The unrealized promise of infant statistical word–referent learning

Linda B. Smith, Sumarga H. Suanda, and Chen Yu

Psychological and Brain Sciences, Program in Cognitive Science, Indiana University, Bloomington, IN 47405, USA

Recent theory and experiments offer a new solution regarding how infant learners may break into word learning by using cross-situational statistics to find the underlying word–referent mappings. Computational models demonstrate the in-principle plausibility of this statistical learning solution and experimental evidence shows that infants can aggregate and make statistically appropriate decisions from word–referent co-occurrence data. We review these contributions and then identify the gaps in current knowledge that prevent a confident conclusion about whether cross-situational learning is the mechanism through which infants break into word learning. We propose an agenda to address that gap that focuses on detailing the statistics in the learning environment and the cognitive processes that make use of those statistics.

Introduction

The world offers data to novice word learners in the form of word–object co-occurrences. These data may be noisy with many spurious co-occurrences (Figure 1). Thus a core problem for theories of early word learning is determining how infants manage to find the right word–referent pairs in the noise. The evidence indicates that by their first birthday, if not before, infants have already found a considerable number of these correspondences [1–3]. Older word learners, 2-year-olds, employ knowledge about social cues, language, and categories to map words to referents; however, this knowledge develops over the course of word learning and may be partly a product of word learning itself [4–8]. Thus the field lacks an understanding of how early word learning starts. Recent theory and experiments offer a new solution: novice learners may break into word learning through the noisy co-occurrence data, using cross-situational statistics to find the underlying word–referent mappings. We begin with a review of the models that show the in-principle plausibility of this solution, followed by the experimental evidence showing that infants aggregate and make statistically appropriate decisions from co-occurrence data. We then turn to the critical question: could this solution work for infants in the real world? The answer depends on a better understanding than we currently have of the relevant statistics in the learning environment and the cognitive processes that make use of those statistics.

Finding structure in co-occurrences

The classic debate in the study of early word learning pits hypothesis testing against associative learning [6,9]. Recent computational advances have blurred the distinction; models from both frameworks can operate over the same co-occurrence data to yield the same learning patterns (Box 1). Accordingly, we ignore this classic divide to focus on four new contributions.

There is discoverable structure in word–scene co-occurrence data

Several researchers have applied statistical learning algorithms to word–scene co-occurrence data taken from audio and video recordings of infants in common everyday activities (i.e., an infant interacting with a parent). The algorithms, expressed as either Bayesian inference models [10] or machine translation models [11–13], readily succeed in discovering the underlying word–referent pairings from real-world co-occurrences. Although these powerful models may not be psychologically realistic [9], they show that there is significant structure in the co-occurrence data such that the co-occurrences –along with statistical learning mechanisms – might be enough for infants to discover the mappings of words to objects.

Statistical learning is about learning a system of co-occurrences

Statistical learning models succeed because they operate on the set of co-occurrence data with the goal of simultaneously learning multiple words and referents. Within these models, the strength of an individual word–referent association (or the probability of a hypothesis) is not strictly a property of that word–referent pair alone. Instead, it interacts with and is embedded in the other regularities in the co-occurrence matrix, regularities that enable the learning machinery to exploit correlations [14–16], coherent covariation [15,17], and the structure in the whole matrix [18–20] to discover the underlying correspondences.

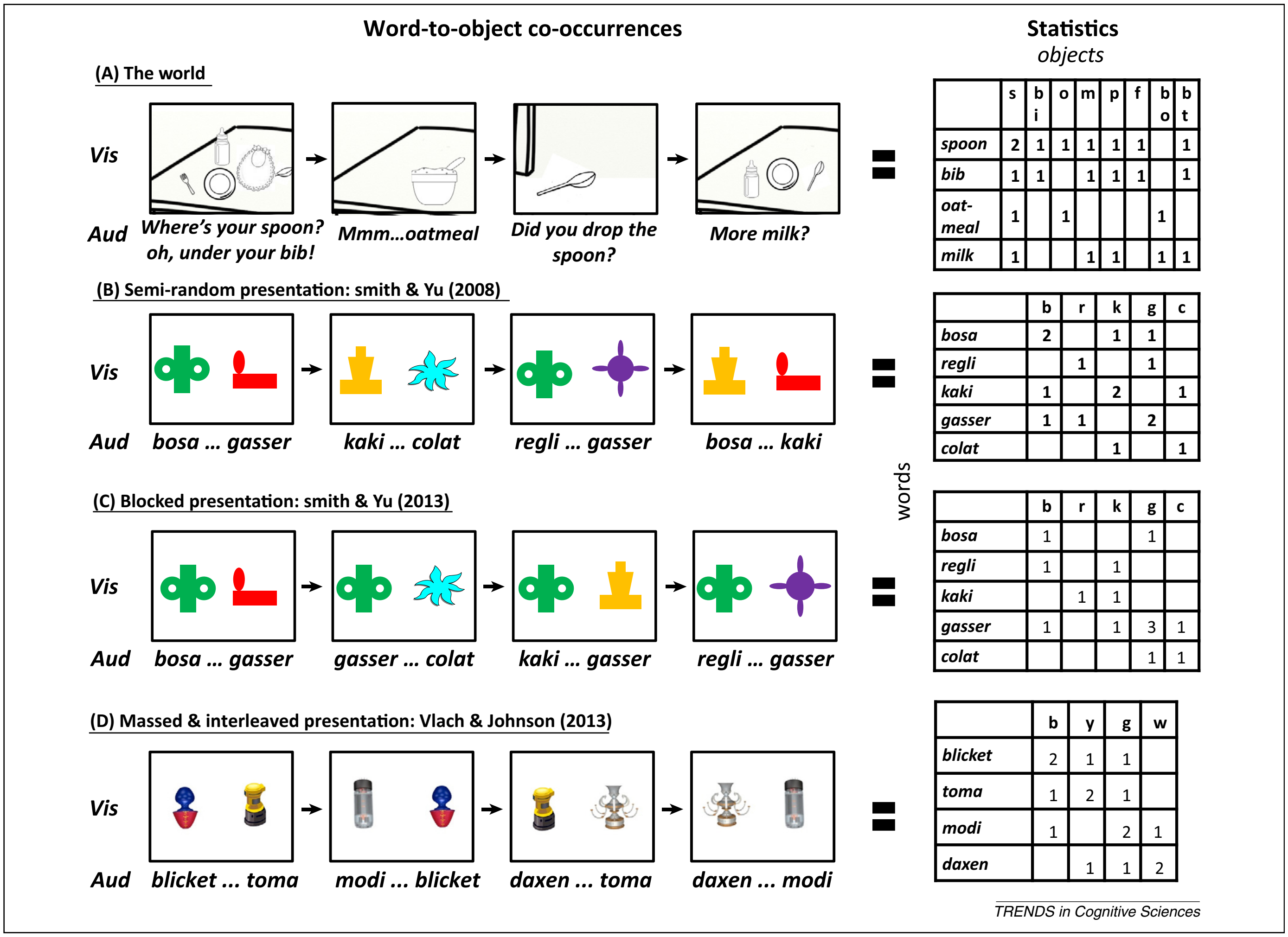
Word–referent pairs compete

One mechanism through which cross-item dependencies can influence learning is competition. Most current statistical models of cross-situational learning optimize solutions in which each unique word is linked to just one referential

Corresponding author: Smith, L.B. (smith4@indiana.edu).

1364-6613/$ – see front matter

http://dx.doi.org/10.1016/j.tics.2014.02.007



	s	b i	o	m	p	f	b o	b t
spoon	2	1	1	1	1	1		1
bib	1	1		1	1	1		1
oat-meal	1		1				1	
milk	1			1	1		1	1

	b	r	k	g	c
bosa	2		1	1	
regli		1		1	
kaki	1		2		1
gasser	1	1		2	
colat			1		1

	b	r	k	g	c
bosa	1			1	
regli	1		1		
kaki		1	1		
gasser	1		1	3	1
colat				1	1

	b	y	g	w
blicket	2	1	1	
toma	1	2	1	
modi	1		2	1
daxen		1	1	2

Figure 1. Series of scenes and words and the co-occurrence matrix. **(A)** Example scenes and co-occurring language as experienced by an infant. **(B)** A sample of the scenes and co-occurring words in Smith and Yu [26]. Within a trial there was no information regarding which object was the referent of each word. Across the 30 trials, each correct word–referent pair occurred ten times and each spurious correspondence only twice. Test trials (not shown) consisted of the presentation of two objects but one word. Looking duration to the target referent was the dependent measure. **(C)** The 'novelty trap' set by Smith and Yu [29] presented blocks of trials in which one object was repeated at the same location within a block and the other location showed non-repeating objects. The final statistics were the same as in Smith and Yu [26] and testing was the same. **(D)** The interleaved trials in the Vlach and Johnson study [34], arranged so that some word–referent pairs were adjacent and others were distant in the series. Final statistics and testing were comparable with the Smith and Yu studies.

category. Sometimes known as the mutual-exclusivity principle, strong word–referent mappings are proposed to block or inhibit other mappings that contain overlapping components [9,21]. Such competition may also be responsible for the disambiguation (also known as 'fast-mapping') phenomenon in young children (Figure 2). However, within the models competition may be implicit and operate on partial (not behaviorally evident) knowledge [21]. For infant learners, this competition could mean that strong evidence for some mappings effectively 'cleans up' the data, helping the learning of weaker and noisier (non-competing) contingencies.

Statistical learning models learn to learn

Infants become better word learners as they learn more words. This is evident in the vocabulary spurt [22], an increase in the rate of receptive and productive vocabulary that typically emerges as infants approach their second birthday, as well as in several other emergent phenomena that are predicted by, and predictive of, infants' vocabularies (Figure 2). Numerous models, including both associationist and probabilistic inference, have shown how these properties of early word learning may be driven by the statistics of word–referent co-occurrences [14–16,18,19,22–25]. Thus, statistical learning – and the structure in word–referent co-occurrence data – might not just start infant word–referent learning but might also build the knowledge-based reduction of uncertainty that is evident in older word learners.

In sum, the extant models serve as compelling demonstration proofs. They show that there is structure in the co-occurrence data. They shift the empirical question from how infants learn individual word–referent pairs to how infants operate on the statistical regularities within a system of words and referents to learn multiple words simultaneously. They highlight the potential importance of mutual exclusivity and competition among word–referent pairings as critical to the process and they yield developmental patterns of learning that are consistent with patterns observed in infant learners. With a few

Box 1. Hypothesis testing and associative learning

A the computational level, models of statistical word–referent learning may be characterized and formulated in terms of three components, as in Figure IA: assumptions about the co-occurrence data on which the learning system operates, the statistical computations that are performed, and the learning outcomes that are achieved. The proposed statistical computations are generally seen as the main claim being made by the model. Within associative theories, these computations emerge from the strengthening and weakening of associations as a function of co-occurrence strength and competition among associations. Within hypothesis-testing theories, conceptually coherent hypotheses are confirmed or disconfirmed through various statistical procedures. These two frameworks thus offer fundamentally different characterizations of what it means to be a statistical learner. However, to simulate performance in a task, models of both classes must (implicitly or explicitly) specify several separable processing components, as illustrated in Figure IB. First, they must specify the information selected within a learning moment. The information selected within a learning trial could be narrow (one word and one referent per learning moment) or broad (many co-occurring word–referent pairs) or even change with learning (beginning broad and then becoming narrower as more is known). Many or few hypothesized or association pairs may be stored in memory and thus tracked. As part of the statistical computation, models must specify how the tracked information is aggregated and represented, including, for example, whether that stored information (knowledge) is represented in an all-or-none or probabilistic fashion. Finally, to simulate the learning outcome, the model must also specify the decision processes at test, including how information is retrieved. For example, the participant may make all-or-none decisions from graded statistics or make graded responses from the same data, and these decision processes, from the same represented knowledge, could vary with testing context. In a series of simulations, Yu and Smith [9] showed that these component decisions interact in complex ways within both classes of associative-learning and hypothesis-testing models. Indeed, very simple associative models could mimic sophisticated hypothesis-testing models, producing the same learning patterns although with different internal components. In brief, the two classes of model cannot be discriminated when formulated at the computational level in Figure IA but only by direct assessments of the proposed learning mechanism in the context of explicitly specified supporting processing components, as shown in Figure IB, and all of them should be informed by empirical evidence on infant attention and memory systems, as well as decision processes.

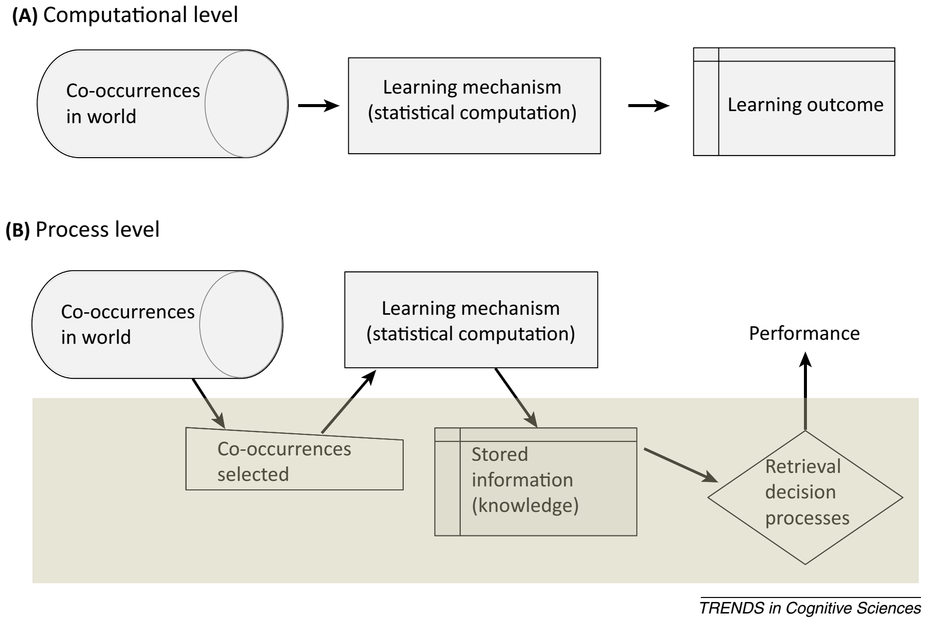
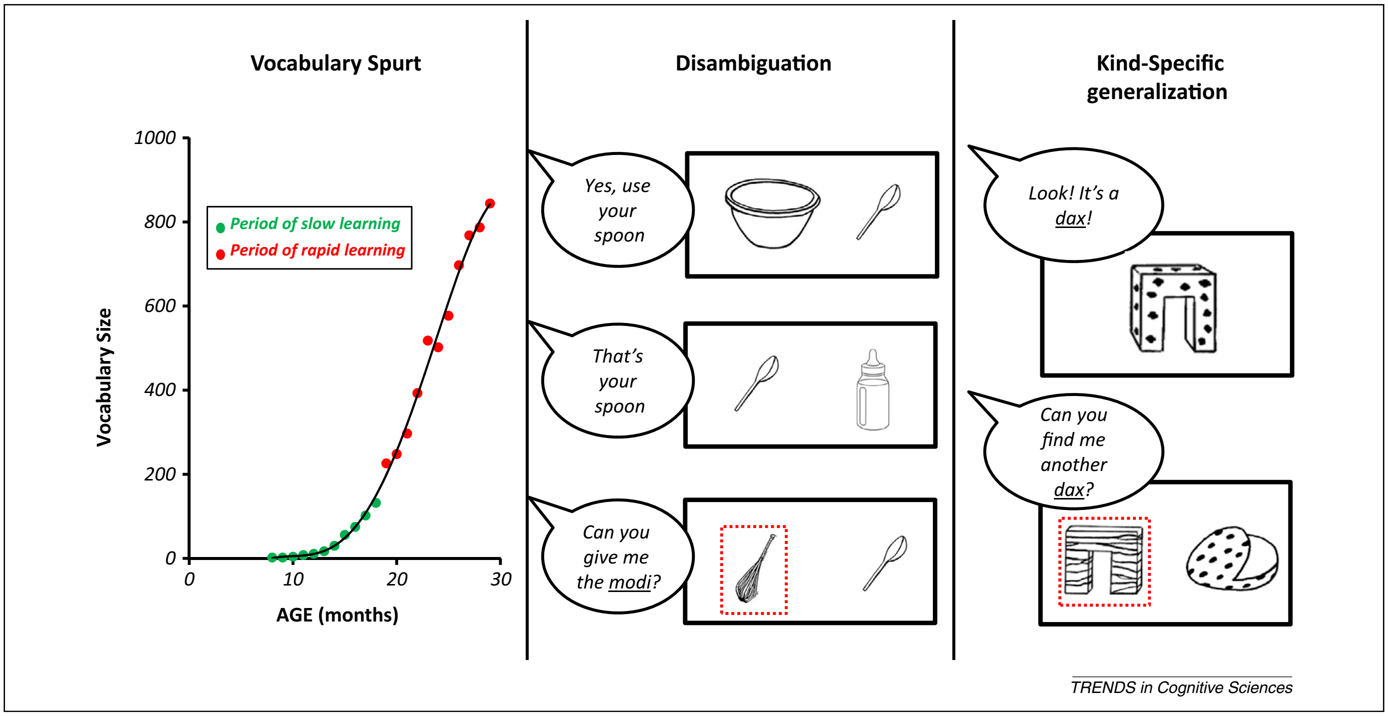

(A) Computational level
Co-occurrences in world
Learning mechanism (statistical computation)
Learning outcome
(B) Process level
Co-occurrences in world
Learning mechanism (statistical computation)
Performance
Co-occurrences selected
Stored information (knowledge)
Retrieval decision processes
TRENDS in Cognitive Sciences

Figure I. Many current models of cross-situational learning are formulated at a high conceptual level with just three components, as shown in **(A)**: the data (assumed or measured co-occurrences in the world); the learning mechanism (a set of statistical computations performed on the data); and the demonstrated learning outcome. However, to simulate the performance of human learners, the model must – implicitly or explicitly – make decisions about other components of the learning system. Because these components interact in complex ways, these decisions can yield multiple but very different paths to the same learning outcomes or performance.

exceptions they do not attempt to model the real-time psychological processes – attention, memory – essential to actually learning words (9,15).

Infants aggregate co-occurrence data

The cross-situational word-learning task was invented to answer the question of whether infants can do what the models propose: learn multiple word–referent pairs from noisy co-occurrence data. On each trial in the task (Figure 1B), infants hear two words and see two objects with no information about which word goes with which object. However, across trials each individual word always co-occurs with just one referent; thus there is across-trial certainty despite within-trial uncertainty, if learners aggregate the co-occurrence data across trials. Smith and Yu [26] (replicated [27]) presented 12- and 14-month-old infants with a randomly ordered stream of 30 such trials with six novel words and six novel objects. At the end of this experience, the infants' word learning was measured using a two-alternative preferential-looking procedure: two visual objects were presented in the context of one spoken word and looking time to the statistically correct referent of that word was measured. The results showed that the infants looked more to the correct referent than the foil on hearing the associated name. Moreover, analyses of individual word pairs and individual infants suggested that most infants learned four or more word pairs. To do this, infants must have attended to, stored, and in some way



TRENDS in Cognitive Sciences

Figure 2. Developmental changes in word learning. Children show knowledge of words and referents early, in both receptive and productive knowledge, with new words added slowly at first and then more rapidly just before the second birthday [1–4,14,22]. The period of rapid learning is also associated with increasingly sophisticated novel word-learning strategies measured in the laboratory, including referential disambiguation [7,14] and kind-specific generalizations [5,15,19]. In the disambiguation task, children are reminded of the names of objects that they know and then a novel object is presented with a novel name; children consistently interpret the novel name as referring to the novel object. This disambiguation or fast mapping of a novel name to an object is sometimes considered an example of the mutual-exclusivity principle of one name per object [9,21]. The shape bias is an example of a kind-specific generalization. When children are told the name of a novel artifact, they generalize that name to other objects by similarity in shape; for substance-like novel objects, however, they systematically generalize the name by material [15].

statistically evaluated the system of word–referent co-occurrences. In a subsequent study, Vouloumanous and Werker [28] showed that 18-month-old infants tracked both high- and low-frequency contingencies within a data set of word-to-object co-occurrences, a result consistent with the notion that infants track a system of regularities, not just the most dominant ones.

Two additional studies revealed that infant statistical learning is constrained by developing attention and memory processes. Smith and Yu [29] disrupted word–referent learning in 12–14-month-old infants by setting a novelty trap such that within each learning trial one potential referent was more salient than the other. They did this by ordering the training trials (Figure 1C) so that one object (and its location) was repeated within blocks of five trials whereas the other location showed a new object on each trial within that block of trials. Across the 30 trials, each object served equally often as the constant object and as one of the non-repeating objects, so that the final co-occurrence statistics were the same as in Yu and Smith's initial [26] study. This manipulation caused about half of the infants to attend visually to only the non-repeating object within a trial; these infants failed to learn the underlying word–referent correspondences. By contrast, the infants who more successfully negotiated the novelty trap so as to sample both potential referents within a trial learned the underlying structure of word–referent mappings. Past research on visual attention indicates that infant attention is often too stimulus driven and too sticky, such that infants may not fully sample the available information in an array [30,31]. Thus, the immature attentional system could pose a serious limit on infant statistical learning because the data available to any statistical learning machinery are not the data in the real world, but only the subset of that data that makes contact with the infant's learning system. Past research has also shown considerable individual differences among infants in the development of visual processing and specifically that more flexible visual attention (less sticky, less stimulus driven) is a strong predictor of later cognitive and vocabulary development [32]. Thus, the development of attentional processes that support broad sampling of the co-occurrence data could be a rate-limiting factor in early vocabulary development (see also [33]).

In a related study, Vlach and Johnson [34] reordered the sequence of learning trials to test the role of memory in statistical word–referent learning. They arranged the series of trials (Figure 1D) so that some word–referent pairs were repeated on successive trials and others were repeated only after many intervening trials. They found that 20-month-old infants learned both classes of pairs, indicating that they could combine information over relatively many interleaved items. By contrast, 16-month-old infants learned only the successively repeated pairs, raising the question of whether the memory systems of young infants are sufficient to aggregate information across long delays between encounters of a word and a referent. Critically, the statistics in all of these experiments are likely to be different from those that characterize infant learning environments. Word–referent pairings in real-world experiences are likely to have a bursty structure [35], with a word–referent pair repeating multiple times in a conversational context followed by periods when it is rarely encountered and then by another burst of repetitions in another coherent conversation. How might this dynamic structure – and the cumulative repetitions over days and weeks – interact with the aggregation of information over time and with the statistical evaluation of that information? Evidence on the dynamic properties of co-occurrences in the world and evidence on how these properties engage the attentional and memory systems of infants is critically needed to determine the scalability of cross-situational word learning (Box 2).

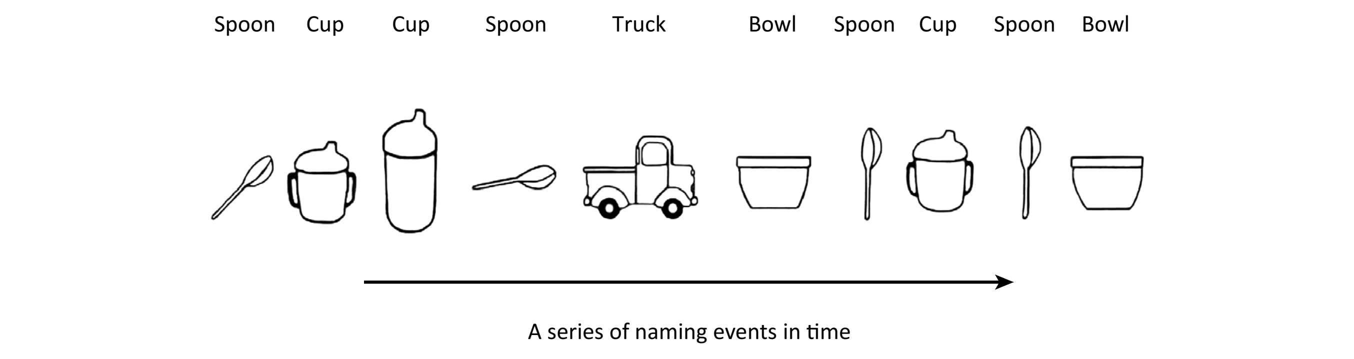
Statistics in the mechanisms and in the world

The models and the infant experiments suggest cross-situational word learning as a potential mechanism for early word learning. However, the models provide computational rather than process accounts of learning. Furthermore, the models do not specifically model infant cross-situational learning data. The infant cross-situational learning task presents streams of word–object co-occurrences that are considerably simpler than those in the world and even so success is limited in some presentation arrangements for learners by their developing attentional and memory processes. Thus, the key question remains open: could cross-situational word–referent learning work in the real world given the statistical regularities in infants' learning environments and given the seemingly limited cognitive skills of infants at the young ages at which initial word–referent learning occurs?

One ongoing discussion in the literature centers on the question of just how noisy the data are. As a counterweight to the perhaps too-powerful statistical learning algorithms, Trueswell, Gleitman, and colleagues [36,37] attempted to measure the uncertainty of word–referent co-occurrences in infant directed speech via what they call the 'Human Simulation Paradigm' (HSP). Instead of putting word–referent co-occurrences into a model, they presented adults with brief video clips of a parent talking to an infant. The audio was removed and a sound was inserted where the target word had been. The adults' task on each trial was explicitly to offer a hypothesis regarding the intended referent by the parent at the moment of the cued sound. Adults were very poor at this and showed no ability to aggregate information about word–referent correspondences across trials. The researchers concluded that the co-occurrences in the real world are much too noisy to be effectively mined by human learners. However, other investigators using variants of the HSP method have come to different conclusions. One study [38] found that about 50% of the naming episodes by mothers to toddlers were not ambiguous to the adults, who could readily guess the target referent. Another study [39] showed that the

Box 2. Statistics in time

Recent studies of infant and toddler retention of an encountered word and referent suggest that learning is incremental and slow. For example, toddlers who map the novel word to the novel object in the fast-mapping paradigm in Figure 2 do not retain that mapping when tested after delays as short as 5 min, although they show savings in later learning [14]. Learning that is incremental and aggregates over time requires that the learning system recognize when an item is a repetition of a previously experienced instance. For simplicity, consider the case, illustrated in Figure I, in which there is no within-situation ambiguity: each moment in time presents the learner with one object and one word. How does the learning system know to aggregate over instances 1 and 4 – that this is a 'repetition'? Research on human memory [51] indicates three stimulus dimensions relevant to the likelihood with which previous memories are activated and combined with current input: similarity, time, and context. For example, if the same person says 'cup' across two instances (rather than if a male speaker names one and a female speaker names another) or if the two cups are identical rather than perceptibly different, the second naming event is more likely to activate and thus strengthen the memory of the first naming event. Likewise, a second instance that follows the first close in time is more likely than one that follows after a delay to activate and be combined with the memory of the first. Finally, human memory systems are highly dependent on contextual cues for activating memories. Thus, naming of a spoon in the context of a cereal bowl on a highchair tray is more likely to reactivate prior memories of spoon-naming events that also occurred in that context than would be the naming of a spoon that is in the flatbed of a toy truck [52]. Similarity, time, and context have all been demonstrated to play a role in toddler memory and retention of object names [53,54], with delays (and partial forgetting) playing a positive role in building more abstract memories and generalizable knowledge [55], but, with the exception of Vlach and Johnson [34], have not been studied in the context of cross-situational word learning in the laboratory. The structure of the statistics of natural learning experiences along these dimensions also has not been studied.



TRENDS in Cognitive Sciences

Figure I. A series of naming events and referents in time.

degree of word–referent ambiguity as measured by the HSP varied considerably across parents and that this ambiguity was predictive of later word learning, such that toddlers who experienced less ambiguity went on to develop larger vocabularies. Finally, in an ideal observer analysis of word–referent co-occurrences in parents' interactions with their 6–18-month-old infants, Frank and colleagues [35] found that parents mostly create unambiguous naming moments. If much of the early data are relatively clean, infants could simply ignore moments that are too uncertain [40] and use relatively simple statistical mechanisms to determine the word–referent pairs from this cleaner sample of co-occurrence data.

To date, the methods for measuring the structure in the learning environment have taken word–referent co-occurrences from the world (videos of parent naming events) and submitted these co-occurrences to powerful algorithms or adult coders, analyzing the problem at the computational level. There are two limits to this approach. First, the data relevant to a statistical learner are not the data in the world but the data in the world filtered through the learners' sensory, attentional, and memory systems (Box 3). Second, these filters – sensory, attentional, and memory systems – are themselves statistical learners (e.g., [41–43]) that may not simply let information through to the learner but may also weight that information in ways directly linked to statistical computation and to the covariation in the data. If these ideas are correct, we will need to change the way we conceptualize the scope of statistical word–referent learning and the way we measure the statistical regularities in the learning environment, and we need to study how statistics are filtered and aggregated in multiple mechanisms from attention, to memory, to decision processes.

We may also need to ask what is the signal and what is the noise in the co-occurrence data. Consider again the stream of learning experiences illustrated in Figure 1A; in terms of the depicted co-occurrence matrix, these are very noisy data with many 'spurious' correspondences. However, in another sense these are not spurious correspondences but examples of the coherent covariation that characterizes learning environments; for infants, spoons (and their name) may be typically experienced in contexts of oatmeal and sippy cups (and their names). What is already known about visual statistical learning [42,42], about cued attention [44,45], about the priming of memories [46], and about the statistical structure of language [8] is that this structure may actively help learners find the right correspondences. For example, the word 'bowl' (or the sight of a bowl) might predict the likely presence of spoons

Box 3. Three lessons from the infant's view of word learning

Developmental scientists have begun to take an 'infant's perspective' on word learning by placing mini-lightweight head cameras [56–60] or head-mounted eye trackers [61–63] on infants as they engage in various tasks (Figure I). These new methods have documented unique properties of infant views. The growing body of work employing this technique has led to significant advancements in our understanding early natural-vision [58], motor [61,63], social [62], object-processing [64], and language [10,59,60] development.

For scholars of early word learning, the existing findings from infants' views have offered at least three important lessons. First, an analysis of infants' views raises the possible need to reconsider fundamental assumptions about the problem of early word learning. That is, multiple studies have demonstrated that the infants' views of events are significantly less cluttered than adults' views of the same events [1,2,4]. This insight is significant considering the assumption that there is a high degree of referential uncertainty caused by cluttered everyday environments [11]. Second, an analysis of infants' views during naming events highlights the potential role of various visual properties in early word learning. In a series of studies in which parents played with and talked about novel objects with their toddlers, Yu, Smith, and their colleagues observed that the visual properties of the target object during naming events (e.g., its image size relative to other objects, its centeredness in the visual field) predicted toddlers' novel object-name learning [4,5], suggesting a possible larger role for bottom-up cues to word learning than previously suggested. This perceptual information may be directly linked to internal statistical computations. Finally, the information available in infants' views during naming events may be more conducive to and induce particular learning mechanisms. In a recent study, Yurovsky and colleagues [38] found that when naming events were viewed from an infant perspective (as opposed to a third-person perspective), adult word learners were more likely to engage in an aggregative, statistical word-learning process. Together, these findings illustrate the need to take the infant's views into account – statistical information that makes contact with the infant's learning system – when developing mechanistic theories of early word learning.

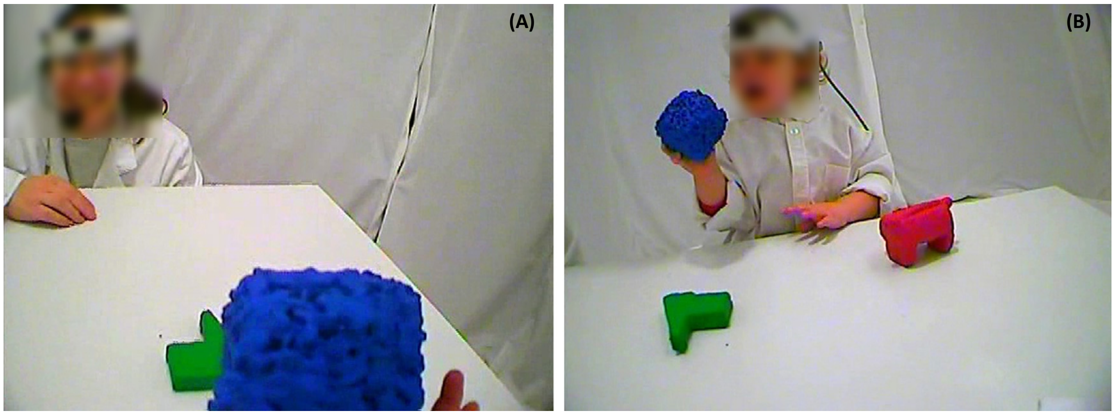


Figure I. Image frames from an infant-view head camera and from a parent-view head camera as the parent and infant engage in joint toy play.

Box 4. Outstanding questions

- What are the linguistic regularities in early naming events: isolated words, frequent frames, co-occurring object names?
- What are the visual regularities in early naming events: visually isolated objects, saliency properties, co-occurring objects and contexts?
- What are the regularities across words and objects and visual contexts?
- What are the dynamic properties of repeated naming events: within the seconds and minutes of working-memory processes, within the hours, days, and weeks of infant learning experiences?
- Can these linguistic and visual regularities in the real world be represented in the framework of cross-situational learning?

and enable the learner to find that referent in the visual clutter; the word 'bowl' (or the sight of a bowl) may prime and activate memories of spoons and their names, thereby supporting the aggregation of information over time and contexts.

Theories of infant word–referent learning treat the co-occurrences between to-be-learned words and referents as the signal and all else as noise. However, from another perspective the 'noise' contains information: regularities that interact with the sensory, attentional, and memory processes on which cross-situational learning depends. It may be through the interactions of multiple statistically sensitive processes that novice learners simultaneously solve multiple and mutually constraining tasks of mapping words to referents while building semantic networks [47–50]. This proposal sets a possible agenda for future research (Box 4).

Acknowledgments

This work was supported by NIH R01HD056029 to C.Y. and L.B.S. and NSF SMA-1203420 to S.H.S.

References

1 Bergelson, E. and Swingley, D. (2012) At 6-9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3253–3258

2 Tincoff, R. and Jusczyk, P.W. (2012) Six-month-olds comprehend words that refer to parts of the body. *Infancy* 17, 432–444

3 Mani, N. and Plunkett, K. (2010) Twelve-month-olds know their cups from their keps and tups. *Infancy* 15, 445–470

4 Tomasello, M. and Tomasello, M. (2009) *Constructing a Language: A Usage-based Theory of Language Acquisition*, Harvard University Press

5 Smith, L.B. *et al.* (2010) Knowledge as process: contextually cued attention and early word learning. *Cogn. Sci.* 34, 1287–1314

6 Waxman, S.R. and Gelman, S.A. (2009) Early word-learning entails reference, not merely associations. *Trends Cogn. Sci.* 13, 258–263

7 Bion, R.A. *et al.* (2012) Fast mapping, slow learning: disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition* 126, 39–53

8 Romberg, A.R. and Saffran, J.R. (2010) Statistical learning and language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 906–914

9 Yu, C. and Smith, L.B. (2012) Modeling cross-situational word–referent learning: prior questions. *Psychol. Rev.* 119, 21–39

10 Frank, M.C. *et al.* (2009) Using speakers' referential intentions to model early cross-situational word learning. *Psychol. Sci.* 20, 579–585

11 Yu, C. *et al.* (2005) The role of embodied intention in early lexical acquisition. *Cogn. Sci.* 29, 961–1005

12 Yu, C. and Ballard, D.H. (2007) A unified model of early word learning: integrating statistical and social cues. *Neurocomputing* 70, 2149–2165

13 Yu, C. (2008) A statistical associative account of vocabulary growth in early word learning. *Lang. Learn. Acquis.* 4, 32–62

14 McMurray, B. *et al.* (2012) Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychol. Rev.* 119, 831–877

15 Colunga, E. and Smith, L.B. (2005) From the lexicon to expectations about kinds: a role for associative learning. *Psychol. Rev.* 112, 347

16 Li, P. *et al.* (2007) Dynamic self-organization and early lexical development in children. *Cogn. Sci.* 31, 581–612

17 McClelland, J.L. *et al.* (2010) Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* 14, 348–356

18 Kachergis, G. *et al.* (2012) An associative model of adaptive inference for learning word–referent mappings. *Psychon. Bull. Rev.* 19, 317–324

19 Xu, F. and Tenenbaum, J.B. (2007) Word learning as Bayesian inference. *Psychol. Rev.* 114, 245

20 Tenenbaum, J.B. *et al.* (2011) How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285

21 Yurovsky, D. *et al.* (2013) Competitive processes in cross-situational word learning. *Cogn. Sci.* 37, 891–921

22 McMurray, B. (2007) Defusing the childhood vocabulary explosion. *Science* 317, 631

23 Fazly, A. *et al.* (2010) A probabilistic computational model of cross-situational word learning. *Cogn. Sci.* 34, 1017–1063

24 Regier, T. (2005) The emergence of words: attentional learning in form and meaning. *Cogn. Sci.* 29, 819–865

25 Hidaka, S. and Smith, L.B. (2011) Packing: a geometric analysis of feature selection and category formation. *Cogn. Syst. Res.* 12, 1–18

26 Smith, L. and Yu, C. (2008) Infants rapidly learn word–referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568

27 Yu, C. and Smith, L.B. (2011) What you learn is what you see: using eye movements to study infant cross-situational word learning. *Dev. Sci.* 14, 165–180

28 Vouloumanos, A. and Werker, J.F. (2009) Infants' learning of novel words in a stochastic environment. *Dev. Psychol.* 45, 1611–1617

29 Smith, L.B. and Yu, C. (2013) Visual attention is not enough: individual differences in statistical word–referent learning in infants. *Lang. Learn. Dev.* 9, 25–49

30 Colombo, J. (1995) On the neural mechanisms underlying developmental and individual differences in visual fixation in infancy: two hypotheses. *Dev. Rev.* 15, 97–135

31 Oakes, L.M. (2011) *Infant Perception and Cognition: Recent Advances, Emerging Theories, and Future Directions*, Oxford University Press

32 Ellis, E.M. *et al.* (2013) Visual prediction in infancy: what is the association with later vocabulary? *Lang. Learn. Dev.* 10, 36–50

33 Fitneva, S.A. and Christiansen, M.H. (2011) Looking in the wrong direction correlates with more accurate word learning. *Cogn. Sci.* 35, 367–380

34 Vlach, H.A. and Johnson, S.P. (2013) Memory constraints on infants' cross-situational statistical learning. *Cognition* 127, 375–382

35 Frank, M.C. *et al.* (2013) Social and discourse contributions to the determination of reference in cross-situational word learning. *Lang. Learn. Dev.* 9, 1–24

36 Medina, T.N. *et al.* (2011) How words can and cannot be learned by observation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9014–9019

37 Trueswell, J.C. *et al.* (2013) Propose but verify: fast mapping meets cross-situational word learning. *Cogn. Psychol.* 66, 126–156

38 Yurovsky, D. *et al.* (2013) Statistical word learning at scale: the baby's view is better. *Dev. Sci.* 16, 959–966

39 Cartmill, E.A. *et al.* (2013) Quality of early parent input predicts child vocabulary 3 years later. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11278–11283

40 Kidd, C. *et al.* (2012) The Goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE* 7, e36399

41 Turk-Browne, N.B. and Scholl, B.J. (2009) Flexible visual statistical learning: transfer across space and time. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 195

42 Alvarez, G.A. (2011) Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn. Sci.* 15, 122–131

43 Brady, T.F. and Alvarez, G.A. (2011) Hierarchical encoding in visual working memory ensemble statistics bias memory for individual items. *Psychol. Sci.* 22, 384–392

44 Patai, E.Z. *et al.* (2012) Long-term memories bias sensitivity and target selection in complex scenes. *J. Cogn. Neurosci.* 24, 2281–2291

45 Summerfield, C. and Egner, T. (2009) Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* 13, 403–409

46 Squire, L.R. (2004) Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* 82, 171–177

47 Thiessen, E.D. (2010) Effects of visual information on adults' and infants' auditory statistical learning. *Cogn. Sci.* 34, 1093–1106

48 Yeung, H.H. *et al.* (2013) Referential labeling can facilitate phonetic learning in infancy. *Child Dev.* http://dx.doi.org/10.1111/cdev.12185

49 Hay, J.F. *et al.* (2011) Linking sounds to meanings: infant statistical learning in a natural language. *Cogn. Psychol.* 63, 93–106

50 Hills, T.T. *et al.* (2010) The associative structure of language: contextual diversity in early word learning. *J. Mem. Lang.* 63, 259–273

51 Tulving, E.E. and Craik, F.I. (2000) *The Oxford Handbook of Memory*, Oxford University Press

52 Perry, L. *et al.* (2013) Highchair philosophers: the impact of seating context-dependent exploration on children's naming biases. *Dev. Sci.* http://dx.doi.org/10.1111/desc.12147

53 Vlach, H.A. and Sandhofer, C.M. (2012) Distributing learning over time: the spacing effect in children's acquisition and generalization of science concepts. *Child Dev.* 83, 1137–1144

54 Vlach, H.A. and Sandhofer, C.M. (2011) Developmental differences in children's context-dependent word learning. *J. Exp. Child Psychol.* 108, 394–401

55 Werchan, D.M. and Gómez, R.L. (2014) Wakefulness (not sleep) promotes generalization of word learning in 2.5-year-old children. *Child Dev.* 85, 429–436

56 Yoshida, H. and Smith, L.B. (2008) What's in view for toddlers? Using a head camera to study visual experience. *Infancy* 13, 229–248

57 Smith, L.B. *et al.* (2011) Not your mother's view: the dynamics of toddler visual experience. *Dev. Sci.* 14, 9–17

58 Aslin, R.N. (2009) How infants view natural scenes gathered from a head-mounted camera. *Optom. Vis. Sci.* 86, 561–565

59 Yu, C. and Smith, L.B. (2012) Embodied attention and word learning by toddlers. *Cognition* 125, 244–262

60 Pereira, A. *et al.* (2014) A bottom-up view of toddler word learning. *Psychon. Bull. Rev.* 21, 178–185

61 Franchak, J.M. *et al.* (2011) Head-mounted eye tracking: a new method to describe infant looking. *Child Dev.* 82, 1738–1750

62 Yu, C. and Smith, L.B. (2013) Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye–hand coordination. *PloS ONE* 8, e79659

63 Kretch, K. *et al.* (2012) What infants see depends on locomotor posture. *J. Vis.* 12, 182

64 James, K.H. *et al.* (2013) Young children's self-generated object views and object recognition. *J. Cogn. Dev.* http://dx.doi.org/10.1080/15248372.2012.749481