Cell PRESS

Perceptual and memory constraints on language acquisition

Ansgar D. Endress[1,*], Marina Nespor[2] and Jacques Mehler[3]

[1] Harvard University, William James Hall 984, 33 Kirkland St, Cambridge, MA 02138, USA
[2] University of Milano Bicocca, Centro Linceo Beniamino Segre, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy
[3] International School for Advanced Studies, Via Beirut 2-4, 34151 Trieste, Italy

A wide variety of organisms employ specialized mechanisms to cope with the demands of their environment. We suggest that the same is true for humans when acquiring artificial grammars, and at least some basic properties of natural grammars. We show that two basic mechanisms can explain many results in artificial grammar learning experiments, and different linguistic regularities ranging from stress assignment to interfaces between different components of grammar. One mechanism is sensitive to identity relations, whereas the other uses sequence edges as anchor points for extracting positional regularities. This piecemeal approach to mental computations helps to explain otherwise perplexing data, and offers a working hypothesis on how statistical and symbolic accounts of cognitive processes could be bridged.

Specialized mechanisms in language acquisition

For the last fifty years, language has been conceived as a complex, rule-governed system with a strong innate basis [1,2]. At the same time, researchers have uncovered powerful (and ubiquitous) statistical learning machinery that infants might use to acquire the grammar of their maternal languages. This machinery might thus dispense with the need for at least some of the postulated innate constraints for language acquisition [3,4]. Here, we suggest a different (and complementary) approach to language acquisition, focusing on perceptual and memory functions that constrain how serially presented materials such as speech utterances are processed. Specifically, we propose that originally non-linguistic mechanisms were recruited by the language faculty, and now constrain the kinds of regularities language learners can acquire from their input.

We first review evidence illustrating that, to understand many animal behaviors in different species, one must assume the existence of a variety of specialized learning mechanisms. Given this evidence, we argue that it would be strange if humans were not similarly endowed with specialized learning abilities. We then show that many results from artificial grammar learning experiments rest upon specialized mechanisms, which might derive in part from perceptual or memory organization. Moreover, the same kinds of specialized mechanisms have important roles in natural language acquisition and processing. We conclude that language might use a toolbox of such specialized 'perceptual or memory primitives' (POMPs), which might complement and constrain both rule-based and statistical accounts of language acquisition and use (see the Glossary and Figure 1 for brief descriptions of the terminology). In other words, although no other animal has language, certain linguistic computations might draw on specialized, phylogenetically pre-existing mechanisms, and these mechanisms might determine what kinds of linguistic structures humans can learn.

The ubiquity of specialized mechanisms

Many animals can perform surprisingly complex computations based on highly specialized mechanisms [5,6]. For example, bees can predict the solar ephemeris [7], although they never learnt about Newtonian physics. Rather, they seem to have a highly specialized ability for orienting themselves during foraging. Specialized mechanisms of

Glossary

In the following, we provide informal definitions of terms used throughout this article. The different terms are also illustrated in Figure 1 in the main text. However, we are agnostic as to whether all logically possible combinations of these concepts are realized in the mind.

Specialized mechanism: specialized mechanisms subserve only one (and no other) function. For instance, a mechanism devoted to processing identity-relations can detect repetitions, but cannot be used for, say, determining the position of an item in a sequence. Specialized mechanisms may or may not be domain-specific.

Domain-specific mechanism: a domain-specific mechanism is a specialized or non-specialized mechanism that exists exclusively in one particular domain. For instance, a mechanism to recognize colors despite illumination changes will be found only in vision, but not in other domains.

Domain-general mechanism: a domain-general mechanism is a specialized or non-specialized mechanism that can be accessed from different domains. Attentional processing might be an example.

Domain-bound mechanisms: domain-bound mechanisms are similar to domain-specific mechanisms, except that they shared by at least two domains. For example, one can imagine distinct mechanisms for processing identity-relations in different domains such as vision, audition and language that are all independent of one another. That is, although these mechanisms might have very similar properties (they all process identity-relations), each of these mechanisms might operate in only one specific domain. For example, the identity-processing mechanism in vision might not process auditory identity-relations, and vice-versa. As domain-bound mechanisms are available in various domains, they are hard to distinguish from domain-general mechanisms (and, to our knowledge, have not been distinguished from them). The crucial difference, however, is that each of the mechanisms is implemented in such a way that it operates in only one particular domain.

Perceptual or memory primitive: a perceptual or memory primitive is a specialized, domain-bound mechanism that subserves a variety of computations, for example to learn grammatical regularities. It might often derive from phylogenetically pre-existing perceptual or memory processes, and might (but needs not) be shared with non-human animals.

---

*Corresponding author.Endress, A.D. (ansgar.endress@m4x.org)

1364-6613/$ – see front matter © 2009 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2009.05.005

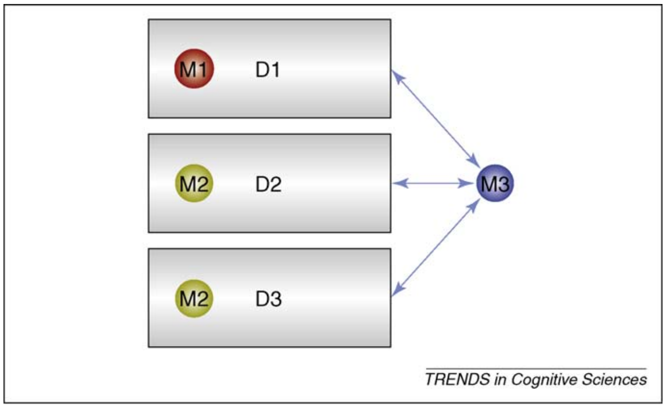

Figure 1. Boxes represent domains, circles represent mechanisms. Mechanism M1 (red) is 'domain-specific' because it exists exclusively in domain D1 and in no other domain. Mechanism M3 (blue) is 'domain-general' because it can be accessed by all domains. Mechanism M2 (yellow) is 'domain-bound' because 'copies' of this mechanism exist in more than one domain (here in D2 and D3). Each of the copies of M2 can operate only within its own domain.

this kind abound at all levels of cognitive processing, from social cognition [8] to decision-making [9,10], but also more generally in biology [11]. They also include many 'learning' mechanisms. Even classical conditioning – the prototypical example of a learning mechanism often thought to apply to all stimuli alike – is in fact highly specialized. For example, rats easily associate tastes with visceral sickness, and external events (such as lights) with electro-shocks, but not tastes with electro-shocks or external events with sickness [12]. This specialization is adaptive because sickness usually results from ingesting toxic food, and pain from external events; it demonstrates that even supposedly general-purpose learning mechanisms can in fact be highly specialized.

Here, we argue that the mechanisms required to learn some grammatical aspects of language are just as specialized. In particular, we suggest that a toolbox of surprisingly simple and specialized mechanisms might be used to learn some grammatical regularities, and might also constrain the kinds of regularities that can be learned. They might both provide the machinery necessary to learn certain regularities and act as filters to the input of other grammar learning mechanisms. Furthermore, we argue that these mechanisms might be based on phylogenetically pre-existing perceptual and memory processes that were at some point recruited for grammatical purposes. These mechanisms are, thus, not necessarily adaptations to particular ecological niches, but seem to exist in different animals.

Specialized mechanisms in artificial grammar learning

In artificial grammar learning experiments, participants are exposed to materials constructed according to certain simple rules. These rules often mirror key aspects of language acquisition, but are hugely simplified relative to realistic language learning situations. How do learners succeed in such experiments? Here, we review studies suggesting that they rely on some POMPs, that is, computational mechanisms specialized for one particular function (e.g. computing identity relations), and that might often derive from pre-existing perceptual or memory mechanisms. The availability of the appropriate primitives might thus determine the kinds of regularities learners can acquire most naturally.

We thus attempt here to explain why some structures (but not others) are easily learnable. This differentiates our approach from previous suggestions that memory limitations (such as a reduced working memory capacity) are instrumental in language acquisition [13] because such accounts predict that all structures should be equally easy to learn as long as they fit a learners' memory-span.

Repetition-based structures in artificial grammar learning

We start by illustrating how the best known demonstration of rule learning in infants can be explained if one assumes the existence of a primitive that is sensitive to identity relations. Marcus *et al.* [14] showed that 7-month-old infants can generalize (repetition-based) structures such as AAB, ABB and ABA. For example, after familiarization with syllable triplets conforming to one of the structures (e.g. *ba-ba-de* for AAB), infants were tested on new triplets with novel syllables that conformed either to the familiar structure or to a new one. Marcus *et al.* [14] showed that infants recognize the structure despite the use of new syllables during the test phase. They concluded that the syllable positions in the triplets act as variables, and that infants discover relations among such variables.

These results raise the question of whether young infants can extract arbitrary relations between variables (at least as long as these relations are not overly 'complex'), or whether repetitions might rather be special relations that can be extracted and generalized much more readily than others.

Further experiments support the latter interpretation. Indeed, repetition-based generalizations seem to be computed by a specialized mechanism that only repetitions can trigger. For example, Endress *et al.* [15] asked whether participants could generalize only repetition-based structures, or also other, comparably simple, structures. Using piano tones, they tested whether participants would generalize only the repetition-based structures ABA and ABB, or whether they would also learn other melodic structures such as Lowest-Highest-Middle or Middle-Highest-Lowest. Participants performed well for the repetition-based structures, but very poorly (although slightly above chance) for the other structures. Next, Endress *et al.* [15] showed that these two types of structures are formally equally 'complex.' Hence, although participants had no *a priori* reason to learn one type of structure rather than the other, they performed much better on the repetition-based structures, suggesting that there is something special about repetitions.

Further support for the 'repetitions are special' hypothesis comes from artificial grammar learning experiments in which participants are typically familiarized with (visual) consonantal letter sequences. Unbeknownst to the participants, these sequences conform to a simple underlying grammar. Such a familiarization enables participants to classify new consonant strings as grammatical or ungrammatical – even when the test strings use a 'different' consonant set than the one used in the training strings [16,17]. Further research revealed a surprising basis for these generalizations. The grammars typically used in artificial grammar learning experiments generate characteristic repetition patterns in the strings. Participants seem to use these repetition patterns for their generalizations: when these are removed, participants can no longer generalize to new strings using a new consonant set [18,19].

These results indicate that humans use a specialized primitive devoted to processing identity relations. Although this POMP responds much more strongly to repetitions of adjacent elements, it also seems to process repetition at a (short) distance (Á.M. Kovács, PhD thesis, SISSA, Trieste, Italy, 2008). The POMP seems to be innate, and might be present in different modalities. Not only are human neonates sensitive to repetition structures [20] but even bees are sensitive to visual and olfactory identity relations [21]. It thus seems that a repetition-processing primitive is available in different animals [21–23], and in humans even at birth [20].

Edge-based positional regularities in artificial grammar learning

A second POMP turned out to be important in all artificial grammar learning experiments in which participants had to learn the positions of items in a sequence. Such positions seem to be encoded relative to the sequence edges as anchor points.

Indeed, people do not only memorize that an item has occurred in a sequence but also, to some extent, where in the sequence it occurred [24–28]. These two processes – remembering the occurrence of an item, and remembering where in a sequence that item occurred – seem to be rather independent. In memory experiments, in which people have to memorize ordered lists of items, a common error is to recall an item in its correct sequential position – but in a wrong list, that is, in another list than the one currently recalled [24,25]. Accordingly, most current models of serial memory assume that participants have access to a mechanism allowing them to memorize the positions of items. Figure 2 shows possible models of how the positions of items in sequences are encoded. Although the specific implementations differ rather widely across models, almost all of them assume, in one form or another, that positions are encoded relative to edges [25,27,28], and that only edges have proper positional codes. As all other positions are encoded relative to edges, they are encoded much less reliably. For instance, in Henson's model [25], the activity of a 'start' marker decreases during a sequence, whereas the activity of an 'end' marker increases; their relative strengths represent the position of an item in that sequence. Edge positions can thus be encoded very accurately, whereas non-edge positions are encoded less well.

Edge-based positional codes turned out to be important also in many artificial grammar-learning experiments. For instance, in Marcus *et al.*'s [14] experiments, infants might (i) detect that there is a repetition and (ii) link this repetition to the appropriate edge-based positional code. That is, they might simply learn that there is a repetition, and that this repetition has to be in either the left or the right edge of the sequence. Indeed, using longer sequences and adult participants, Endress *et al.* [29] showed that repetition-based structures are generalized only when the repetitions are in sequence-edges (e.g. in structures such as ABCDE*FF*, in which each letter stands for a syllable), but not when the repetitions are located in middle positions (e.g. in ABC*DD*EF).

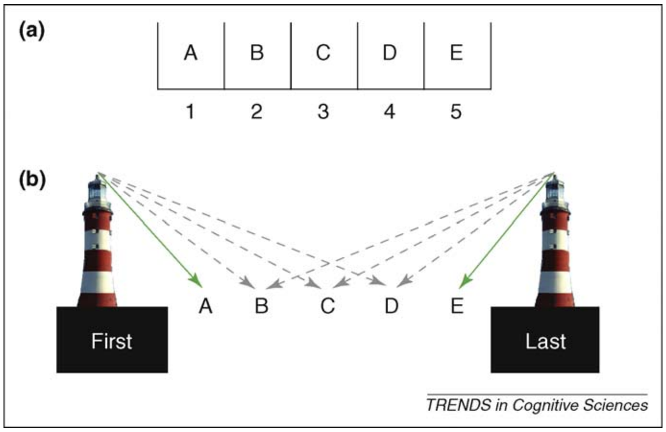

Figure 2. Possible mechanisms for memory for positions in sequences. For memorizing the positions of the elements of the sequence *ABCDE*, different models have been proposed. **(a)** Each position might correspond to a single slot (represented by a box), and each element might become associated with one slot. **(b)** Each sequence might have two marker points, corresponding to the first and the last position. All positions are then encoded relative to these marker points. As a result, the first and the last position can be encoded very accurately, whereas all other positions are encoded less accurately. Much evidence suggests that humans use a mechanism such as the one illustrated in (b) for memorizing the positions of items (see Refs [25,26] for reviews). Photograph by D. Skinner, licensed under the Creative Commons Attribution and Share Alike License 2.0.

Note that the aforementioned results are not just due to the salience of the edges. Indeed, participants discriminated sequences that differed in their middle syllables just as well as sequences that differed in their edge syllables (in fact, they were at ceiling in both cases). This would be unexpected if the generalization deficit were simply due to problems perceiving middle syllables. Rather, it seems that participants had difficulties encoding and generalizing the repetitions and their positions when these were not at the sequence edges. Thus, Endress *et al.* [29] concluded that the only positions that act as variables are the edges. In other words, participants might use one POMP to detect the repetitions, and then another POMP to determine where in the sequence the repetitions occur.

Edge-based positional codes turned out to be important also for virtually all recent artificial grammar learning experiments in which participants had to learn the items' positions. For example, in some experiments [30,31], participants were familiarized with a stream of trisyllabic nonsense 'words' in which the first syllable predicted the last syllable with certainty; the middle syllable, by contrast, was variable. Although participants could use this dependency for segmenting the 'words' from the continuous familiarization stream, they could not generalize the dependency to new words even after arbitrarily long familiarizations. By contrast, when 'words' in the streams were separated by 25 ms silences, a familiarization of only two minutes was sufficient for participants to generalize a rule; they learned that some syllables had to occur in the first positions of words, whereas others had to occur in the last position [31]. Further research showed that, when longer 'words' were used, participants readily generalized a similar rule when the crucial syllables occupied the edge positions, but not when they were in middle positions. Statistical processes, by contrast, proceed fairly well also for middle syllables [32]. That is, these results illustrate the interplay of a fast mechanism registering the positions of items that occurred in word edges, and of a slower one tracking the statistical syllable distribution.

The importance of edges is also reflected in 'phonotactic' constraints. Such constraints govern permissible sound sequences in a language. Using monosyllabic Consonant-Vowel-Consonant (CVC) words, for example, Chambers and collaborators [33,34] showed that adults and young infants can learn that certain consonants must occur in syllable onsets, whereas others must occur in coda positions (that is, in syllable offsets). In monosyllabic words, however, syllable onsets and codas are edges of words; thus, one might ask whether these generalizations are also related to edge-based codes. When using bi-syllabic words, adult participants generalized similar constraints when the critical consonants occupied the word-edges, but not when they were in word-internal positions. Again, control conditions showed that these results were not due to a failure to perceive middle consonants [35]. Hence, Chambers *et al.*'s [33,34] participants might have learned to link particular consonants to positional codes for word-edges.

To sum up, the two aforementioned POMPs – one computing identity relations and one computing positional codes in edges – surface in a wide array of artificial grammar learning results. In the next section, we review evidence suggesting that the very same POMPs also have an important role in natural languages.

Specialized mechanisms in natural language learning

Compared to the complex input to infant learners, artificial grammar learning studies use highly simplified stimuli. Still, the two mechanisms just reviewed are important to account not only for diverse artificial grammar learning results but also for many linguistic phenomena.

Repetitions in natural language

Reduplication (that is, the repetition of a word root or parts of it) is a prominent feature of child-directed speech [36–38], and might thus be important for word learning. As reduplication entails the repetition of phonetic material, it might well appeal to a mechanism that can process identity relations. Reduplication is also extensively used in both derivational and inflectional morphology across languages [39]. In Micronesian languages, for example, word-final syllable reduplication is used extensively in derivational morphology (e.g. to derive verbs or adjectives from nouns, as in Marshallese *takin* 'sock', *takinkin* 'to wear socks' [40]). Likewise, in Classical Greek, consonantal left edge reduplication is used in verbal inflection for the perfect tense, as in **π**αιδεύω [**p**aideuo] 'I educate' versus **π**ε**π**αίδευκα [**pep**aideuka] 'I educated'; specifically, words are prefixed with a syllable whose first consonants is a copy of the word-initial consonant, and whose vowel is [e]. Although more research is needed to establish whether this type of phenomenon relies on identity-detecting mechanisms similar to those found in the experiments reviewed here, it is at least striking that repetitions have a prominent role both in artificial grammar learning and in morphology.

Edge-based positional regularities in natural language

Edges seem to have a crucial role in a variety of grammatical phenomena. Word stress, for example, is always defined with respect to one of the edges; it can fall on the left edge (as in Hungarian) or on the right edge (as in French), or on another syllable that is counted from the right edge – as in Italian, in which stress falls on one of the last three syllables – most frequently on the second syllable from the right. Stress is never assigned relative to positions other than the word edges [41].

Edges are also crucial to most morphological processes. Reduplications, for example, are found mainly at edges of morphological constituents. Across the languages of the world, reduplications are widespread either in initial or in final position. Middle reduplications, by contrast, are rarely attested. Moreover, when these arise, they occur at edges of phonological constituents; that is, they are at edges of a different level of representation [42].

Affixation provides another example of the importance of edges in morphology. Prefixes and suffixes are cross-linguistically much more frequent than infixes [43]; this observation follows naturally if one assumes that edge-based positional codes are used for affixation (for a similar proposal, see Ref. [44]).

Edges might also be important for the hierarchical organization that is crucial to language (and many other cognitive functions; e.g. Refs [45–47]). For example, the morphosyntactic and phonological hierarchies do not always coincide, as in cases where a morpheme is not a syllable (e.g. the English plural [s]). Even though the morphosyntactic and the phonological hierarchies do not match, at least one of the edges of their constituents do; in the case of the plural [s], for instance, the right edge of the morpheme is always aligned with the right edge of a syllable (that is, their right edges coincide). Likewise, at the phrasal level, at least one edge of a syntactic phrase is always aligned with an edge of a phonological phrase [48]. Edges thus seem to be important not only for morphological processes but also to coordinate different hierarchies [48,49]. Hence, both repetitions and edges seem to be important not only for explaining numerous artificial grammar learning results but seem to be exploited also by many languages for various grammatical processes.

A toolbox for alleviating some learning problems

The data reviewed earlier indicate that an operation sensitive to identity relations and edge-based positional codes might be involved in many artificial grammar learning experiments, and in a variety of linguistic phenomena. Conceivably, the language faculty might use a toolbox of these and other primitives for its own purposes. If such primitives are innate – which is plausible at least for repetitions because already human neonates are sensitive to them [20], they might alleviate at least some learning problems. Certain specialized primitives – some of which might have been recycled from memory and perceptual mechanisms – might thus facilitate the acquisition of some grammatical regularities.

In principle, each POMP might be either specific to a particular domain, or accessible to different domains. Figure 1 illustrates the different possibilities. Take the primitive processing identity relations as an example. Plausibly, there might be a single, generic domain-general process that detects identity relations. Alternatively, however, this primitive might be what we call '*domain-bound*' in Figure 1: when analyzing speech input, acoustic, phonological, morphological and other computations might all have their own way of processing identity relations, and these operations might all be independent of one another. Hence, even if an operation can be performed in different modules, the corresponding primitive is not necessarily a generic, domain-general mechanism, but it might be domain-bound.

The hypothesis that each module has its own independent and domain-bound toolbox of primitives receives support from linguistic theory. In Semitic languages, for example, consonants might be repeated at the right edge of a word root, but not at its left edge [50–52]. The standard explanation is that the phonological representation of word roots forbids repeated consonants [52]. However, if a root has only two consonants rather than three, as most roots, the second consonant is reduplicated at the end to fit the canonic word pattern (which has three slots for consonants); this reduplication thus leads to a consonant repetition in the word's final syllable. However, the reduplication occurs at the morphological rather than at the phonological level, in which the prohibition of repeated consonants holds. Hence, both the phonological and the morphological level have to appeal to identity relations; they can do so if each level has its own operations processing identity-relations, whereas the data are more difficult to explain if they appeal to one single, generic identity-processing mechanism.

The hypothesis that each level of representation might have its own set of primitives also gains support when edges are considered. Based on the data reviewed earlier, edge-based positional codes define the positions of different grammatical units (such as syllables and phonemes) and of repetitions; in other words, although syllables, phonemes and repetitions of them are very different objects, they could all be linked to positional codes. Likewise, linguistic theory establishes that constituents on all levels of representation have their own edges and that these must be aligned in specific ways [48,49]. We thus speculate that each level of representation and each module might have its own set of primitives, and that some modules might, but need not, share some primitives.

Conclusion

The results reviewed here indicate that the language faculty might use a collection of POMPs that might be in part specific to language, and in part recycled from other domains, in particular from perception and memory. This view receives support from proposals that some abstract grammatical principles might derive from processing constraints [53,54]. We have provided evidence for two examples of such primitives, namely a mechanism processing identity-relations, and one giving positional tags to items occurring at edges of sequences. Although POMPs like those highlighted in this article do not pinpoint what is special about language, they provide a conceptual and methodological foundation for exploring specific mechanisms mediating language processing and their psychological basis, and might shed light on some of the constraints that make certain grammatical structures more learnable than others (Box 1).

Box 1. Outstanding questions

- Which aspects of grammar can be explained based on POMPs, and which aspects evolved for specifically linguistic purposes, with no homologues in other domains or other animals?
- How do POMPs get integrated with the specifically linguistic aspects of language (e.g. phonological, syntactic or semantic representations)?
- Can non-human animals learn language-like structures based on POMPs, and what are the limits to their learning abilities?
- What, if any, is the contribution of the environment to the ontogenetic unfolding of POMPs?
- In addition to domain-bound mechanisms, are there domain-general mechanisms that are truly accessible *to* different domains rather than being localized copies in different domains?
- Can domain-bound POMPs interact with POMPs in other domains?

Acknowledgements

This research has been supported by McDonnell Foundation Grant 21002089 and the European Commission Special Targeted Project CALACEI (contract N° 12778 NEST) and by the Mind, Brain, and Behavior Interfaculty Initiative at Harvard University. We are grateful to J. Gervain, M. Hauser, Á. Kovács and J. M. Toro for helpful discussions and comments on earlier versions of this manuscript.

References

1 Chomsky, N. (1980) *Rules and Representation.* Basil Blackwell
2 Lenneberg, E.H. (1967) *Biological Foundations of Language.* John Wiley and Sons
3 McClelland, J.L., Rumelhart, D.E. and The PDP Research Group, eds (1986). *Parallel Distributed Processing, Volume 2: Psychological and Biological Models*, MIT Press
4 Saffran, J.R. (2003) Statistical language learning: mechanisms and constraints. *Curr. Dir. Psychol. Sci.* 12, 110–114
5 Gallistel, C.R. (2000) The replacement of general-purpose learning models with adaptively specialized learning modules. In *The Cognitive Neurosciences* (2nd edn) (Gazzaniga, M.S., ed.), pp. 1179–1191, MIT Press
6 Hauser, M.D. (2000) *Wild Minds: What Animals Really Think.* Henry Holt
7 Dyer, F.C. and Dickinson, J.A. (1994) Development of sun compensation by honeybees: How partially experienced bees estimate the sun's course. *Proc. Natl. Acad. Sci. USA* 91, 4471–4474
8 Sugiyama, L.S. *et al.* (2002) Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proc. Natl. Acad. Sci. USA* 99, 11537–11542
9 Tversky, A. and Kahneman, D. (1974) Judgement under uncertainty: Heuristics and biases. *Science* 185, 1124–1131
10 Gigerenzer, G. *et al.* Group (1999). Simple *heuristics that make us smart*, Oxford University Press
11 Jacob, F. (1977) Evolution and tinkering. *Science* 196, 1161–1166
12 Garcia, J. *et al.* (1974) Behavioral regulation of the milieu interne in man and rat. *Science* 185, 824–831
13 Newport, E.L. (1990) Maturational constraints on language learning. *Cogn. Sci.* 14, 11–28
14 Marcus, G.F. *et al.* (1999) Rule learning by seven-month-old infants. *Science* 283, 77–80
15 Endress, A.D. *et al.* (2007) Perceptual constraints and the learnability of simple grammars. *Cognition* 105, 577–614

16 Altmann, G.T.M. *et al.* (1995) Modality independence of implicitly learned grammatical knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 899–912

17 Reber, A.S. (1969) Transfer of syntactic structure in synthetic languages. *J. Exp. Psychol.* 81, 115–119

18 Gómez, R.L. *et al.* (2000) The basis of transfer in artificial grammar learning. *Mem. Cognit.* 28, 253–263

19 Tunney, R.J. and Altmann, G.T.M. (2001) Two modes of transfer in artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 614–639

20 Gervain, J. *et al.* (2008) The neonate brain detects speech structure. *Proc. Natl. Acad. Sci. USA* 105, 14222–14227

21 Giurfa, M. *et al.* (2001) The concepts of 'sameness' and 'difference' in an insect. *Nature* 410, 930–933

22 Hauser, M.D. *et al.* (2002) Rule learning by cotton-top tamarins. *Cognition* 86, B15–B22

23 Murphy, R.A. *et al.* (2008) Rule learning by rats. *Science* 319, 1849–1851

24 Conrad, R. (1960) Serial order intrusions in immediate memory. *Br. J. Psychol.* 51, 45–48

25 Henson, R.N. (1998) Short-term memory for serial order: The Start-End Model. *Cognit. Psychol.* 36, 73–137

26 Henson, R.N. (2001) Serial order in short-term memory. *Psychologist* 14, 70–73

27 Ng, H.L.H. and Maybery, M.T. (2002) Grouping in short-term verbal memory: is position coded temporally? *Q. J. Exp. Psychol. A* 55, 391–424

28 Hitch, G.J. *et al.* (1996) Temporal grouping effects in immediate recall: a working memory analysis. *Q. J. Exp. Psychol. A* 49, 116–139

29 Endress, A.D. *et al.* (2005) The role of salience in the extraction of algebraic rules. *J. Exp. Psychol. Gen.* 134, 406–419

30 Peña, M. *et al.* (2002) Signal driven computations in speech processing. *Science* 298, 604–607

31 Endress, A.D. and Bonatti, L.L. (2007) Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition* 105, 247–299

32 Endress, A.D. and Mehler, J. Primitive computations in speech processing. Q. J. Exp. Psychol. (in press) doi: 10.1080/17470210902783646.

33 Chambers, K.E. *et al.* (2003) Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87, B69–B77

34 Onishi, K.H. *et al.* (2002) Learning phonotactic constraints from brief auditory experience. *Cognition* 83, B13–B23

35 Endress, A.D. and Mehler, J. Perceptual Constraints in Phonotactic Learning. *J. Exp. Psychol. Hum. Percept. Perform.*(in press).

36 Ferguson, C. (1964) Baby talk in six languages. *Am. Anthropol.* 66, 103–114

37 Haynes, L.M. and Cooper, R.L. (1986) A note on Ferguson's proposed baby-talk universals. In *The Fergusonian Impact: In Honor of Charles A. Ferguson. Volume 1: From Phonology to Society* (Fishman, J. *et al.*, eds), pp. 127–134, Mouton de Gruyter

38 Shi, R. *et al.* (1998) Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *J. Child Lang.* 25, 169–201

39 McCarthy, J.J. and Prince, A. (1999) Faithfulness and identity in prosodic morphology. In *The Prosody Morphology Interface* (Kager, R. *et al.*, eds), pp. 218–309, Cambridge University Press

40 Moravcsik, E. (1978) Reduplicative constructions. In *Universals of Human Language: Word Structure* (Volume 3) (Greenberg, J.H., ed.), In pp. 297–334, Stanford University Press

41 Hayes, B. (1995) *Metrical Stress Theory: Principles and Case Studies.* University of Chicago Press

42 Broselow, E. and McCarthy, J. (1983) A theory of internal reduplication. *Ling. Rev.* 3, 25–88

43 Greenberg, J.H. (1957) *Essays in Linguistics.* University of Chicago Press

44 Slobin, D. (1973) Cognitive prerequisites for the development of grammar. In *Studies of Child Language Development* (Ferguson, C. and Slobin, D., eds), pp. 175–208, Holt, Reinhart & Winston

45 Hauser, M.D. *et al.* (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579

46 Marr, D. and Nishihara, H.K. (1992) Visual information processing: artificial intelligence and the sensorium of sight. In *Frontiers in Cognitive Neuroscience* (Kosslyn, S.M. and Andersen, R.A., eds), pp. 165–186, MIT Press

47 Cooper, R.P. and Shallice, T. (2006) Hierarchical schemas and goals in the control of sequential behavior. *Psychol. Rev.* 113, 887–916

48 Nespor, M. and Vogel, I. (2008) *Prosodic Phonology*. Mouton de Gruyter

49 McCarthy, J.J. and Prince, A. (1993) Generalized alignment. In *Yearbook of Morphology 1993* (Booij, G. and van Marle, J., eds), pp. 79–153, Kluwer

50 Berent, I. *et al.* (2007) Roots, stems, and the universality of lexical representations: Evidence from Hebrew. *Cognition* 104, 254–286

51 Frisch, S.A. *et al.* (2004) Similarity avoidance and the OCP. *Nat. Lang. Ling. Theory* 22, 179–228

52 McCarthy, J.J. (1979). *Formal Problems in Semitic Phonology and Morphology*, PhD thesis, MIT, Cambridge.

53 Bever, T.G. (1970) The cognitive basis for linguistic structures. In *Cognition and the Development of Language* (Hayes, J.R., ed.), pp. 279–362, John Wiley and Sons

54 Kimball, J. (1973) Seven principles of surface structure parsing in natural language. *Cognition* 2, 15–47