poor readers [35]. Changes in behavioral cues to the structure of speech, such as manipulating reactions to infants' vocalizing in parent–infant vocal turn-taking, leads to rapid changes in vocal learning [23,24].

The ACCESS principles also apply to the task of learning structure in vision. Unlike speech, which unfolds over time, the units and relations that must be learned in vision

Box 2. Alignment in CDS: the ADIOS model

The procedure illustrated in Figure I is based on the ADIOS algorithm for language acquisition, which can detect patterns in corpora and generate new utterances that fit the learned patterns [17]. The data are presented in the form of a graph, with the utterances already aligned on significant morphemes (these too can be detected automatically; the algorithm is capable of bootstrapping from raw, unannotated speech, e.g. transcribed into a phonetic notation).

We believe that performance of corpus-based algorithms such as ADIOS can be improved (and its dependence on batch processing reduced) by using variation set structure within a restricted time window to boost the relevant statistics and by drawing on extralinguistic behavioral cues, as indicated by the ACCESS framework. Such an algorithm should be able to detect significant patterns even in a relatively small corpus and use those patterns to generate acceptable utterances. This approach to modeling language learning in temporally restricted, socially situated contexts stands in contrast to current efforts to model unsupervised learning of grammatical structure. For example, the reliance of U-DOP and ConText on batch processing of entire corpora means that those algorithms cannot easily incorporate temporal structure or cues indicating behavioral significance [18,19].

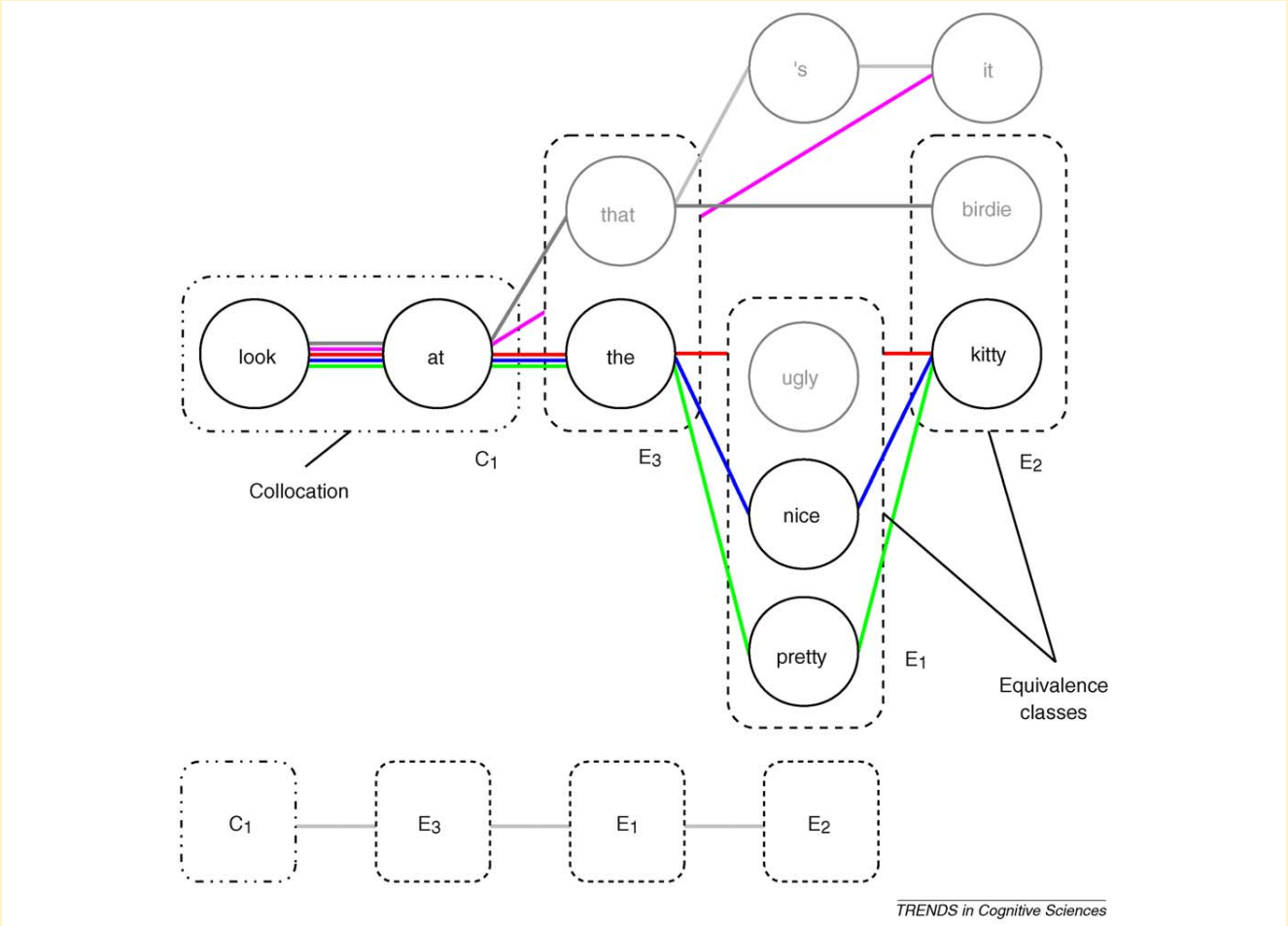

Figure I. The vertices of the graph are labeled by the morphemes (''words'') and the color-coded edges correspond to the utterances. For instance, the red path represents the utterance ''look at the kitty''. (TOP) Aligning the utterances reveals structures that the algorithm might deem statistically significant. Some of these are collocations; for instance, the significance of the hypothesis that ''look at the __ kitty'' constitutes a unit could be based on a test of binomial probabilities, which compares the number of paths that reach the final word (''kitty'') to the number of paths that leave the initial one (''look''). Other structures are equivalence classes; for instance, ''nice'', ''fluffy'', and ''__'' (empty slot) are substitutable in the context of the collocation ''look at the __ kitty''. It is very difficult to determine from the corpus data alone the extent to which such equivalence relations should be generalized. For example, although ''that'' and ''the'' are equivalent in some contexts, ''the'' cannot be substituted for ''that'' in the phrase ''that's it''. (BOTTOM) Collocations and equivalences that are deemed significant join the growing lexicon/grammar as new units, which can subsequently participate in the search for further structure. This recursive process gives rise to hierarchical, tree-structured representations and to constructions (limited-scope 'rules') that can be used productively [17]. Here, a sequence formed by one of the collocations and three equivalence classes becomes a partially lexicalized construction ''look at E3 E1 E2''. The performance of the learned grammar can be measured by its recall, or coverage (the percentage of sentences in a withheld test corpus that a parser derived from the grammar accepts) and precision, or productivity (the percentage of sentences it generates that are judged acceptable by human evaluators). On both these counts, the ADIOS algorithm achieved hitherto unprecedented performance [17].

are spatial and temporal. Behavioral studies yield evidence of *unitization* – a process whereby a class of visual patterns becomes an entity over which subsequent computations, including statistical learning, can be carried out [36]. Importantly, statistical learning of structural relationships in vision is subject to spatial constraints [37] that resemble the temporal constraints mentioned above in the context of learning language.

Neurally informed computational models can learn visual structure. For example, discovering would-be units in variation-set-like contexts can be carried out by Hebbian neurons [38]. A working computational model of systematic

Box 3. ACCESS and autism

Although the causes of autism are not yet clear, there is increasing agreement that it is characterized by a failure to attend to basic social cues, such as eye gaze, human voice and social motion, with presumed cascading consequences for both social and cognitive development [63,64]. Simulations and theoretical analysis indicate a link between these characteristics and an atypical application of the ACCESS principles of learning, in which perceptual defects lead to inappropriate data streams (and hence inappropriate statistical regularities), and social deficits lead to inappropriate behavioral cues [22]. Specifically, failure to use social cues skews the distribution of the incoming data, which affects the discovered regularities and results in: (i) a failure to detect or accept as significant patterns that most learners would view as common and (ii) accepting 'false' patterns as significant, as a result of experiencing them relatively too often. The first error could occur due to inattention to social contingencies that highlight patterns for typically-developing children. To understand the second type of error, consider the following three utterances:

don't touch the stove
don't spill the milk
don't touch the milk

If they all enter the alignment and comparison process with the appropriate frequency, they are likely to be segmented as 'don't', 'touch', 'spill', 'the stove', and 'the milk'. However, if for some reason a child pays attention only to the phrase 'don't touch the stove', he or she might learn it as a single undivided unit.

Such segmentation errors are common in autistic children (cf. echolalia [65,66]). Autistic children might frequently use an entire phrase rather than a single appropriate word when they see an object or a person related to this phrase. Often these phrases are lines in songs, or phrases taken from favorite television shows. This type of behavior is predicted by our approach. Autistic children are usually not very attentive to human speech [65,67]. This should increase the probability that a phrase to which they repeatedly pay attention will not have the opportunity to be aligned and compared to other phrases during the normal time window of the alignment and comparison process. The ACCESS principles indicate that appropriately controlling the richness of variation sets in learning methods used with autistic children could improve their learning [22]. Future models that test ACCESS principles could simulate learning given the processing deficits associated with autism. Input to the model could then be restructured to control the richness of variation sets to test whether learning improves with additional structure.

treatment of spatial configurations of basic units has also been reported [39]. Both these models implicitly rely on assumptions that ACCESS makes explicit (such as alignment and comparison). ACCESS also offers more effective ways to capitalize on those assumptions. For example, visual scene interpretation might be learned better by resorting to attentional selection to fixate the relevant part of the scene to begin with, paralleling the effectiveness of joint attention in language learning.

The ACCESS framework contributes to one of the central debates in cognitive sciences; the principles it proposes might help obviate the need for innate structures and be simple and general enough to evolve and become part of the brain's genetic script. Infants have available many sources of structure from which to learn, some of them surprising. Recent work on infants' object completion abilities demonstrates significant contributions of locomotor experience to cognitive development; independent, unsupported sitting and developmental changes in visuo-manual object exploration promote the formation of expectations about what unseen parts of objects should look like [40]. Thus the ACCESS framework offers an intriguing take on the 'poverty of the stimulus' idea: if the learner is sensitive to multiple converging streams of statistical and social cues, bringing them all to bear on the problem of structure discovery, then language learning becomes computationally easier, not more difficult (cf. [41]). The stream of experience is rich in salient, tractable structure; ACCESS posits that infants are capable of detecting and using that structure to build complex and adaptive skills.

Possible neurocomputational mechanisms behind ACCESS learning

Neural mechanisms of language are traditionally deemed neocortical [42,43]. By contrast, the more general task of sequential structure learning, studied in many animal models, is clearly associated with the hippocampus [44], an archicortical area that supports episodic memory [45,46]. Imaging studies show that medial temporal lobe areas, including the hippocampus, are involved in learning novel words [47] and hierarchically structured sensorimotor sequences [48], indicating that the hippocampus might play a key role in language acquisition. Indeed, infants diagnosed with bilateral hippocampal sclerosis do not acquire language, or lose it if the morbidity occurs at a young age [49]. Episodic memories tied to spatial information might also support language learning by linking objects and labels via locations in space. In one study, an experimenter gave young children an object in one location, removed the object, and then drew the children's attention to the (now empty) location while saying a word. After a short delay, the children associated the word with the object through their physical location, although the two never co-occurred [41]. These findings are consistent with the ACCESS principles, which hold that the learning of sequential structures involves alignment, statistical tracking, and eventual consolidation of cues that essentially are episodic memories.

In addition to areas in the medial temporal lobe (the hippocampus and the entorhinal cortex), in the frontal lobe, and in the thalamus, the system that mediates structured learning from episodic information includes, most significantly, the basal ganglia. As revealed by behavioral and neuropsychological studies, the basal ganglia in humans are involved in supporting learning and execution of various cognitive tasks that require flexible coordination of sequential structure processing and working memory [50], including language [51,52]. The basal ganglia circuits also handle the social-motivational aspects of complex learning (Syal, S. and Finlay, B. Motivating language learning: Thinking outside the cortex, unpublished manuscript). Although basal ganglia circuitry receives much attention from neuroscientists and computational modelers [29,53,54], its role in social cognitive computing, as called for by the ACCESS framework, is rarely mentioned.

Recent studies propose several specific mechanisms of neural plasticity that might underlie learning hierarchically structured sequence representations. First, serial

Box 4. Outstanding questions

- Do the ACCESS principles represent an evolutionarily early, conserved set of constraints on the acquisition of structure?
- How widespread are the ACCESS principles across animal species?
- How general are the ACCESS principles across cognition? Can they be linked to alignment-based techniques in visual object recognition [68]? Are they applicable to motor learning?
- Can ACCESS principles boost the learning of grammatical dependencies (e.g. agreement between noun and verb [69]) or visual dependencies (e.g. co-occurrences of specific objects in a scene [70])?
- Do the ACCESS principles of alignment and comparison apply over longer time intervals (hours or days) and deeper hierarchies (such as those formed by complex conceptual structures)? If yes, long-term memories must be retrieved and processed similarly to recent inputs, as might indeed be the case [71].
- Can a cognitively plausible computational model of language acquisition based on the ACCESS principles outperform traditional approaches to grammar inference [72]?
- Can ACCESS principles reduce the combinatorial explosion of hypotheses arising in the process of matching words to their referents? Typically, there are several potential word-to-world mappings, and the learner has to solve this mapping problem in the face of uncertainty. Although relying on statistical regularities alone can yield learning of referential regularities [73], preliminary experimental evidence in our labs indicates that variation sets and social cues can facilitate word-referent learning.
- Can ACCESS principles be used to extend explanations of language acquisition across cultures? Many theories of language development are based on experimental or observational studies of middle-class Western mother–child dyads. However, children learn language even when the amount of speech directed to them is relatively low [74,75]. Contingent interactions characterize caregiver–infant interactions in a wide range of cultures [76]. Infants can also learn while overhearing adult-directed speech, which also contains partial overlap across utterances [77], although less than in IDS. Adults' variation sets might facilitate language development in non-Western children; differences between adult- and infant-directed variation sets could partially explain differences in rate of acquisition.
- Can ACCESS principles be used to boost learning of second languages? For example, teaching materials could be arranged such that partial repetitions of to-be-learnt target constructions are presented in a social setting that promotes meaningful interaction. The combination of intrinsic variation and social motivation should make structure more salient to the learner, stimulating alignment and comparison of adjacent sentences, without overtaxing working memory capacity.
- In what ways does the differential weighting of various input streams change over developmental time? For example, variation sets are important to language learning from 14 to 30 months of age (H. Waterfall, PhD thesis, University of Chicago, 2006). Accurate predictions of learning in situations where behavioral and statistical cues interact should be the goal of future modeling efforts.

order might be coded by synfire chains: volleys of action potentials or spikes of activity propagating down a staggered set of cliques of neurons [55]. Second, the formation of such chains could be supported by spike timing-dependent plasticity (STDP), a form of experience-driven Hebbian learning [56]. Third, learning from experience can be made dependent on contingent reward and on social cues (as stipulated by the ACCESS framework) through diffuse dopaminergic modulation of the STDP circuits [57] and through appropriately tuned STDP-driven network dynamics [58].

Summary and future directions

We outlined a general framework for learning structure, ACCESS, the central tenet of which is that candidate structures drawn from a continuous stream of experience must pass two 'tests' to be learned. First, they must occur with statistical regularity, relative to a baseline of chance alignments, within a small time window. Second, they must be behaviorally significant, as indicated by external cues. Unlike statistical significance, which is formulated in terms of abstract information patterns, behavioral significance is embodied in interactive mechanisms of perception and action, and situated in the world. If structural elements pass both tests, they become likely to be learned. They can then be used recursively to discover further structure, resulting in hierarchical representations and developmental cascades of learning.

Although computer algorithms that apply the alignment, comparison and statistical testing principles described above exist and have yielded advances in generative grammar induction from large natural-language corpora [17,18], such algorithms currently do not deal with multimodal data, nor do they allow incremental learning (Box 2). Because these algorithms have proved especially effective when applied to transcripts of CDS [59], one promising avenue for their modification that we are pursuing is to make them utilize the rich information in CDS, such as variation sets (Box 4). We are also currently conducting behavioral experiments that test the effectiveness of variation sets for learning novel nouns and verbs.

The ACCESS principles are readily applicable to language and vision; we suspect that they will also apply to other domains. To identify their neural underpinnings, links need to be established between cortical processes such as synfire chains and STDP-mediated learning and subcortical structures such as basal ganglia. At the behavioral level, the interplay between statistical and social significance needs to be explored in both typical and atypical development. Such efforts will yield a more comprehensive and computationally grounded understanding of situated and embodied cognition.

Acknowledgements

MHG and JAS were supported by NSF BCS 0844015 and NICHD R03 HD061524-01. JYH was supported in part by NSF grants ITR-0325453, IIS-0534064, IIS-0812045 and IIS-0911036, by AFOSR grants FA9550-08-1-0438 and FA9550-09-1-0266, and ARO grant W911NF-09-1-0281. SE was supported in part by World Class University program at Korea University, funded by the National Research Foundation of Korea through the Ministry of Education, Science and Technology grant R31-2008-000-10008-0.

References

1 James, W. (1890) *The Principles of Psychology*, Holt
2 Saffran, J.R. *et al.* (1996) Statistical learning by 8-month-old-infants. *Science* 274, 1926–1928
3 Harris, Z.S. (1954) Distributional structure. *Word* 10, 146–162
4 Shepard, R.N. (1987) Toward a universal law of generalization for psychological science. *Science* 237, 1317–1323
5 Tenenbaum, J.B. and Griffiths, T.L. (2001) Generalization, similarity, and Bayesian inference. *Behav. Brain Sci.* 24, 629–641
6 Barsalou, L.W. (1999) Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660

7 Edelman, S. (2008) On the nature of minds, or: Truth and consequences. *J. Exp. Theor. Artif. Intell.* 20, 181–196
8 Lashley, K.S. (1951) The problem of serial order in behavior. In *Cerebral Mechanisms in Behavior* (Jeffress, L.A., ed.), pp. 112–146, Wiley
9 Graf Estes, K. *et al.* (2007) Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260
10 Cameron-Faulkner, T. *et al.* (2003) A construction based analysis of child directed speech. *Cogn. Sci.* 27, 843–873
11 Hume, D. (1739) *A Treatise of Human Nature*, Penguin Books
12 Harris, Z.S. (1991) *A theory of language and information*, Clarendon
13 Edelman, S. (2008) *Computing the mind: how the mind really works*, Oxford University Press
14 Fillmore, C.J. (1985) Syntactic intrusion and the notion of grammatical construction. *Berk. Ling. Soc.* 11, 73–86
15 Goldberg, A.E. (2006) *Constructions at work: The nature of generalization in language*, Oxford University Press
16 Christiansen, M.H. and Chater, N. (2001) Connectionist psycholinguistics: Capturing the empirical data. *Trends Cogn. Sci.* 5, 82–88
17 Solan, Z. *et al.* (2005) Unsupervised learning of natural languages. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11629–11634
18 Waterfall, H.R. et al. An empirical generative framework for computational modeling of language acquisition. *J. Child Lang.* (in press). doi:10.1017/S0305000910000024
19 Bod, R. (2009) From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cogn. Sci.* 33, 752–793
20 Kareev, Y. (1995) Through a narrow window: Working memory capacity and the detection of covariation. *Cognition* 56, 263–269
21 Edelman, S. and Waterfall, H.R. (2007) Behavioral and computational aspects of language and its acquisition. *Phys. Life Rev.* 4, 253–277
22 Lotem, A. and Halpern, J. (2008) A data-acquisition model for learning and cognitive development and its implications for autism. Computing and Information Science Technical Reports, Cornell University
23 Goldstein, M.H. *et al.* (2003) Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8030–8035
24 Goldstein, M.H. and Schwade, J.A. (2008) Social feedback to infants' babbling facilitates rapid phonological learning. *Psychol. Sci.* 19, 515–523
25 Bruner, J.S. (1983) *Child's talk: Learning to use language*, W.W. Norton
26 Goldstein, M.H. *et al.* (2010) Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy* DOI: 10.1111/j.1532-7078.2009.00020.x In: www.isisweb.org)
27 Pereira, A.F. *et al.* (2008) Social coordination in toddler's word learning: Interacting systems of perception and action. *Connect. Sci.* 20, 73–89
28 Fitch, S. and Chater, N. (1991) A hybrid approach to the automatic learning of linguistic categories. *AISB Qtrly.* 78, 16–24
29 Cohen, M.X. and Frank, M.J. (2009) Neurocomputational models of basal ganglia function in learning, memory and choice. *Behav. Brain Res.* 199, 141–156
30 Küntay, A. and Slobin, D. (1996) Listening to a Turkish mother: Some puzzles for acquisition. In *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp* (Slobin, D. *et al.*, eds), pp. 265–286, Lawrence Erlbaum Associates
31 Ristad, E.S. and Yianilos, P.N. (1998) Learning string-edit distance. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 522–532
32 Onnis, L. *et al.* (2008) Learn locally, act globally: Learning language from variation set cues. *Cognition* 109, 423–430
33 Morgan, J.L. *et al.* (1989) Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *J. Mem. Lang.* 28, 360–374
34 Hoff-Ginsberg, E. (1990) Maternal speech and the child's development of syntax: A further look. *J. Child Lang.* 17, 85–99
35 McCandliss, B. *et al.* (2003) Focusing attention on decoding for children with poor reading skills: Design and preliminary tests of the Word Building intervention. *Sci. Stud. Read.* 7, 75–104
36 Goldstone, R. (2000) Unitization during category learning. *J. Exp. Psychol. : Hum. Percept. Perform.* 26, 86–112
37 Conway, C.M. et al. (2007) Spatial constraints on visual statistical learning of multi-element scenes. In Proceedings of the 29th Annual Meeting of the Cognitive Science Society (McNamara, D.S. and Trafton, J.G., eds), pp. 185-190, Austin, TX: Cognitive Science Society
38 Edelman, S. et al. (2002) Unsupervised learning of visual structure. In Proc. 2nd Intl. Workshop on Biologically Motivated Comput. Vis. (Bülthoff, H.H. et al., eds), pp. 629-643, Springer
39 Edelman, S. and Intrator, N. (2003) Towards structural systematicity in distributed, statically bound visual representations. *Cogn. Sci.* 27, 73–109
40 Soska, K.C. *et al.* (2010) Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Dev. Psychol.* 46, 129–138
41 Smith, L. and Gasser, M. (2005) The development of embodied cognition: Six lessons from babies. *Artif. Life* 11, 13–29
42 Pulvermüller, F. (2002) A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. *Prog. Neurobiol.* 67, 85–111
43 Hurford, J.R. (2003) The neural basis of predicate-argument structure. *Behav. Brain Sci.* 26, 261–316
44 Fortin, N.J. *et al.* (2002) Critical role of the hippocampus in memory for sequences of events. *Nat. Neurosci.* 5, 458–462
45 Marr, D. (1971) Simple memory: A theory for archicortex. *Philos. Trans. R. Soc. Lond. B., Biol. Sci.* 262, 23–81
46 Eichenbaum, H. *et al.* (1999) The hippocampus, memory and place cells: is it spatial memory or memory space? *Neuron* 23, 209–226
47 Breitenstein, C. *et al.* (2005) Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage* 25, 958–968
48 Schendan, H.E. *et al.* (2003) An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron* 37, 1013–1025
49 Delong, G.R. and Heinz, E.R. (1997) The clinical syndrome of early-life bilateral hippocampal sclerosis. *Ann. Neurol.* 42, 11–17
50 Seger, C.A. (2006) The basal ganglia in human learning. *Neuroscientist* 12, 285–290
51 Lieberman, P. (2002) On the nature and evolution of the neural bases of human language. *Am. J. Phys. Anthropol.* 119, 36–62
52 Ullman, M.T. (2006) Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex* 42, 480–485
53 Dominey, P.F. and Hoen, M. (2006) Structure mapping and semantic integration in a construction-based neurolinguistic model of sentence processing. *Cortex* 42, 476–479
54 O'Reilly, R.C. and Frank, M.J. (2006) Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.* 18, 283–328
55 Ikegaya, Y. *et al.* (2004) Synfire chains and cortical songs: Temporal modules of cortical activity. *Science* 304, 559–564
56 Izhikevich, E.M. (2006) Polychronization: Computation with spikes. *Neural Comput.* 18, 245–282
57 Izhikevich, E.M. (2007) Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb. Cortex* 17, 2443–2452
58 Buonomano, D.V. and Maass, W. (2009) State-dependent computations: Spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125
59 MacWhinney, B. (2000) The CHILDES Project: Tools for Analyzing Talk (Vol. 1: Transcription format and programs; Vol. 2: The Database). Erlbaum
60 Chouinard, M.M. and Clark, E.V. (2003) Adult reformulations of child errors as negative evidence. *J. Child Lang.* 30, 637–669
61 Newport, E.L. *et al.* (1977) Mother, I'd rather do it myself: Some effects and noneffects of maternal speech style. In *Talking to children: Language input and acquisition* (Snow, C.E. and Ferguson, C.A., eds), pp. 109–150, Cambridge University Press
62 Saxton, M. (1997) The Contrast Theory of negative input. *J. Child Lang.* 24, 139–161
63 Klin, A. *et al.* (2003) The enactive mind, or from action to cognition: lessons from autism. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 358, 345–360
64 Klin, A. *et al.* (2009) Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* 459, 257–261
65 Frith, U. (1989) *Autism: Explaining the Enigma*, Blackwell
66 Howlin, P. (1998) *Children with Autism and Asperger Syndrome: A Guide for Practitioners and Carers*, John Wiley and Sons
67 Klin, A. (1991) Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *J. Autism Dev. Disord.* 21, 29–42

68 Ullman, S. (1989) Aligning pictorial descriptions: An approach to object recognition. *Cognition* 32, 193–254
69 Mel'čuk, I. (1988) *Dependency Syntax: Theory and practice*, SUNY Press
70 Oliva, A. and Torralba, A. (2007) The role of context in object recognition. *Trends Cogn. Sci.* 11, 520–527
71 Dudai, Y. (2009) Predicting not to predict too much: How the cellular machinery of memory anticipates the uncertain future. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364, 1255–1262
72 Klein, D. and Manning, C.D. (2004) Corpus-based induction of syntactic structure: Models of dependency and constituency. In Proc. 42nd Annual Meeting Assoc. Comput. Linguist, pp. 478–485
73 Smith, L. and Yu, C. (2008) Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568
74 Evans, G.W. *et al.* (1999) Parental language and verbal responsiveness to children in crowded homes. *Dev. Psychol.* 35, 1020–1023
75 Hart, B. and Risley, T. (1995) *Meaningful differences in the everyday experience of young American children*, Brookes
76 Keller, H. *et al.* (2008) The timing of verbal/vocal communications between mothers and their infants: A longitudinal cross-cultural comparison. *Infant Behav. Dev.* 31, 217–226
77 Pickering, M.J. and Garrod, S. (2004) Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–225
78 Fernald, A. (1992) Meaningful melodies in mothers' speech to infants. In *Nonverbal vocal communication: Comparative and developmental approaches* (Papousek, H. *et al.*, eds), Cambridge University Press
79 Goldstein, M.H. and West, M.J. (1999) Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *J. Comp. Psychol.* 113, 52–58
80 Tomasello, M. (2003) *Constructing a language: A usage-based theory of language acquisition*, Harvard University Press