Statistical Learning by 8-Month-Old Infants

Jenny R. Saffran; Richard N. Aslin; Elissa L. Newport

Science, New Series, Vol. 274, No. 5294 (Dec. 13, 1996), 1926-1928.

infection of murine cells (*15*) and transgenic mice expressing human CD4 (*16*) and provides a rationale for transgenic approaches to developing animal models of HIV disease.

REFERENCES AND NOTES

1. F. Cocchi *et al.*, *Science* **270**, 1811 (1995).
2. Y. Feng, C. C. Broder, P. E. Kennedy, E. A. Berger, *ibid.* **272**, 872 (1996).
3. M. Samson, O. Labbe, C. Mollereau, G. Vassart, M. Parmentier, *Biochemistry* **35**, 3362 (1996); C. J. Raport, J. Gosling, V. L. Schweickart, P. W. Gray, I. F. Charo, *J. Biol. Chem.* **271**, 17161 (1996).
4. H. Choe *et al.*, *Cell* **85**, 1135 (1996); B. J. Doranz *et al.*, *ibid.*, p. 1149.
5. T. Dragic *et al.*, *Nature* **381**, 667 (1996); H. Deng *et al.*, *ibid.*, p. 661; G. Alkhatib *et al.*, *Science* **272**, 1955 (1996).
6. S. Gartner *et al.*, *Science* **233**, 215 (1986).
7. R. Atchison *et al.*, unpublished observations.
8. L. Boring *et al.*, *J. Biol. Chem.* **271**, 7551 (1996).
9. We cloned cDNAs encoding human or murine CCR5 into the expression vector pcDNA3 (Invitrogen) after engineering the FLAG epitope into the NH_2-terminus as described (*13*). Expression of each construct was determined by FACS with an antibody to FLAG (anti-FLAG) (Boehringer Mannheim), and relative expression for each (see below) was calculated as the percentage of cells expressing human CCR5 on the cell surface normalized to the expression of hCCR5 (defined as 100%), with standard errors of the mean. The mean fluorescence intensity of the positive cells from any single sample never varied from the average by more than 30% in a single experiment. Therefore, neither the relative number of positive cells nor the absolute expression levels within transfected cells explains the differences in coreceptor activity. Chimeric receptors were prepared by the overlap polymerase chain reaction (PCR) method (*17*). hCCR5 (*HHHH*), human CCR5 (100% relative expression); mCCR5 (*MMMM*), murine CCR5 (126 ± 49%); *HMMM*, NH_2-terminus of human CCR5 [amino acids (aa) 1 to 32] fused to murine CCR5 (aa 35 to 354) (77 ± 22%); *MHHH*, NH_2-terminus of murine CCR5 (aa 1 to 34) fused to human CCR5 (aa 33 to 352) (73 ± 17%); *MHMM*, extracellular loop 1 and a portion of transmembrane domain 3 of human CCR5 (aa 86 to 118) replacing the corresponding segment of the murine receptor (aa 88 to 120) (37 ± 22%); *MMHM*, extracellular loop 2 and adjacent portions of human CCR5 (aa 134 to 210) replacing the corresponding region of the murine receptor (aa 136 to 212) (81 ± 30%); *MMHH*, NH_2-terminal half of mCCR5 (aa 1 to 162) fused to the COOH-terminal half of hCCR5 (aa 161 to 352) (80 ± 39%).
10. I. F. Charo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2752 (1994).
11. C. Franci, L. M. Wong, J. Van Damme, P. Proost, I. F. Charo, *J. Immunol.* **154**, 6511 (1995).
12. We cloned cDNAs encoding human CCR2B or chimeras into the expression vector pCMV4 (*18*) after engineering the FLAG epitope into the NH_2-terminus as described (*13*). Expression of each construct (see below) was determined as described earlier. Chimeric receptors were prepared by the overlap PCR method (*17*). *5555*, human CCR5 (100% relative expression); 2222, human CCR2B (87 ± 2%); 5222, NH_2-terminus of CCR5 (aa 1 to 32) fused to CCR2B (aa 45 to 360) (27 ± 5%); 2555, NH_2-terminus of CCR2B (aa 1 to 44) fused to CCR5 (aa 33 to 352) (108 ± 17%); 2255, CCR2B (aa 1 to 136) fused to CCR5 (aa 124 to 352) (119 ± 33%).
13. F. S. Monteclaro and I. F. Charo, *J. Biol. Chem.* **271**, 19084 (1996); F. S. Monteclaro *et al.*, unpublished observations.
14. J. Gosling *et al.*, unpublished observations.
15. P. J. Maddon *et al.*, *Cell* **47**, 333 (1986).
16. P. Lores *et al.*, *AIDS Res. Hum. Retroviruses* **8**, 2063 (1992).
17. S. N. Ho, H. D. Hunt, R. M. Horton, J. K. Pullen, L. R. Pease, *Gene* **77**, 51 (1989).
18. S. Andersson, D. L. Davis, H. Dahlback, H. Jornvall, D. W. Russell, *J. Biol. Chem.* **264**, 8222 (1989).
19. M. A. Goldsmith, M. T. Warmerdam, R. E. Atchison, M. D. Miller, W. C. Greene, *J. Virol.* **69**, 4112 (1995).
20. COS-7 cells were transfected with 2 μg of plasmid DNA per well in a six-well plate as described (*19*). DNA samples consisted of appropriate combinations of 0.5 μg of a human CD4 expression plasmid [pCD4Neo (*19*)] or plain vector, and 1.5 μg of a chemokine receptor–expressing plasmid or plain vector. About 30 hours after addition of DNA, the medium in each well was replaced with 1.0 ml of medium containing HIV-1 Ba-L (~100 to 170 ng of p24 per sample; source: NIH AIDS Reagent Repository, passaged on primary human macrophages). About 10 hours later, an additional 1.0 ml of medium was added to each well. After 30 hours, the cells were recovered from the dish as described (*19*) and analyzed with a FacScan (Becton Dickinson). Staining for intracytoplasmic HIV-1 p24 was carried out with the Fix and Perm reagents (Caltag Laboratories), with a monoclonal antibody to p24 (Coulter Immunology) and goat anti-mouse fluorescein isothiocyanate (FITC)–conjugated secondary antibody (Becton Dickinson). Cells were further stained with phycoerythrin (PE)–conjugated anti-CD4 (Becton Dickinson). Appropriate controls indicated that the appearance of double-positive cells (FITC + PE) was dependent on cotransfection with both CD4 and human CCR5 expression plasmids and on the presence of HIV-1 Ba-L.
21. H. Arai and I. F. Charo, *J. Biol. Chem.* **271**, 21814 (1996).

22. We acknowledge the advice of M. Warmerdam (transfection-infection assay), E. Weider (FACS studies), and L. Boring, H. Arai, and R. Speck (scientific interpretation). We appreciate the assistance of J. Carroll and M. Ceniceros in the preparation of this manuscript. Supported in part by NIH grant HL52773 (I.F.C.) and by Pfizer (M.A.G.).

24 September 1996; accepted 24 October 1996

Statistical Learning by 8-Month-Old Infants

Jenny R. Saffran, Richard N. Aslin, Elissa L. Newport

Learners rely on a combination of experience-independent and experience-dependent mechanisms to extract information from the environment. Language acquisition involves both types of mechanisms, but most theorists emphasize the relative importance of experience-independent mechanisms. The present study shows that a fundamental task of language acquisition, segmentation of words from fluent speech, can be accomplished by 8-month-old infants based solely on the statistical relationships between neighboring speech sounds. Moreover, this word segmentation was based on statistical learning from only 2 minutes of exposure, suggesting that infants have access to a powerful mechanism for the computation of statistical properties of the language input.

During early development, the speed and accuracy with which an organism extracts environmental information can be extremely important for its survival. Some species have evolved highly constrained neural mechanisms to ensure that environmental information is properly interpreted, even in the absence of experience with the environment (*1*). Other species are dependent on a period of interaction with the environment that clarifies the information to which attention should be directed and the consequences of behaviors guided by that information (*2*). Depending on the developmental status and the task facing a particular organism, both experience-independent and experience-dependent mechanisms may be involved in the extraction of information and the control of behavior.

In the domain of language acquisition, two facts have supported the interpretation that experience-independent mechanisms are both necessary and dominant. First, highly complex forms of language production develop extremely rapidly (*3*). Second, the language input available to the young child is both incomplete and sparsely represented compared to the child's eventual linguistic abilities (*4*). Thus, most theories of language acquisition have emphasized the critical role played by experience-independent internal structures over the role of experience-dependent factors (*5*).

It is undeniable that experience-dependent mechanisms are also required for the acquisition of language. Many aspects of a particular natural language must be acquired from listening experience. For example, acquiring the specific words and phonological structure of a language requires exposure to a significant corpus of language input. Moreover, long before infants begin to produce their native language, they acquire information about its sound properties (*6*). Nevertheless, given the daunting task of acquiring linguistic information from listening experience during early development, few theorists have entertained the hypothesis that learning plays a primary role in the acquisition of more complicated aspects of language, favoring instead experience-independent mechanisms (*7*). Young humans are generally viewed as poor learners, suggesting that innate factors are primarily responsible for the acquisition of language.

Here we investigate the nature of the

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA.

experience-dependent factors involved in language acquisition. In particular, we ask whether infants are in fact better learners than has previously been assumed, thus potentially reducing the extent to which experience-independent structures must be posited. The results demonstrate that infants possess powerful mechanisms suited to learning the types of structures exemplified in linguistic systems. Experience may therefore play a more important role in the acquisition of language than existing theories suggest.

One task faced by all language learners is the segmentation of fluent speech into words. This process is particularly difficult because word boundaries in fluent speech are marked inconsistently by discrete acoustic events such as pauses (*8*). Although it has recently been demonstrated that 8-month-old infants can segment words from fluent speech and subsequently recognize them when presented in isolation (*9*), it is not clear what information is used by infants to discover word boundaries. This problem is complicated by the variable acoustic structure of speech across different languages, suggesting that infants must discover which, if any, acoustic cues correlated with word boundaries are relevant to their native language (*10*); there is no invariant acoustic cue to word boundaries present in all languages.

One important source of information that can, in principle, define word boundaries in any natural language is the statistical information contained in sequences of sounds. Over a corpus of speech there are measurable statistical regularities that distinguish recurring sound sequences that comprise words from the more accidental sound sequences that occur across word boundaries (*11*). Within a language, the transitional probability from one sound to the next will generally be highest when the two sounds follow one another within a word, whereas transitional probabilities spanning a word boundary will be relatively low (*12*). For example, given the sound sequence *pretty#baby*, the transitional probability from *pre* to *ty* is greater than the transitional probability from *ty* to *ba*. Previously, we showed that adults and children can use information about transitional probabilities to discover word boundaries in an artificial language corpus of nonsense words presented as continuous speech, with no acoustic cues to word boundaries (*13*).

We asked whether 8-month-old infants can extract information about word boundaries solely on the basis of the sequential statistics of concatenated speech. We used the familiarization-preference procedure developed by Jusczyk and Aslin (*9*). In this procedure, infants are exposed to auditory material that serves as a potential learning experience. They are subsequently presented with two types of test stimuli: (i) items that were contained within the familiarization material and (ii) items that are highly similar but (by some critical criterion) were not contained within the familiarization material. During a series of test trials that immediately follows familiarization, infants control the duration of each test trial by their sustained visual fixation on a blinking light (*14*). If infants have extracted the crucial information about the familiarization items, they may show differential durations of fixation (listening) during the two types of test trials (*15*). We used this procedure to determine whether infants can acquire the statistical properties of sound sequences from brief exposures.

In our first experiment, 24 8-month-old infants from an American-English language environment were familiarized with 2 min of a continuous speech stream consisting of four three-syllable nonsense words (hereafter, "words") repeated in random order (*16*). The speech stream was generated by a speech synthesizer in a monotone female voice at a rate of 270 syllables per minute (180 words in total). The synthesizer provided no acoustic information about word boundaries, resulting in a continuous stream of coarticulated consonant-vowel syllables, with no pauses, stress differences, or any other acoustic or prosodic cues to word boundaries. A sample of the speech stream is the orthographic string *bidakupadotigolabubidaku*. . . . The only cues to word boundaries were the transitional probabilities between syllable pairs, which were higher within words (1.0 in all cases, for example, *bida*) than between words (0.33 in all cases, for example, *kupa*).

To assess learning, each infant was presented with repetitions of one of four three-syllable strings on each test trial. Two of these three-syllable strings were "words" from the artificial language presented during familiarization, and two were three-syllable "nonwords" that contained the same syllables heard during familiarization but not in the order in which they appeared as words (*17*).

The infants showed a significant test-trial discrimination between word and nonword stimuli (*18*), with longer listening times for nonwords (Table 1). This novelty preference, or dishabituation effect, indicates that 8-month-olds recognized the difference between the novel and the familiar orderings of the three-syllable strings. Thus, 8-month-old infants are capable of extracting serial-order information after only 2 min of listening experience.

Of course, simple serial-order information is an insufficient cue to word boundaries. The learner must also be able to extract the relative frequencies of co-occurrence of sound pairs, where relatively low transitional probabilities signal word boundaries. Our next experiment examined whether 8-month-olds could perform the more difficult statistical computations required to distinguish words (that is, recurrent syllable sequences) from syllable strings spanning word boundaries (that is, syllable sequences occurring more rarely). To take an English example, *pretty#baby*, we wanted to see if infants can distinguish a word-internal syllable pair like *pretty* from a word-external syllable pair like *ty#ba*.

Another 24 8-month-old infants from an American-English language environment were familiarized with 2 min of a continuous speech stream consisting of three-syllable nonsense words similar in structure to the artificial language used in our first experiment (*19*). This time, however, the test items for each infant consisted of two words and two "part-words." The part-words were created by joining the final syllable of a word to the first two syllables of another word. Thus, the part-words contained three-syllable sequences that the infant had heard during familiarization but statistically, over the corpus, did not correspond to words (*20*). These part-words could only be judged as novel if the infants had learned the words with sufficient specificity and completeness that sequences crossing a word boundary were relatively unfamiliar.

Despite the difficulty of this word versus part-word discrimination, infants showed a significant test-trial discrimination between the word and part-word stimuli (*21*), with longer listening times for part-words (Table 1). Thus, 2 min of exposure to concatenated speech organized into "words" was suffi-

Table 1. Mean time spent listening to the familiar and novel stimuli for experiment 1 (words versus nonwords) and experiment 2 (words versus part-words) and significance tests comparing the listening times.

Experiment	Mean listening times (s)		Matched-pairs *t* test
	Familiar items	Novel items	
1	7.97 (SE = 0.41)	8.85 (SE = 0.45)	$t(23) = 2.3$, $P < 0.04$
2	6.77 (SE = 0.44)	7.60 (SE = 0.42)	$t(23) = 2.4$, $P < 0.03$

cient for 8-month-old infants to extract information about the sequential statistics of syllables. Moreover, this novelty preference cannot be attributed to a total lack of experience with the three-syllable sequences forming part-words, as was the case with the nonwords in the first experiment. Rather, infants succeeded in learning and remembering particular groupings of three-syllable strings—those strings containing higher transitional probabilities surrounded by lower transitional probabilities.

The infants' performance in these studies is particularly impressive given the impoverished nature of the familiarization speech stream, which contained no pauses, intonational patterns, or any other cues that, in normal speech, probabilistically supplement the sequential statistics inherent in the structure of words. Equally impressive is the fact that 8-month-old infants in both experiments were able to extract information about sequential statistics from only 2 min of listening experience. Although experience with speech in the real world is unlikely to be as concentrated as it was in these studies, infants in more natural settings presumably benefit from other types of cues correlated with statistical information.

Our results raise the intriguing possibility that infants possess experience-dependent mechanisms that may be powerful enough to support not only word segmentation but also the acquisition of other aspects of language. It remains unclear whether the statistical learning we observed is indicative of a mechanism specific to language acquisition or of a general learning mechanism applicable to a broad range of distributional analyses of environmental input (*22*). Regardless, the existence of computational abilities that extract structure so rapidly suggests that it is premature to assert a priori how much of the striking knowledge base of human infants is primarily a result of experience-independent mechanisms. In particular, some aspects of early development may turn out to be best characterized as resulting from innately biased statistical learning mechanisms rather than innate knowledge. If this is the case, then the massive amount of experience gathered by infants during the first postnatal year may play a far greater role in development than has previously been recognized.

REFERENCES AND NOTES

1. Certain species-specific skills develop without any experiential input, including bat echolocation [E. Gould, *Dev. Psychobiol.* **8**, 33 (1975)] and cricket song [R. Hoy, *Am. Zool.* **14**, 1067 (1974)].
2. Examples of behaviors mediated by early experience are imprinting [E. Hess, *Imprinting* (Van Nostrand, New York, 1973); M. Leon, *Physiol. Behav.* **14**, 311 (1975)] and suckling responses in newborn rats [M. H. Teicher and E. M. Blass, *Science* **198**, 635 (1977)].
3. These milestones have been well-documented both in English [for example, R. Brown, *A First Language* (Harvard Univ. Press, Cambridge, MA, 1973)] and cross-linguistically [for example, E. Lenneberg, *Biological Foundations of Language* (Wiley, New York, 1967); D. Slobin, Ed., vols. 1 to 3 of *The Crosslinguistic Study of Language Acquisition* (Erlbaum, Hillsdale, NJ, 1985, 1987, 1992)].
4. This "argument from the poverty of the stimulus" remains widely accepted [for example, N. Chomsky, *Aspects of the Theory of Syntax* (MIT Press, Cambridge, MA, 1965); S. Crain, *Behav. Brain Sci.* **14**, 597 (1991)].
5. D. Bickerton, *Behav. Brain Sci.* **7**, 173 (1984); N. Chomsky, *Rules and Representations* (Columbia Univ. Press, New York, 1981); J. Fodor, *Modularity of Mind* (MIT Press, Cambridge, MA, 1983); L. Gleitman and E. Newport, in *Language: An Invitation to Cognitive Science*, L. Gleitman and M. Liberman, Eds. (MIT Press, Cambridge, MA, 1995), pp. 1–24.
6. Examples include vowel structure [P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, B. Lindblom, *Science* **255**, 606 (1992)], phonotactics [P. Jusczyk, A. Friederici, J. Wessels, V. Svenkerud, A. Jusczyk, *J. Mem. Lang.* **32**, 401 (1993)], and prosodic structure [P. Jusczyk, A. Cutler, N. Redanz, *Child Dev.* **64**, 675 (1993)].
7. Exceptions include research on prenatal exposure to maternal speech [A. DeCasper, J.-P. Lecanuet, M.-C. Busnel, C. Granier-Deferre, R. Maugeais, *Infant Behav. Dev.* **17**, 159 (1994)] and early postnatal preferences [J. Mehler *et al.*, *Cognition* **29**, 149 (1988)].
8. R. Cole and J. Jakimik, in *Perception and Production of Fluent Speech*, R. Cole, Ed. (Erlbaum, Hillsdale, NJ, 1980), pp. 133–163.
9. P. Jusczyk and R. Aslin, *Cognitive Psychol.* **29**, 1 (1995).
10. A. Christophe, E. Dupoux, J. Bertoncini, J. Mehler, *J. Acoust. Soc. Am.* **95**, 1570 (1994); A. Cutler and D. Carter, *Comput. Speech Lang.* **2**, 133 (1987).
11. Z. Harris, *Language* **31**, 190 (1955); J. Hayes and H. Clark, in *Cognition and the Development of Language*, J. Hayes, Ed. (Wiley, New York, 1970). See M. Brent and T. Cartwright [*Cognition* **61**, 93 (1996)] for a discussion of related statistical cues to word boundaries.
12. The transitional probability of

$$Y|X = \frac{\text{frequency of } XY}{\text{frequency of } X}$$

13. J. Saffran, E. Newport, R. Aslin, *J. Mem. Lang.* **35**, 606 (1996); ———, R. Tunick, S. Barrueco, *Psychol. Sci.*, in press.
14. Each infant was tested individually while seated on the parent's lap in a sound-attenuated booth. Synthetic speech was generated off-line by the MacinTalk system and stored on disk at a sampling rate of 22 kHz for on-line playback through an Audiomedia board in an Apple Quadra 650 computer. An observer outside the testing booth monitored the infant's looking behavior with the use of a color video system, using a buttonbox connected to the computer to initiate trials and score head-turn responses. Both the parent and the observer listened to masking music over headphones to eliminate bias. During the 2-min familiarization phase, the infant's gaze was first directed to a blinking light located on the front wall of the testing booth, and then the sound sequence was presented from two loudspeakers located on the side walls. The infant's gaze was directed to one of two blinking lights on these side walls during familiarization, but there was no relation between lights and sound. Immediately after familiarization, 12 test trials were presented (six words and six nonwords). Each test trial began with the central blinking light. When the observer signaled with a button press that the infant had fixated on the central light, one of the two side blinking lights was turned on and the center light was extinguished. When the infant faced the side light (a head turn of at least 30° in the direction of the light), the three-syllable test string was played and repeated until the infant looked away from the light for 2 s or until 15 s of looking had occurred. The observer simply recorded the direction of the infant's head turn, and the computer measured looking times, determined when the 2-s lookaway criterion had been met, and controlled the randomization and presentation of stimuli. Cumulative looking time across each of the two types of test trials provided the measure of preference.
15. The direction of the fixation preference depends on the degree of familiarity with the stimuli. If the infants have become highly familiar with the stimuli, they show dishabituation behavior, preferring the novel stimuli.
16. Two counterbalanced stimulus conditions were generated. For each condition, 45 tokens of each of four trisyllabic nonsense words (condition A: *tupiro*, *golabu*, *bidaku*, and *padoti*; condition B: *dapiku*, *tilado*, *burobi*, and *pagotu*) were spoken in random order to create a 2-min speech stream, with the stipulation that the same word never occurred twice in a row.
17. Test stimuli: *tupiro*, *golabu*, *dapiku*, and *tilado*. In condition A, the first two strings were words and the last two strings were nonwords (the transitional probabilities between the syllables in the nonwords were all zero relative to the exposure corpus, as these syllable pairs had never occurred during familiarization). In condition B, the first two strings were nonwords and the last two strings were words. This between-subjects counterbalanced design ensured that any observed preferences for words or nonwords across both conditions would not be artifacts of any general preferences for certain syllable strings. Each of the four test strings were presented (repeated with a 500-ms interval between test strings) on three different trials, resulting in a total of 12 test trials per infant.
18. There were no significant differences between the infants in condition A and condition B: $t(22) = 0.31$. The data from the two groups were thus combined for the other analyses.
19. Condition A words: *pabiku*, *tibudo*, *golatu*, and *daropi*; condition B words: *tudaro*, *pigola*, *bikuti*, and *budopa*.
20. Test stimuli: *pabiku*, *tibudo*, *tudaro*, and *pigola*. In condition A, the first two strings were words and the second two strings were part-words. For example, the part-word *pigola* spanned the word boundary between *daropi#golatu* and thus was heard during exposure. In condition B, the first two strings were part-words and the second two strings were words. The part-words were thus three-syllable sequences that the infants had heard during the course of the exposure period. The difficulty of this test discrimination can be seen by comparing the transitional probabilities between the syllables in the words (1.0 between syllables 1 and 2 and between syllables 2 and 3) to the transitional probabilities between the syllables in the part-words (0.33 between syllables 1 and 2 and 1.0 between syllables 2 and 3).
21. There were no significant differences between the infants in condition A and condition B: $t(22) = 0.49$. The data from the two groups were thus combined for the other analyses.
22. For example, this same general mechanism could be used to find an object, such as a human face, in the environment.

23. We thank J. Gallipeau, J. Hooker, P. Jusczyk, A. Jusczyk, T. Mintz, K. Ruppert, and J. Sawusch for their help with various aspects of this research, and P. Jusczyk, S. Pollak, M. Spivey-Knowlton, and M. Tanenhaus for their helpful comments on a previous draft. Supported by an NSF predoctoral fellowship (J.R.S.), NSF grant SBR9421064 (R.N.A.), and NIH grant DC00167 (E.L.N.). The parents of all participants gave informed consent.

10 May 1996; accepted 30 September 1996