Zen in the Art of Language Acquisition: Statistical Learning and the Less is More Hypothesis

David Ludden (david-ludden@uiowa.edu)
Department of Psychology, University of Iowa
Iowa City, IA 52242 USA

Prahlad Gupta (prahlad-gupta@uiowa.edu)
Department of Psychology, University of Iowa
Iowa City, IA 52242 USA

Abstract

It seems an obvious truth that children are better language learners than adults. Children seem able to master a second language with ease, while adults are rarely successful at second language acquisition. Newport's (1990) Less is More hypothesis represented an attempt to explain these observations by invoking general cognitive mechanisms. This hypothesis takes as its starting point the observation that children exhibit reduced working memory capacity relative to adults and suggests that this reduction serves as a filter to aid children in deducing the structure of the language they are learning. We present two experiments testing a specific prediction that follows from the Less is More hypothesis, namely that adults will perform better on language learning tasks if their available working memory capacity is reduced. The experiments examined the learning of word boundaries and syntactic agreement, each with and without a concurrent cognitive load. The results of these experiments were contrary to the Less is More prediction, suggesting that other explanations must be found for the observed superior language learning performance of children over adults.

Introduction

In *Zen in the Art of Archery* (Herrigel, 1953), the German philosopher Eugen Herrigel recounts his endeavor to learn Japanese archery from a great Zen master. During his many years of training, Herrigel struggles with his need to control the bow, to think about the target. The more he tries to control the bow, the Master tells him, the less control he will have over it; the more he thinks about the target, the less accurate his shots will be. Rather, the Master tells Herrigel to strive for less control, less conscious thought of the act of releasing the arrow from the bow, for only then will he gain mastery over the bow. The idea that Less is More is a dominant theme of Zen Buddhism. However, this notion is not limited to the realm of Oriental philosophy; it is one that pops up again in Western developmental psychology, in particular, in the area of language acquisition.

It is generally acknowledged that children's cognitive abilities, working memory in particular, are considerably constrained relative to those of adults (Gathercole, Willis, Baddeley & Emslie, 1994). At the same time, it has also been widely observed that children are much more successful at learning language than adults. After all, virtually all children learn a first language, while few adults ever master a second. Furthermore, children seem to learn language with such effortlessness, as opposed to the great expense of effort necessary for an adult to acquire even marginal proficiency in a second language. To Newport (1990), this seemed a perfect example of a Less is More situation: children have less cognitive ability yet learn language more easily, while adults have more cognitive ability yet learn language less easily. This observation led to the formulation of the Less is More hypothesis (Newport, 1990), which not only maintained that the restricted working memory of children was an advantage to language acquisition, but also attempted to explain the causal nature of that relationship.

The idea for the Less is More hypothesis came out of studies on critical periods in first and second language acquisition performed by Newport and her colleagues. For example, Johnson & Newport (1989) conducted a study of 46 native Chinese and Korean speakers living in the United States who had learned English as a second language. The participants were divided into two groups, those who had arrived in the US before age 15 (dubbed Early Arrivals), and those who had arrived in the US after age 17 (dubbed Late Arrivals). All participants had spent at least the last three years prior to the experiment in the US. Johnson & Newport (1989) found an inverse linear relationship between age of arrival in the US and ultimate performance in English. In that study, only those participants who had arrived in the US by the age of seven achieved native speaker proficiency as measured on a grammaticality judgment task. Beyond age seven, performance on this task decreased as age or arrival increased. Johnson & Newport (1989) also examined attitudinal variables, but found statistical significance for the age of arrival variable over and above and other variables they looked at.

In a subsequent study, Newport (1990) examined critical periods in first language acquisition, in particular in American Sign Language (ASL). She studied three groups of congenitally or pre-lingually deaf adults who used ASL as their predominant language and had limited skills in English. The first group, dubbed Native Learners, had been exposed to

ASL since birth or shortly thereafter. The second group, dubbed Early Learners, had been first exposed to ASL between the ages of 4 and 6. The last group, dubbed Late Learners, had been exposed to ASL only after age 12. All participants were tested on their knowledge of ASL sentence structure and morphology. While all three groups performed at ceiling on the sentence structure test, there was a significant decline in performance in morphology across the three groups from Native to Early to Late Learners.

Newport (1990) took these findings as clear evidence for a critical period in language acquisition, and in an attempt to explain the mechanism responsible for this critical period she posited the Less is More hypothesis. The hypothesis takes as its starting point the notion that the working memory capacity of children is limited compared to that of adults. The hypothesis then proposes that this limitation is actually advantageous. According to Newport (1990), language acquisition requires a componential analysis. Adults take in too much of the language input at one time because of their expanded working memory capacity relative to children. This wider perceptual window in adults leads to a combinatorial explosion of possible analyses for the language input, hence the likelihood of hitting upon the right analysis is small. Children, with their limited working memories, are constrained by the size of the input they do take in to a more limited number of possible analyses. Hence, the likelihood of hitting upon the right analysis is greatly increased. According to Newport (1990), it is the limitations of the child's ability to process information that provides the basis for successful language acquisition.

Elman (1993) tested the Less is More hypothesis with a connectionist model of syntactic agreement acquisition. He trained a simple recurrent network (Elman 1990) on a corpus of sentences based on a simplified English grammar. In this grammar, subjects and verbs agreed in number, verbs differed in argument expectations, and sentences could contain multiple embeddings. The corpus contained sentences such as *cats chase dogs* and *dogs see boys who cats who mary feeds chase*. The context units of this simple recurrent network represented the working memory of the network, and the capacity of this working memory was a parameter that could be varied. When the network was trained on the entire corpus with working memory at full capacity, the network failed to learn.

Elman (1993) then tried incrementing the capacity of the working memory of the network. Working memory capacity was manipulated by an automatic flushing of the context units after every three or four words. As training progressed, the interval between flushings was gradually increased. The result was that the network was then able to learn how to process the input. Elman (1993) interpreted this finding as consistent with the Less is More hypothesis.

Relevant empirical data came from a study by Santelmann & Jusczyk (1998), who used a headturning paradigm with 15- and 18-month-old infants to test their sensitivity to morphosyntactic dependencies in English. The experimental condition consisted of well-formed English sentences with the structure *...is...*<verb>*ing*, while the control condition consisted of ill-formed sentences with the structure *...can...*<verb>*ing*, such as *Everybody is baking* vs. **everybody can baking.* Santelmann & Jusczyk (1998) also varied the distance in syllables between auxiliary verb (*is* or *can*) and main verb by the insertion of adverbs, as in *Everybody is often baking*. They found that at distances of 1-3 syllables, 18-month-old infants preferred well-formed over ill-formed sentences. However, at distances over 3 syllables, the 18-month-olds showed no preference for either form, nor did the 15-month-olds at any distance. Santelmann & Jusczyk (1998) concluded that their findings were "consistent with the hypothesis that 18-month-olds are working with a limited processing window, and that they are only picking up relevant dependencies that fall within this window." Although the authors found no evidence to determine whether these limitations in processing space facilitated or hampered language acquisition, the Santelmann & Jusczyk (1998) study nevertheless does lend support to a basic premise of the Less is More hypothesis, namely that infants are processing the language input in shorter chunks than adults are, justifying the approach Elman (1993) took in modeling the syntax-acquisition process.

Work on statistical learning by Saffran and her colleagues has also been relevant to the Less is More hypothesis. Saffran, Newport & Aslin (1996a) asked adult participants to listen to a nonsense language that contained words but no meanings or grammar. The task was to try to figure out where the word boundaries were. At the end of 21 minutes of exposure, the participants were asked to choose which of two items sounded more like a word from that language. The participants performed significantly above chance, with a mean score of 76% (chance was 50%). This type of exposure condition was referred to as the explicit learning condition in this and later Saffran studies.

Saffran et al. (1997) tested the learning of word boundaries in an incidental learning condition. In this condition, participants were asked to draw a picture while the stimulus played in the background. Subjects were told nothing about the stimulus. After 21 minutes of exposure, the participants were administered the same test as in the explicit condition. Saffran et al. (1997) tested two groups, adults (college students) and children (6-7 years old). Mean percent correct identification scores for each group were significantly above chance (50%) at around 59%, with no significant difference between children and adults. Because of the low scores after one exposure period, the experiment was redone with two exposure periods on consecutive days. In this second experiment, adults averaged 73% and children 68%, with the difference between adults and children being nonsignificant. Saffran et al. (1997) concluded that passive exposure was

sufficient at least for some aspects of the language acquisition process.

In her dissertation, Saffran (1997) extended her research in statistical learning to the acquisition of syntax, in particular, hierarchical phrase structure. The stimulus set in each experiment consisted of a sample of sentences from an artificial language, with the only cues to syntactic structure being statistical. In an explicit learning condition, the participants were exposed to the stimulus for 30 minutes a day for two days, and tested on their knowledge of the phrase structure at the end of each listening period. Mean adult performance in this explicit learning task was 68%[1]. (No children were run in this condition.) In an incidental learning condition in which participants listened to the stimulus while drawing a picture, both adults and children (aged 6-9) showed performance significantly above chance after the first exposure period, with no significant improvement after the second session.[2] Children's scores (57%) were significantly worse than those of the adults (67%). There was no significant difference between adults in the explicit and incidental conditions.

The results of these various experiments by Saffran and her colleagues seem inconsistent with the Less is More hypothesis. Specifically, the Less is More hypothesis predicts that children will perform better than adults in language learning tasks, and furthermore that adults will perform better in an incidental learning task than an explicit one. But these predictions are belied by the data. Not only was there no significant difference in performance between children and adults in the incidental word boundary learning task (Saffran et al. 1997), children in fact fared worse than adults in the implicit syntax learning task (Saffran, 1997). Furthermore, there was no significant difference between explicit and incidental conditions in adult performance on the syntax learning task (Saffran 1997).

The various results described above paint an inconsistent picture of the impact of working memory resources in language learning. The studies by Elman (1993) and Santelmann & Juszczyk (1998) appear to support the Less is More hypothesis, while the results of the studies by Saffran et al. (1996) and Saffran et al. (1997) are inconsistent with that hypothesis.

The experiments described below were aimed at examining the following question: Is adult performance on a language learning task superior when working memory resources are reduced, as the Less is More hypothesis would predict? Although the results of Saffran et al. (1997) are inconsistent in this regard, they are difficult to interpret because they were obtained under different experimental conditions. The present experiments attempt to address this question systematically. Experiment I addresses this question in the domain of word boundary learning while experiment II addresses this question in the domain of syntax learning.

[1] This and the following three composite scores were calculated from the data in Saffran (1997).

[2] Saffran (1997) acknowledges that this incidental task was not as incidental as it was in the word boundary experiments. In the incidental learning condition of the phrase structure experiment, participants were told about the nature of the background stimulus and the test they would be given at the end of the drawing period.

Experiment I: Word Boundaries

Because the exposure periods in Saffran et al.'s explicit (1996a) and incidental (1997) word boundary learning tasks were not equivalent, a direct comparison cannot be made. Experiment I represents an attempt to replicate these two experiments under identical exposure conditions. To insure this, the difference between these two tasks was reduced to the presence or absence of a concurrent cognitive load (drawing a picture) during the exposure to the stimulus. For this reason, in this and the following experiment, Saffran et al.'s (1996a) explicit condition is referred to by the more theory-neutral term No Load, while Saffran et al.'s (1997) incidental condition is referred to as the Load condition. If the Less is More hypothesis is true, then we would expect superior performance in the Load vs. the No Load condition.

Method

Thirty-two participants were recruited for the experiment from the University of Iowa Psychology Department subject pool. The participants received partial credit toward fulfilling requirements for an introductory psychology course. These participants were randomly assigned to two groups of 16 each, constituting the No Load and Load groups for this experiment. The exposure and test materials were reconstructed per the specifications given in Saffran et al. (1997).

In the No Load condition, participants were informed that they would be listening to an artificial language that consisted of a small number of words but no meanings or grammar. They were not told the exact number of words or anything about the structure of those words. The participants were asked to listen to the language and try to figure out where the word boundaries were. They were also told that they would be tested on their knowledge of the word boundaries later in the experiment. These instructions were made as similar as possible to those given in Saffran et al. (1996).

In the Load condition, participants were asked to draw a picture using a computer drawing program. They were informed that an auditory stimulus would play while they drew, and that the experimenter was looking at a certain effect that would be explained to them later in the experiment. The participants were told nothing at the outset of the experiment about the content of the stimulus, nor were they informed that they would be given a test based on the auditory stimulus later on. These instructions were made as similar as possible to those given in Saffran et al. (1997).

The exposure procedure was identical for both the No Load and the Load groups, and consisted of three seven-minute listening periods with five-minute breaks between. This exposure procedure is the same as that used in Saffran et al.

(1996). After the three listening sessions were finished, the experiment proceeded to the test phase, in which the participants were asked to listen to each pair of sound items and to decide which of the two sounded more like it came from the stimulus.

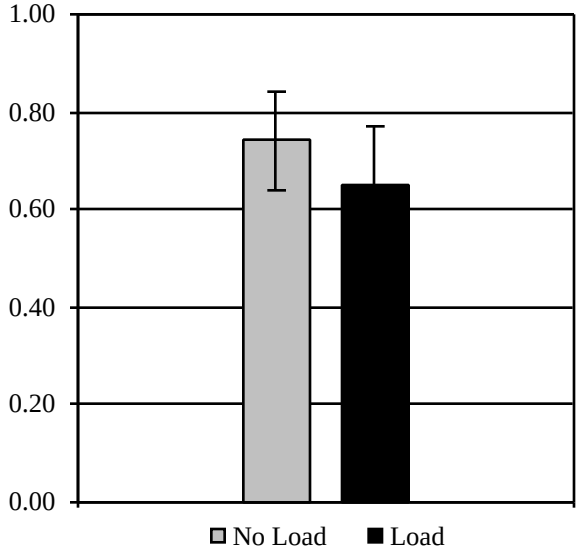

Figure 1: Results of Experiment I, Statistical Learning of Word Boundaries (Mean percent correct on vertical axis).

Results and Discussion

The results of both groups are in line with the results of the word boundary learning experiments of Saffran et al. (1996) and Saffran et al. (1997), and are shown in Figure 1. The mean score for the No Load group was 74%. A single-sample *t* test (two-tailed) showed that performance was significantly above chance, $t\ (15) = 9.94$, $p < .01$. The mean score for the Load group was 65%, which was significantly above chance as well, $t\ (15) = 5.20$, $p < .01$. A two-sample *t* test of the No Load vs. Load means was significant ($p < .03$), indicating superior performance on the part of the No Load group over the Load group.

In the present experiment, both groups exhibited learning. However, the Load group did not outperform the No Load group, contrary to the prediction of the Less is More hypothesis. Rather, this finding is consistent with a More is More hypothesis, that is, with the idea that enhancing cognitive resources enhances cognitive performance.

Experiment II: Syntactic Agreement

While the results of Experiment I were inconsistent with the Less is More hypothesis, it could be argued that segmenting words from a stream does not involve the sort of componential analysis that Newport (1990) considered necessary for successful language acquisition. Experiment II consists of a task that would require such a componential analysis: the learning of the pattern of syntactic agreement in an artificial language. The Less is More hypothesis predicts that participants with a reduced working memory capacity (experimentally induced by the imposition of a cognitive load) will perform better than participants with no reduction of working memory.

Method

Thirty-two adult college students were recruited from the subject pool of the University of Iowa Psychology Department and randomly assigned to two groups (No Load or Load) as in Experiment I.

An artificial language with a small vocabulary and a simple grammar was created for this experiment. The vocabulary of this language consisted of twenty one-syllable noun roots and twenty one-syllable verb roots. The grammar consisted of two rules. First, all sentences were two words in length, each composed of a noun followed by a verb. Second, the noun and verb of each sentence agreed in number, with singular nouns marked with the suffix *-bo*, plural nouns with *-za*, singular verbs with *-ki*, and plural verbs with *-nu*. Thus, the noun *da* and the verb *me* could form both the sentence *da-bo me-ki* (singular) and the sentence *da-za me-nu* (plural). The exposure and test corpuses were set up such that all the words in the test phase had been heard in the exposure phase, but that all of the test sentences were new.

For the sake of comparison across Experiments, the instructions and procedures in Experiment II were made as parallel as possible to those used in Experiment I.

Subjects in the No Load condition were told that they would hear an artificial language consisting of a series of two-word sentences. They were told nothing about the number of different words or the length of the words. The participants were told that this language was spoken by a computer speech-synthesis program that did not put pauses between words or sentences. Their task, then, was to listen to the language and see if they could figure out where the sentence breaks were supposed to be. They were also told that they would be tested on their ability to find the sentence breaks at the end of the experiment. The rationale for giving the participants this task during exposure was twofold: One was to keep the participants focused on the stimulus, and the other was to keep the procedures in Experiment II as parallel as possible to those in Experiment I. During the test phase, the participants were told to listen to each pair of sound items and decide which of the two sounded more like the training stimulus.

Subjects in the Load condition were given a drawing task and cover story as in Experiment I. During the test phase, they were asked to decide which of the two items in each trial sounded more like the stimulus that played while they were drawing. However, at no time were they told about the content of the stimulus.

Results and Discussion

The results of Experiment II are shown in Figure 2. Mean performance of the No Load group was 56%. A single-sample

t test showed that performance was significantly above chance, t (15) = 4.57, $p < .01$. Mean performance of the Load group was 52%. This performance was not significantly above chance, t (15) = 1.23, n.s. A two-sample *t* test comparing No Load vs. Load means, however, was significant, $p < .05$.

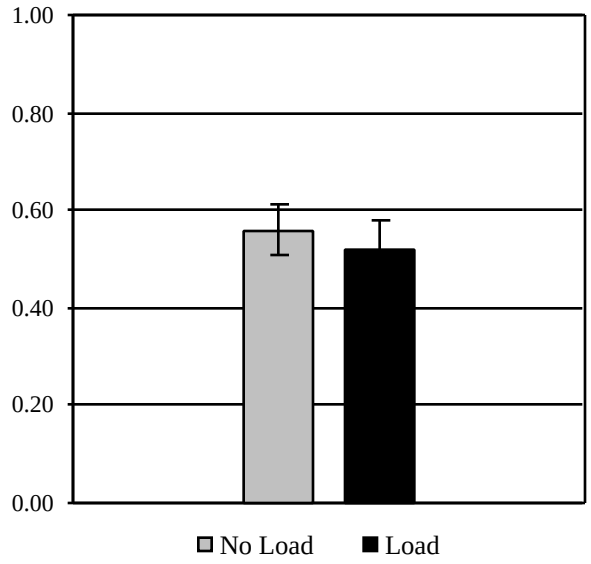

Figure 2: Results of Experiment II, Statistical Learning of Syntactic Agreement (Mean percent correct on vertical axis).

Contrary to the predictions of the Less is More hypothesis, the No Load group did not outperform the Load group; on the contrary, the Load group significantly outperformed the No Load group. However, the null result in the Load condition makes the findings of this experiment hard to interpret. It is not clear whether the Load group failed to learn due to the inherent difficulty of the task, or if they would have exhibited learning had they been given a longer exposure period. When Saffran et al. (1997) increased exposure in their incidental word boundary learning task to two 21-minute sessions on consecutive days, the participants' performance improved significantly. However, the participants in Saffran's (1997) phrase-structure learning experiments showed no significant improvement from Day 1 to Day 2 in either the explicit (No Load) (Saffran, 1997) or incidental (Load) (Saffran, 1997) conditions. At the very least, the results of the present experiment suggest that increased working memory capacity leads to better performance. There could also be a role for attention in the acquisition of syntax, as the null result in the Load condition suggests that syntax may not be learnable at all without attention.

Results across Experiments I and II were analyzed in a two-way ANOVA of task (Word Boundaries vs. Syntactic Agreement) by condition (No Load vs. Load). The results, as shown in Figure 3, indicate main effects for both task and load, but no interaction. In other words, performance in the Word Boundary task was significantly better than in the Syntactic Agreement task, regardless of condition. Likewise, performance in the Load condition was significantly worse than performance in the No Load condition, regardless of task. The results of this ANOVA suggest, first of all, that the syntactic agreement task was inherently more difficult than the word boundary task was. In addition, they suggest that the imposition of a cognitive load leads to reduced performance in either of these tasks, a finding that runs counter to the predictions of the Less is More hypothesis.

General Discussion

Experiment I clearly indicates that, at least with regard to the segmentation of words in a speech stream based on statistical regularities, the Less is More hypothesis does not hold. Under that hypothesis, we would expect to see better performance on the part of the Load participants. What we see instead is significantly better performance on the part of the No Load participants, exactly the opposite of what we would expect if the Less is More hypothesis were true. The same pattern of results obtained in Experiment II, suggesting that the Less is More hypothesis does not hold in the domain of syntax acquisition, either.

A key issue in the Less is More hypothesis is the issue of the role of working memory in language acquisition. The Less is More hypothesis posits that the restricted working memory in children aids them in language learning, and furthermore maintains that the larger working memory capacity of adults hinders their language learning ability. An alternative to the Less is More hypothesis would be a More-is-More hypothesis predicting that the greater the cognitive resources available, the better the language learning (or any other) performance will be. The data presented in this paper are consistent with that view.

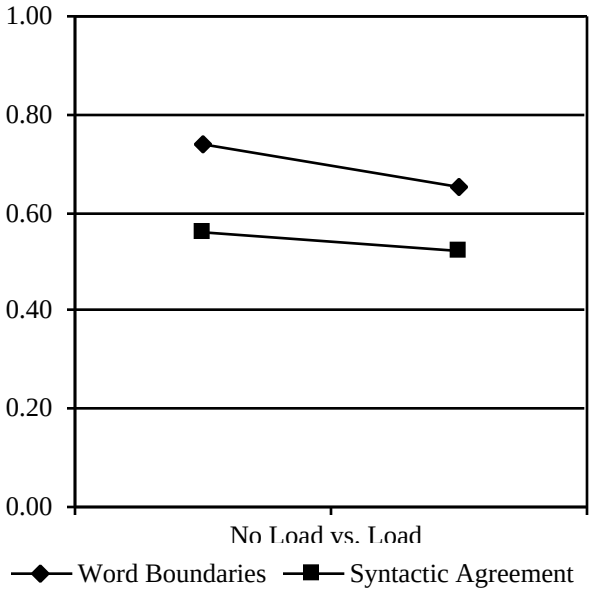

Figure 3: Result of ANOVA across Experiments I and II.

Even if it were the case that the lower performance in the No Load condition were in part because participants were not attending to the stimulus (and not merely because of working memory limitations imposed), this would still in some sense represent a reduction of available cognitive resources and thus, according to the Less is More hypothesis, should still result in better performance. We recognize that there is a potential confound in the two experiments presented here between the effect of the manipulation on working memory and on attention. Research currently underway in our lab is testing the separate effects of working memory and attention on the acquisition of syntax.

We would also like to address three concerns with the present research that have been brought to our attention. The first concern is that Experiment I may not be relevant to the Less is More hypothesis. However, a similar result in both experiments suggests similar mechanisms at work in both the word boundary learning and syntactic agreement learning tasks, making Experiment I relevant to our argument. The second concern is that the dependencies were too close, creating a situation in which the Less is More hypothesis would predict no advantage for a limited working memory capacity. But if this were the case, then we would expect no difference in performance between the Load and No Load groups. Rather, we see that even when dependencies are only a syllable apart, the Load group performs significantly worse than the No Load group. The third concern is that the low performance in Experiment II may be due to having too many words to learn; in other words, syntax acquisition was hindered by the demands of vocabulary acquisition. However, it is not at all clear that it is necessary to learn the words in order to learn the syntax. For example, it is doubtful that the infants in Santelman & Jusczyk (1998) knew all the words in the sentences they heard; nevertheless, they were sensitive to the long-distance dependency being tested for.

To the extent that the Less is More hypothesis is challenged, the question is raised of how to account for the observed critical period effect (Lenneberg 1967) in language acquisition. The Less is More hypothesis makes the implicit assumption that the only relevant difference between children and adults approaching the language learning task is working memory capacity. However, it is very likely that there are other differences, motivational in particular, between the conditions under which children and adults enter a language learning situation besides just working memory capacity (Schuman, 1975, as cited in Johnson & Newport, 1989). In fact, empirical evidence suggests that when motivational factors are held constant in a laboratory situation, children fare worse than adults, as they did, for example, in the experiments reported by Saffran (1997).

Conclusion

After seven years in Japan, Herrigel (1953) finally masters the bow, learning to send the arrow to its target with apparent effortlessness. Yet behind that appearance of ease lies seven years of struggle. Seeming effortlessness is the goal in mastering the bow, not the means to mastering it. Likewise in mastering a language. Facility in a language is achieved only by an arduous, extended process. The language learning process demands a great expense of cognitive effort, and it only stands to reason that the more cognitive resources one has available, the more likely one is to succeed at the task. This premise is borne out by the evidence presented here: adults performed better at language learning tasks when there were no other cognitive demands placed on them. At least for the aspects of language acquisition examined here, it is clear that less is less, not more.

Acknowledgments

We would like to thank Steven Luck and Rochelle Newman for many helpful discussions.

References

Elman, J.L. (1990). Finding structure in time. *Cognitive Science 14,* 179-211.

Elman, J.L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition 48,* 71-99.

Gathercole, S.E., Willis, C.S., Baddeley, A.D. & Emslie, H. (1994). The children's test of nonword repetition: a test of phonological working memory. *Memory 2,* 103-127.

Herrigel, E. (1981). *Zen in the art of archery.* New York: Vintage Books.

Johnson, J.S. & Newport, E.L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology 21,* 60-99.

Lenneberg, E. (1967). *Biological foundations of language.* New York: Wiley.

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science 14,* 11-28.

Saffran, J.R. (1997). *Statistical learning of syntactic structure: mechanisms and constraints.* Doctoral dissertation, Departments of Brain & Cognitive Sciences and Linguistics, University of Rochester, Rochester, NY.

Saffran, J.R., Newport, E.L. & Aslin, R.N. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language, 35,* 606-621.

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.

Santelmann, L.M. & Jusczyk, P.W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition 69,* 105-134.