



Brief article

Visual statistical learning in infancy: evidence for a domain general learning mechanism

Natasha Z. Kirkham*, Jonathan A. Slemmer, Scott P. Johnson

Cornell University, Ithaca, NY, USA

Received 10 October 2001; accepted 2 January 2002

Abstract

The rapidity with which infants come to understand language and events in their surroundings has prompted speculation concerning innate knowledge structures that guide language acquisition and object knowledge. Recently, however, evidence has emerged that by 8 months, infants can extract statistical patterns in auditory input that are based on transitional probabilities defining the sequencing of the input's components (Science 274 (1996) 1926). This finding suggests powerful learning mechanisms that are functional in infancy, and raises questions about the domain generality of such mechanisms. We habituated 2-, 5-, and 8-month-old infants to sequences of discrete visual stimuli whose ordering followed a statistically predictable pattern. The infants subsequently viewed the familiar pattern alternating with a novel sequence of identical stimulus components, and exhibited significantly greater interest in the novel sequence at all ages. These results provide support for the likelihood of domain general statistical learning in infancy, and imply that mechanisms designed to detect structure inherent in the environment may play an important role in cognitive development. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Visual statistical learning; Infancy; Domain general learning mechanism

1. Introduction

A central question asked by developmental psychologists concerns how infants learn so much in so little time, often with little explicit instruction. The rapidity and ease with which children understand and produce speech, for example, have led to the postulation of an innate device that allows the young child to discover how his or her native language embodies those principles common to all languages (Chomsky, 1965; Gleitman & Wanner, 1982; Pinker, 1984). In object perception tasks, likewise, infants' facility at

* Corresponding author. Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853, USA. Tel.: +1-607-255-6398; fax: +1-607-255-8433.

E-mail address: nzk2@cornell.edu (N.Z. Kirkham).

recognizing apparent violations of physical laws (e.g. the persistence of objects after occlusion) has been cited as evidence for innate reasoning skills that provide some knowledge about fundamental object properties (Carey, 1999; Spelke & Van de Walle, 1993; Wynn, 1998). On such accounts the role of learning is subsidiary to a set of specialized cognitive modules, operational early in infancy, from which domain specific knowledge emerges (cf. Fodor, 1983). Such views may risk neglecting the potential role of environmental structure in guiding development. Natural visual scenes, for example, are richly structured and contain a high degree of statistical redundancy, showing considerable consistency across space and time (Field, 1994), and there is evidence that development of response properties of some visual neurons exploits the statistical nature of the input (Olshausen & Field, 1996). In the present study, we asked whether young infants are able to extract statistical information in visual stimuli, to explore a larger question of the nature of learning in infancy: are early learning mechanisms attuned to input structure in the visual environment?

Several recent experiments have reported evidence that infants readily learn statistically defined patterns in auditory input sequences. Saffran, Aslin, and Newport (1996), for example, investigated 8-month-old infants' word segmentation in a corpus of artificial speech. They noted that in natural speech, adjacent sounds that co-occur with a high probability are usually found within words, whereas low probability sound pairs tend to span word boundaries. This difference in likelihood of co-occurrence provides potential information for word boundaries, and may contribute to early language acquisition by bolstering the ability to segment the speech stream into meaningful units.

To investigate infants' discrimination of high and low probability sound pairs within a corpus of speech, Saffran et al. (1996) presented 8-month-olds with a synthesized speech stream consisting of four three-syllable "words" composed of 12 unique syllables (e.g. *tupiro*, *golabu*, *dapiku*, and *tilado*), presented in random order (e.g. *dapikutupirotiladogolabutupiro...*) for 2 min. The only cues to word boundaries were the transitional probabilities between syllable pairs. For example, the transitional probability of *tu-pi* in this corpus was 1.00, because *pi* always followed *tu* within the word *tupiro* (a within-word syllable pair), whereas the transitional probability of *ro-go* was 0.33, because *golabu* was one of three words that could follow *tupiro* (a between-word syllable pair). Following exposure, the infants then heard isolated instances of familiar words, presented repeatedly. In one test condition, the familiar words were presented randomly within blocks that also contained "non-words", and in another test condition the familiar words were presented with "part-words". Non-words were created from novel combinations of syllables from the familiar speech stream (e.g. *pitugo*), and part-words were created by combining the last syllable of one word with the first two syllables of a second word (e.g. *kutilla*). The infants showed a reliably greater interest in the non-words and part-words than in the words, as revealed by a preferential head-turning paradigm. On the logic that infants often exhibit post-familiarization novelty preferences (Bornstein, 1985), this suggests that they distinguished between the words and the other stimuli based on learning the transitional probabilities defining word boundaries. (See Aslin, Saffran, and Newport (1998) for evidence that the results stem from true computation of input statistics rather than simple frequency counting.)

These findings provide evidence for a functional statistical learning mechanism, avail-

able to infants, that segments speech into words on the basis of computation of input statistics, and gives rise to questions concerning the generality of this ability. Saffran, Johnson, Aslin, and Newport (1999) found that 8-month-olds detected transitional probabilities of non-linguistic tone sequences, indicating that statistical learning is not a purely linguistic mechanism. Further evidence for generality comes from experiments by Hauser, Newport, and Aslin (2001) with non-human primates (cotton-top tamarins, a species of New World monkey). After exposure to the same set of auditory stimuli employed by Saffran et al. (1996), adult monkeys showed reliably greater interest in both non-words and part-words than in the familiar words, suggesting that they were able to extract the statistical information defining word boundaries in the artificial speech, in like manner to human infants. These experiments imply that statistical learning may be a general purpose learning device, but it is unknown at present if young infants can detect statistically defined structure in sequential visual stimuli.

The present study had two goals. First, we probed the question of domain generality of statistical learning in infancy by asking whether it is limited to auditory information. This was accomplished with a task in which infants were presented with sequential visual input that contained probabilistic structure. Second, we tested infants younger than those observed by Saffran et al. (1996, 1999), to probe the developmental time-course of statistical learning during the first year after birth. We used a visual habituation procedure, an effective tool for investigating perceptual and cognitive processes in infants as young as neonates (Slater, 1995).

2. Method

2.1. Participants

Forty-eight full-term infants (22 females) composed the final sample, 16 2-month-olds (M age = 64.9 days, SD 7.4), 16 5-month-olds (M age = 151.6 days, SD 8.3), and 16 8-month-olds (M age = 240.9 days, SD 15.9). Nine additional infants were observed but not included in the analyses due to fussiness ($n = 5$), sleepiness ($n = 2$), or equipment failure ($n = 2$). The infants were recruited by letter and telephone from hospital records and birth announcements in the local newspaper. Parents and infants received a small gift (a baby t-shirt or toy) for their participation.

2.2. Apparatus and stimuli

A Macintosh G4 computer and 53 cm color monitor were used to present stimuli and collect looking time data. An observer viewed the infant on a second monitor and entered looking judgments with a keypress on the computer keyboard. The observer was unaware of the stimulus sequence viewed by the infant. The computer presented displays, recorded looking times, calculated the habituation criterion for each infant, and changed displays after the criterion was met. Stimuli consisted of six colored shapes (turquoise square, blue cross, yellow circle, pink diamond, green triangle, and red octagon) presented one at a time in a continuous stream, with no break or delay between shapes. Each shape was presented for 1 s and loomed from 4 to 24 cm in height (2.4–14.6°). (Pilot testing revealed that

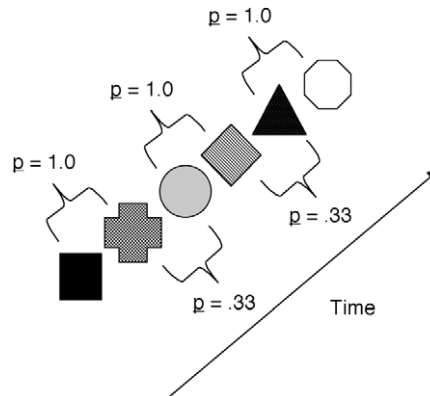


Fig. 1. Schematic depiction of the familiar stimulus sequence showing the transitional probabilities defining pairs. (Each stimulus was a unique color, rather than the patterns shown here.)

looming was effective for maintaining infants' attention throughout the experiment.) The stream of stimuli was shown as long as the infant attended to the monitor. A trial ended when the infant looked away for 2 s, or had looked for 60 s (the maximum trial duration was lengthened to 90 s for the 2-month-olds). Between trials, a beeping target was shown to attract attention back to the screen.

2.3. Procedure

Infants were tested individually and sat on a parent's lap 95 cm from the computer monitor. The parent was instructed not to interact with the infant or watch the monitor. The stimuli appeared in a continuous stream of randomly-ordered pairs (e.g. pair 1: turquoise square followed by blue cross; pair 2: yellow circle followed by pink diamond; pair 3: green triangle followed by red octagon; see Fig. 1), with only transitional probabilities defining between-stimulus boundaries (the transitional probability within pairs was 1.0 and between pairs it was 0.33). Shape pairing was randomized by the computer for each infant. The initial member of a shape pair always predicted the next member, and the next stimulus after a pair was constrained to be the initial member of one of the three allowable pairs. For an individual infant, the pairs were always the same, but the order of the pairs within the sequence was random. The infants were habituated to this sequence until habituation of looking occurred or 12 trials had elapsed. The habituation criterion was defined as a decline in looking times across a block of four trials adding up to less than 50% of looking times during the first four trials. After habituation, infants viewed six test displays alternating between familiar sequences, composed of the same three pairs of shapes, and novel sequences, produced by randomly ordering the same shapes. In the novel sequences, the single constraint on stimulus order was that there were never two identical stimuli in a row. The only difference between familiar and novel sequences was the transitional probabilities between the shapes. This ensured that any looking time difference observed would necessarily be related to the statistical structure of the

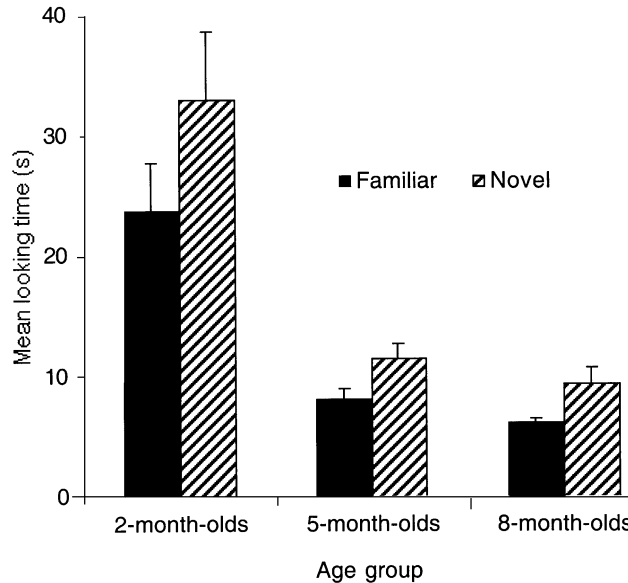


Fig. 2. Looking times after habituation. Infants at all three ages looked reliably longer at the novel sequence, relative to the familiar sequence.

sequence. Ordering of test trials was counterbalanced across infants so that half the infants saw a familiar trial first and half the infants saw a novel trial first.

3. Results

Infants in all three age groups exhibited longer looking at the novel sequence, relative to the familiar sequence, after habituation (see Fig. 2). Twelve of the 16 2-month-olds showed this preference (Wilcoxon matched pairs test, $z = 2.38$, $P < 0.05$), as did 11 of the 16 5-month-olds ($z = 2.33$, $P < 0.05$) and 12 of the 16 8-month-olds ($z = 2.02$, $P < 0.05$). (The Wilcoxon statistic takes account of the magnitude of differences in performance, and is therefore a more sensitive test of infants' preferences than, say, a sign test.) These conclusions were confirmed with parametric analyses. Looking time data in some cells were positively skewed (which is often the case in visual habituation procedures with young infants), and all data were therefore log-transformed prior to analysis; data shown in Fig. 2 are based on raw scores. A 2 (sex: male vs. female) \times 3 (age: 2, 5, or 8 months) \times 2 (order: familiar vs. novel sequence seen first after habituation) \times 2 (test display: familiar vs. novel sequence) mixed ANOVA yielded a significant main effect of age ($F(2, 36) = 21.10$, $P < 0.001$), the result of longer looking overall by the youngest infants. (Very young infants typically exhibit longer looking times than older infants in visual tasks, which may reflect developmental differences in infants' basic information processing skills (see Johnson, 1996).) There was also a significant main effect of test display ($F(1, 36) = 14.67$, $P < 0.001$), the result of longer looking overall at

the novel sequence. There were no other significant main effects or interactions. Planned comparisons (simple effects tests) revealed a reliable preference for the novel sequence in each age group (2-month-olds, $F(1, 36) = 4.30$, $P < 0.05$; 5-month-olds, $F(1, 36) = 7.00$, $P < 0.05$; 8-month-olds, $F(1, 36) = 4.15$, $P < 0.05$). (Simple effects tests take into account the omnibus error term of the ANOVA, and are the functional equivalent of a series of more conservative t -tests to examine preferences in each age group.)

4. Discussion

Infants at 2, 5, and 8 months of age were familiarized with a series of discrete visual stimuli whose ordering was defined solely by statistical regularities, and subsequently demonstrated a reliable preference for novel sequences whose ordering violated the transitional probability that defined grouping of the original stimuli. There were no significant differences in performance between the age groups; 2-month-olds demonstrated the same preference for the novel random sequences as 5- and 8-month-olds. The only cue to stimulus sequence lay in its statistical nature: there were no pauses between pairings, each discrete stimulus was the same duration, and there was no a priori relation between stimuli to provide further information for co-occurrence. Infants' performance was likely based on a true sensitivity to the transitional probabilities defining shape pairs rather than to the frequency of occurrence of these pairs. Aslin, Slemmer, Kirkham, and Johnson (2001) recently conducted a "frequency balanced" version of the present study to eliminate such an explanation (see Aslin et al., 1998 for further description of the frequency balanced paradigm), in which 8-month-olds distinguished between legal and illegal pairs that contained the same frequency of occurrence. It is possible, of course, that infants picked up on the higher order probabilities inherent in the paired stimuli sequences (i.e. 1st, 2nd, 3rd, ..., n th order conditional probabilities). In addition, given that our paradigm is markedly different from that of Saffran et al. (1996), it is possible that the infants responded primarily to the existence of alternations in the sequences (e.g. 33% of the time a pair will repeat). However, the most parsimonious explanation would simply require tracking of the highest sets of probabilities, those between pairs of colored shapes.

These results are consistent with the existence of a domain general statistical learning device that is available to even very young infants (indeed, we found no evidence that the older infants were better able to compute the statistical structure in the input than the youngest infants we tested). Given the youngest age tested in addition to the lack of observed development, it seems reasonable to posit an associative mechanism that is functional with the onset of visual experience. This statistical learning mechanism is powerful enough to ascertain visual input structure after only a few minutes of exposure in a highly constrained, unnatural setting. Moreover, it appears to operate outside other potential contributions to learning in infancy, such as reinforcement, that would normally supplement patterns of input in the environment. In everyday situations, infants would presumably benefit from other kinds of structure in the course of cognitive development, such as intermodal information (e.g. the consistent pairing of certain sights and sounds), as well as direct instruction and other social interactions.

At present we have no information concerning how the infants computed the statistics of

the stimulus sequences. Mechanisms to accomplish such computations may be rather primitive, and can be instantiated in simple recurrent networks (e.g. Cleeremans & McClelland, 1991). However, not all sequence learning can be construed as primitive. Studies of implicit learning in adults, for example, reveal the existence of mechanisms that detect input structure of complex sequences known as artificial grammars, as indicated by accelerated response times upon repeated exposure (Reber, 1989). In these tasks, learners are often unable to report the rules underlying the sequence. Nevertheless, adults are remarkably facile at such tasks (see Stadler & Frensch, 1998 for review), and evidence is beginning to emerge that children also can learn intricate stimulus sequences without explicit knowledge of the statistical structure of the input (Meulemans, Van der Linden, & Perruchet, 1998), and even without attending to the stimuli (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Moreover, 12-month-old infants have been found to generalize their knowledge of complex patterns in artificial grammars (Gomez & Gerken, 1999; cf. Marcus, Vijayan, Bandi Rao, & Vishton, 1999). We do not know whether the infants we observed learned the sequences implicitly or explicitly, or even if such distinctions are applicable to such a young population. In spite of these remaining questions, documentation of statistical learning in young infants contributes an important piece of information to our knowledge base concerning human learning, in addition to implicit and rule learning in children and adults.

Obviously, statistical learning is unlikely to account for all aspects of cognitive development in humans. There are constraints on what can be learned, and who can learn it: only humans show the full range of language acquisition and production skills, for example, despite some cross-species commonalities in mechanisms for word segmentation (Hauser et al., 2001). Nevertheless, our findings, in conjunction with those of Gomez and Gerken (1999) and Saffran et al. (1996, 1999), are consistent with the thesis that early development is highly attuned to the multifaceted structure of the infant's environment, and suggest that learning the statistical regularities of the environment may be a critical part of the cognitive apparatus with which infants make sense of the world.

Acknowledgements

This research was supported by NSF grant BCS-0094814 to S.P.J. The authors are indebted to Richard Aslin for invaluable assistance with all phases of this project. We also thank the undergraduates in the Cornell Baby Lab for assistance with recruitment of our participants, and Morten Chistiansen, Shimon Edelman, David Field, Kerri Lawson Johnson, and Michael Spivey for comments on an earlier draft of this paper. We are especially grateful for the contributions of the infants and parents who participated in the studies. The stimuli used in the experiment can be viewed at <http://babylab.psych.-cornell.edu/research/statlearning>.

References

- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321–324.

- Aslin, R. N., Slemmer, J. A., Kirkham, N. Z., & Johnson, S. P. (2001, April). *Statistical learning of visual shape sequences*. Paper presented at the meeting of the Society for Research in Child Development, Minneapolis, MN.
- Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: a methodological overview* (pp. 253–300). Norwood, NJ: Ablex.
- Carey, S. (1999). Sources of conceptual change. In E. K. Scholnick, K. Nelson, S. A. Gelman & P. H. Miller (Eds.), *Conceptual development: Piaget's legacy* (pp. 293–326). Mahwah, NJ: Erlbaum.
- Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235–253.
- Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, *6*, 559–601.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Gleitman, L. R., & Wanner, E. (1982). Language acquisition: the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: the state of the art* (pp. 3–48). Cambridge: Cambridge University Press.
- Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–135.
- Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*, B53–B64.
- Johnson, S. P. (1996). Habituation patterns and object perception in young infants. *Journal of Reproductive and Infant Psychology*, *14*, 207–218.
- Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77–80.
- Meulemans, T., Van der Linden, M., & Perrouchet, P. (1998). Implicit learning in children. *Journal of Experimental Child Psychology*, *69*, 199–221.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: MIT Press.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101–105.
- Slater, A. (1995). Visual perception and memory at birth. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (pp. 107–162). Vol. 9. Norwood, NJ: Ablex.
- Spelke, E. S., & Van de Walle, G. (1993). Perceiving and reasoning about objects: insights from infants. In N. Eilan, R. A. McCarthy & B. Brewer (Eds.), *Spatial representation: problems in philosophy and psychology* (pp. 132–161). Oxford: Blackwell.
- Stadler, M. A. & Frensch, P. A. (1998). Handbook of implicit learning Thousand Oaks, CA: Sage.
- Wynn, K. (1998). Psychological foundations of number: numerical competence in human infants. *Trends in Cognitive Sciences*, *2*, 296–303.