Statistical learning in infants

Gerry T. M. Altmann*

Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom

Statistical learning has become the subject of some considerable debate within cognitive psychology (1, 2). The debate reached particular prominence with the advent of sophisticated computational models of such learning based on neurally inspired notions of spreading activation within highly distributed systems of interacting units, so-called neural networks (3). One important development in this field was the notion that representations of higher-level structure might "emerge" on the basis of an initial sensitivity to low-level statistical cooccurrence phenomena (4). Thus, within the field of language research, higher-level theoretical constructs such as the grammatical class of a word (as noun or verb, for example) would emerge within a system that was sensitive only to the statistical distributions of words within sentences. Importantly, this emergence represented little more than simple statistical clustering; the internal representations of words that would tend to occur in similar distributional contexts would cluster together, and because nouns tend to occur in particular sentential contexts and verbs in others, the clustering of words into these two classes (and others with even finer distinctions between the classes) was in some sense a statistical inevitability. One important feature of such models was that although learning within these models was based on a sensitivity to statistically predictable variation in their input, beyond that, the real-world nature of that input (whether pertaining to words, sounds, letters, or shapes) did not matter; the experimenter might deem a particular pattern of activation across the "input units" to represent a particular kind of linguistic stimulus, but these inputs could be deemed just as easily to represent visual stimuli. It is thus significant that Fiser and Aslin (5) have demonstrated the sensitivity of infants to statistical properties of visual input.

A variety of studies have demonstrated effects that are equivalent in some respects to those reported by Fiser and Aslin but have been observed in the linguistic domain. For example, infants can use statistical regularities (conditional probabilities) in the linguistic input to develop sensitivity to word boundaries (6) and aspects of grammatical structure (7). These effects cannot be explained simply in terms of sensitivity to the frequency of occurrence of individual elements, because in these studies they depended on the conditional probabilities between one element and another, a distinction that is explored and confirmed by Fiser and Aslin in the visual domain. Such evidence of a statistical underpinning to aspects of cognition has provoked considerable controversy with respect to language learning because of the claim that language acquisition requires an algebraic component, in which mental representations corresponding to algebra-like expressions are assumed to support the acquisition of language grammars (8). Infant abilities that have been claimed to demand algebraic processing (9) have been shown recently to be susceptible to statistical modeling (10, 11) and even more recently have been observed in cotton-top tamarin monkeys (12). The findings of Fiser and Aslin in the visual domain open up the possibility that a common statistical-learning device serves (aspects of) both language and vision. Of course, the full range of abilities that serves language processing, or visual scene interpretation, may well require learning abilities other than those best described as "statistical" and may well be constrained by innate factors also, but it is nonetheless useful to consider how far statistical learning can go in respect of accounting not only for learning in the visual domain or, separately, in the language domain, but for learning the coupling between language and vision itself.

Evidence of a statistical underpinning to aspects of cognition has provoked considerable controversy with respect to language learning.

There exists a very tight coupling between language and vision, as evidenced by the speed with which eye movements around a visual scene can be mediated by linguistic input, whether constituting a command to manipulate objects in the scene (13) or a description of what may happen or may have happened to objects in the scene (14). Indeed, the coupling is so tight that it appears as if eye movements toward objects in the scene are planned as soon as the mental representations corresponding to words referring to those objects are themselves activated (15). The rapidity with which the eyes can be directed toward particular objects is not confined only to occasions when words are used that refer directly to those objects. For example, hearing a sequence such as "the man will ride" while viewing a scene portraying a man, a child (a girl), a motorbike, and a fairground carousel causes the eyes to be directed toward the motorbike during the verb "ride," but when the sequence is "the girl will ride," the eyes are directed toward the carousel instead (16). These effects reflect the rapid integration of various sources of information including the meaning of the verb, the meaning of its grammatical subject ("the man" or "the girl"), real-world knowledge about the plausibility of alternative scenarios (who is doing the riding and what they are likely to ride), and information contained within the visual scene (concerning, among other things, the objects that are ridable and the actual individuals who may be doing the riding). The relevance of this tight coupling for results such as those described by Fiser and Aslin is that these interactions between language and vision reflect experiential knowledge regarding the roles that entities can play in the event to which a verb refers (17) and, in doing so, reflect conditional probabilities in respect to the ways in which entities in the real world interact with one another. The coupling is predicated on a correspondence between conditional probabilities regarding elements in the language and conditional probabilities regarding elements in the real world. As such, one can ask whether the associative processes observed in infants by Fiser and Aslin and others (18) might also be responsible for the early mapping, in infancy, between statistical regularities in

See companion article on page 15822.

*E-mail: g.altmann@psych.york.ac.uk.

www.pnas.org/cgi/doi/10.1073/pnas.262659399

the language and statistical regularities in the world which that language describes.

That infants can map between language and the world is not in doubt. Language use is predicated on that mapping, and young infants are able to integrate information they receive at a verb such as "drink" to direct attention toward drinkable objects portrayed in a visual scene before them (A. Fernald, personal communication). Moreover, their ability to do this is mediated by the size of their vocabulary (even though they might know the meaning of the specific verbs used) and, by extension, their grammatical competence (19). However, although it is self-evident that infants are able to compute such mappings, it is less clear on what basis they do this. One possibility is suggested by recent work with infants who were required to learn conditional probabilities between elements in a small artificial language. Whereas in the Fiser and Aslin study the conditional probabilities were "instantiated" in different spatial configurations between different visual elements, in these language studies the conditional probabilities were instantiated in different possible sequences of syllables based on which syllables could follow or be adjacent to which other syllables. Infants who were habituated to a set of sequences instantiating one set of conditional probabilities dishabituated if given test sequences instantiating different conditional probabilities (7). However, the critical finding in these studies was that it did not matter if the test sequences used the same syllables as had been used for the habituation phase or different ones (7, 9). In terms of the Fiser and Aslin study, this procedure would be equivalent to habituating to one set of visual symbols but then being tested on new symbols that had not been seen before but which either had the same statistical properties as the habituation stimuli or different ones. The relevance of this finding is in the mechanism that might underlie this ability. Several studies have modeled these data by using neural networks (see ref. 10 for review), which as indicated earlier are agnostic as to the nature of the real-world inputs they receive. Thus, the ability to map between conditional probabilities expressed in one set of syllables and those expressed in another in such models would translate just as easily into an ability to map between conditional probabilities pertaining to linguistic elements and those pertaining to visual elements. Two questions remain. Are infants able to map between statistical regularities in an artificial language and statistical regularities in an artificial visual world (as in the Fiser and Aslin study)? If they can, do they do so according to the same statistical principles identified in recent statistical-learning models?

In the absence of empirical data, we can only conjecture. The Fiser and Aslin study does suggest, however, that at least the first of these two questions is empirically addressable; they have demonstrated one way in which conditional probabilities can be manipulated within the visual domain, and in principle it should be possible to test infants with multimodal stimuli and thus explore whether there is indeed a statistical basis for the coupling between language and vision. If infants, like adults (20), could indeed map statistical regularities in one modality onto statistical regularities in another, it would remain an open question whether they did so according to the principles embodied in recent models of statistical learning.

The Fiser and Aslin study is timely for a variety of reasons, not the least of which is the promise it holds for future research. It is a first step toward a more detailed exploration of statistical learning by infants in the visual domain as well as a more detailed exploration of the commonalities between statistical learning in the visual and linguistic domains. Nonetheless, the statistical nature of the visual world in which infants develop is rather different from that manipulated in the Fiser and Aslin study; the conditional probabilities to which infants are sensitive apply not simply to spatial configuration but also to changes that happen to objects across time. They apply to the manner in which objects interact and the manner in which the state of the world dynamically changes across time. The challenges that face researchers in visual cognition, language development, and the interface between the two are to explore further the infant's ability to extract statistical information from the visual and ever-changing world and to determine how such information might underpin the formation of concepts regarding the objects in that world and the interactions between those objects. These concepts necessarily underpin language use. Contemporary theories of language acquisition stress the manner in which particular kinds of structure in the visual world focus attention on particular kinds of structure within the sentences that describe that world and, conversely, the manner in which particular structures within sentences focus attention on particular kinds of structure in the visual world (21). This symbiotic relationship between language and vision may well be based on statistical learning. The Fiser and Aslin study is an important step toward determining whether this is the case.

1. Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996) *Rethinking Innateness: A Connectionist Perspective on Development* (MIT Press/Bradford Books, Cambridge, MA).
2. Marcus, G. F. (1998) *Cognit. Psychol.* **37,** 243–282.
3. Rumelhart, D. E. & McClelland, J. L. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (MIT Press, Cambridge, MA).
4. Elman, J. L. (1990) *Cognit. Sci.* **14,** 179–211.
5. Fiser, J. & Aslin, R. N. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 15822–15826.
6. Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996) *Science* **274,** 1926–1928.
7. Gomez, R. L. & Gerken, L. A. (1999) *Cognition* **70,** 109–136.
8. Marcus, G. (2001) *The Algebraic Mind* (MIT Press, Cambridge, MA).
9. Marcus, G. F., Vijayan, S., Bandi Rao, S. & Vishton, P. M. (1999) *Science* **283,** 77–80.
10. Altmann, G. T. M. (2002) *Cognition* **85,** B43–B50.
11. Altmann, G. T. M. & Dienes, Z. (1999) *Science* **284,** 875.
12. Hauser, M. D., Weiss, D. & Marcus, G. (2002) *Cognition* **86,** B15–B22.
13. Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. (1995) *Science* **268,** 1632–1634.
14. Altmann, G. T. M. & Kamide, Y. (1999) *Cognition* **73,** 247–264.
15. Allopenna, P. D., Magnuson, J. S. & Tanenhaus, M. K. (1998) *J. Mem. Lang.* **38,** 419–439.
16. Altmann, G. T. M. & Kamide, Y. (2003) in *The Interface of Language, Vision, and Action: What We Can Learn from Free-Viewing Eyetracking*, eds. Henderson, J. & Ferreira, F. (Psychology Press, New York).
17. McRae, K., Ferretti, T. R. & Amyote, L. (1997) *Lang. Cognit. Proc.* **12,** 137–176.
18. Gomez, R. L. & Gerken, L. (2000) *Trends Cogn. Sci.* **4,** 178–186.
19. Bates, E. & Goodman, J. C. (1997) *Lang. Cognit. Proc.* **12,** 507–584.
20. Altmann, G. T. M., Dienes, Z. & Goode, A. (1995) *J. Exp. Psychol. Learn. Mem. Cogn.* **21,** 899–912.
21. Gleitman, L. R. & Gillette, J. (1995) in *The Handbook of Child Language*, eds. Fletcher, P. & MacWhinney, B. (Blackwell, Oxford).