When Less is Less and When Less is More: Starting Small with Staged Input

Christopher M. Conway (cmc82@cornell.edu)
Department of Psychology; Cornell University; Ithaca, NY 14853, USA
Michelle R. Ellefson (M.Ellefson@warwick.ac.uk)
Department of Psychology; University of Warwick; Coventry CV4 7AL, UK
Morten H. Christiansen (mhc27@cornell.edu)
Department of Psychology; Cornell University; Ithaca, NY 14853, USA

Abstract

It has been suggested that external and/or internal limitations may paradoxically lead to superior learning (i.e., the concepts of "starting small" and "less is more"; Elman, 1993; Newport, 1990). In this paper, we explore what conditions might lead to a starting small effect. We report on four artificial grammar learning experiments with human participants. In Experiment 1A, we found an effect of starting small with visual center-embedded, recursive input staged incrementally. Experiment 1B replicated this finding and extended the effect to right-branching recursive structure. Finally, in Experiments 2A and 2B we found no effect for starting small with auditory center-embedded or right-branching input. These results suggest that starting small can confer a learning advantage but perhaps only under certain conditions.

Introduction

Intuitively, learners should learn better when they are unhindered by internal or external limitations, such as those relating to constraints on memory or on the input. However, recent proposals take the somewhat paradoxical stance that cognitive limitations and/or reduced input may confer a computational advantage for learning. These theories, specifically the notion that "*less is more*" (Newport, 1990) and the importance of "*starting small*" (Elman, 1993), are often couched in terms of language acquisition. For analyses involving componential inputs, such as language, limited processing may be advantageous because it acts as a "filter" to reduce the problem space, making it more manageable.

Unfortunately, the evidence related to starting small is far from conclusive. Though it is true that children learn language better than adults, this may be due to any number of factors. Initially, computational work supported the theory of starting small (e.g., Elman, 1993), but more recent simulations appear to contradict those findings (Rohde & Plaut, 1999; in press). Empirical evidence gathered from human participants has also not resolved the issue. Though some of the data support starting small, (Cochran, McDonald, & Parault, 1999; Kareev, Lieberman, & Lev, 1997; Kersten & Earles, 2001), other data do not (Fletcher, Maybery, & Bennett, 2000; Ludden & Gupta, 2000; Rohde & Plaut, 1999; in press).

This paper attempts to understand under what conditions, if any, starting small might have an effect. First, we discuss the inconclusive evidence for starting small. Second, we discuss recursive grammars and why such structures may provide a suitable testbed. Next, we present data from four artificial grammar learning experiments with human participants. Experiment 1A shows that when visual, center-embedded input is staged in a starting small fashion, participants achieve better learning than when the input is presented non-incrementally. Experiment 1B reveals a similar effect of starting small using right-branching recursive structure. Finally, Experiments 2A and 2B provide a test of starting small in the auditory modality using center-embedded and right-branching input. The results of these last two experiment suggest that under some conditions, starting small may not be beneficial. Together, this evidence suggests new ways to interpret the starting small hypothesis and the conditions under which less is less and less is more.

Evidence Relating to Starting Small

The "less is more" and "starting small" hypotheses can actually be thought of as two related but separate ideas. The ideas are similar in that they propose that processing limitations may confer a learning advantage but they differ in terms of the nature of the limitation itself. One possibility is that the processing limitations arise from internal, cognitive (e.g., memory) constraints. A second possibility is that the processing limitations are external in nature, in the form of staged or incremental input. Here we review data related to these two possibilities, starting with the internal constraint viewpoint.

In the context of language acquisition, Newport (1990) proposed that maturational constraints in the form of cognitive limitations are crucial for allowing language to be learned successfully. Elman (1993) tested this idea by training a simple recurrent network (SRN) to learn aspects of an artificial language. Under normal conditions, the network was unable to learn the sequential regularities of the grammar. But when Elman simulated children's working memory limitations by periodically eliminating the network's access to its prior internal states—and allowing the size of this temporal window to increase over time—the neural network's performance improved. Further support comes from studies with human participants. Cochran, McDonald, and Parault (1999) taught adults portions of a modified version of American Sign Language (ASL). In their first two experiments, they simulated cognitive limitations by supplying a simultaneous capacity-limiting task during training. Cochran et al. found that the participants in the no-load condition displayed more rigid learning and were less adept at using the signs in new contexts. Additionally, Kareev, Lieberman, and Lev (1997) explored the relation between working memory capacity and the detection of correlation. Human participants were tested on their ability to predict the relationship between two binary variables. Participants with lower natural working memory were better at detecting the appropriate correlation and performed better on the task than did high memory capacity participants. This evidence appears to lend direct support to the importance of starting small: in some situations, cognitive limitations appear to confer a learning advantage.

However, there are reasons to be critical of this data. For instance, Rohde and Plaut (1999; in press) conducted neural network simulations that contradicted Elman's (1993) results. Using the same architecture, simulation parameters, and training input, Rohde and Plaut failed to get an advantage for starting small. Rohde and Plaut (in press) also give reasons for questioning Cochran et al.'s (1999) and Kareev et al.'s (1997) conclusions, instead arguing that their data does not support the notion that internal limitations benefit learning. Other studies appear to support this perspective. For example, adult participants in an artificial grammar learning task with a capacity-limiting condition failed to show an effect of starting small (Ludden & Gupta, 2000). Relatedly, children early in development do not surpass more developed children in artificial grammar learning tasks (e.g., Fletcher, Maybery, & Bennett, 2000).

There are fewer experiments testing the external limitation version of starting small. This may be partly because of the widespread belief that the language input that children receive is not substantially different from that of adults. However, as Rohde and Plaut (in press) point out, there is evidence that child-directed speech tends to consist of shorter utterances and less complex sentences than adult-directed speech (e.g., Pine, 1994). Therefore, it may be feasible that starting with simplified input grants a learning advantage in language and other domains.

Elman (1993) provided a test of this version of starting small using neural network simulations. In an incremental input condition, Elman organized the network's input so that it was exposed first only to simple sequences; complex sequences were then introduced to the network gradually. When trained in this way, the networks showed a learning advantage[1].

A recent study with human participants also supports the validity of an external constraints view of starting small (Kersten & Earles, 2001). Adults were exposed to an artificial language consisting of both auditory nonsense sentences and visual, animated events. Some of the participants were exposed to a staged input regiment, in which they received input in three phases: first only single words were presented along with the animated events, then sentences composed of two words, then finally three-word sentences. These participants fared better on tests of their understanding of the language compared to participants who were exposed to a non-staged input presentation. Though Kersten and Earles (2001) view this demonstration as supporting the notion of internal limitations providing a starting small advantage, we agree with Rohde and Plaut (in press) that this conclusion may not be warranted. Instead, we view this data as showing the possible benefits of using a staged input training scheme.

In closing, we note three crucial observations. First, the study by Kersten and Earles (2001), though it may not be an accurate depiction of children's language acquisition, does suggest that staged input may confer a learning advantage. Second, we observe that most of the "successful" tests of starting small have incorporated visual input (e.g., Cochran et al., 1999; Kareev et al., 1997; Kersten & Earles, 2001) while most of the evidence refuting starting small has relied on auditory input (e.g., Ludden & Gupta, 2000). Third, the input structures that have been used to date in tests of starting small have been relatively simple. However, people are able to learn structures of greater complexity, such as those found in recursion. It is possible that in these more complex learning situations, the effect of starting small may be more pronounced. Based on these three observations, we explore starting small using a staged input scheme, examining both visual and auditory learning, and using input that is recursively-structured.

Recursive Artificial Grammars

A recursive, grammatical construction is one that is defined by self-reference. Different types of recursion can be found across a variety of linguistic structures. As the amount of self-referencing increases within a recursive construction, the amount of embedding increases. Consider these grammatical English noun-phrases.

a) *The dog* [*on the sidewalk*].
b) *The dog* [*on the sidewalk*] [*near the tree*].

---

[1] However, it should be noted that Rohde and Plaut's (1999; in press) simulations appear to contradict this finding.

c) *The dog* [*on the sidewalk*] [*near the tree*] [*by the house*].

The above sentences involve *right-branching* recursion, in which new prepositional phrases are recursively added onto the right end, creating sentences of potentially infinite length. Sentence (a) comprises 0 levels of embedding, (b), 1 level of embedding, and (c), 2 levels of embedding.

Increased levels of embedding result in slightly decreased comprehensibility of English right-branching recursive sentences. Decreases in comprehension are even larger for a second type of recursive structure: *center-embedding* (e.g., Bach, Brown, & Marslen-Wilson, 1986). Center-embedded recursion grows a sequence by embedding new material in the center. For example, consider:

d) *The boy likes the dog.*

e) *The boy* [*the girl loves*] *likes the dog.*

f) *The boy* [*the girl* [*the woman* [*the man adores*] *admires*] *loves*] *likes the dog.*

Sentence (d) is easy to understand, (e) is harder, but (f) is almost impossible to comprehend.

The difficulty of comprehending and producing deeply center-embedded constructions is well documented (e.g., Bach, et al., 1986). English speakers rarely include them in written or spoken language, despite their conformance to the formal grammatical rules of English. Center-embedded recursive structures might be difficult to comprehend because of the need to learn relationships between non-adjacent elements. With greater levels of embedding, memory is taxed, which may hinder comprehension and learning.

Here, we explore the possibility that starting small may facilitate learning of recursive constructions by focusing learners' attention on the relationship between elements (e.g., as in the number agreement relationship between nouns and verbs). Once this relationship has been learned for simple constructions it can then be generalized to more complex constructions. Thus the purpose of this study was to examine the relative usefulness of starting small when learning recursive structure across two different modalities: vision and audition.

Experiments 1A and 1B: Visual Learning of Recursive Structure

In the first set of experiments, we generated visual stimuli from an artificial grammar having center-embedded (Experiment 1A) or right-branching (Experiment 1B) recursion. Besides the difference in the recursive structures, the two experiments were identical. For each experiment, we created two separate training conditions. In the starting small condition, participants were exposed to three training phases. In the first phase, the input was composed of sentences with 0-level center-embedding. The second phase incorporated sentences with 1-level embedding, while the third phase used sentences with 2-level embedding. In this way, the input "started small" and progressively became more complex.

In the second training condition, participants received the same input though in random order. Experiment 1A also contained a third experimental condition receiving no input; rather, these control participants took part in the testing phase only. Experiment 1B did not contain a control group as experiment 1A indicated that such a control was not necessary. The starting small theory predicts that the starting small input group will learn the recursive center-embedded and right-branching structure better than the corresponding random group.

Method

Subjects For Experiment 1A, twenty-four undergraduate subjects (eight in each condition) were recruited from Psychology classes at Cornell University, earning extra credit. For Experiment 1B, fourteen new subjects (seven in each condition) were recruited in an identical manner.

Materials For both experiments, the stimuli were letter sequences generated from the same artificial grammar as Ellefson (2002). The sequences were based on the repetition of noun-verb pairs within a recursive structure, in which arbitrary letters designated plural and singular nouns and verbs. As with English sentences, nouns and verbs were paired with each other according to grammatical number. For example, the singular noun *cat* might be paired with the singular verb *plays*, but it would not be paired with the plural verb *play*. Likewise, the plural noun *cats* would be paired with *play* but not *plays*. In our experiment, the letter S falls in the plural noun category and the letter T falls in the plural verb category. Therefore, S and T comprised a possible plural noun-verb pair. Likewise, the letters M and Z comprise a possible singular noun-verb letter pair.

Twelve consonants, C, Q, M, P, X, S, W, Z, K, H, T, and L represented the singular and plural nouns and verbs. There were three letters assigned to each of the letter roles of singular noun, plural noun, singular verb, and plural verb[2]. The sequences contained 0-, 1-, and 2-level embeddings. For Experiment 1A, embedding was increased by inserting additional noun-verb pairs into the middle of the center-embedded sequences to achieve higher levels of embedding. An example of a 0-level center-embedded sequence is CW, a 1-level sequence is CPTW, and a 2-level sequence is CPQMTW. For Experiment 1B, embedding was increased by adding new noun-verb pairs to the end of a sequence. For example, CW is a 0-level right-branching sequence, CWPT is a 1-level sequence, and CWPTQM is a

---

[2] Singular nouns: C, Q, and M. Plural nouns: P, X, and S. Singular verbs: W, Z, and K. Plural verbs: H, T, and V.

2-level sequence.

In Experiment 1A, unique sequences were created for the training and testing sessions. Fifty sequences comprised the training session. Of these 50 training sequences, 10 were 0- level embedding, 20 were 1-level embedding, and 20 were 2-level embedding. An additional fifty sequences comprised the test session. Of these testing sequences, 25 were generated from the same grammar as the training sequences (grammatical) and 25 did not follow the grammar (ungrammatical). Ungrammatical sequences were created by changing one letter of a grammatical test sequence. The substituted letter was one that was of the proper noun-verb category but with an incorrect grammatical number. The positions in which the substituted letters occurred in the sequences were distributed evenly across all items. The test session comprised 16 sequences of 0-level embedding, 16 of 1-level embedding, and 18 of 2-level embedding, with each level of embedding having half grammatical and half ungrammatical structures. The sequences used in Experiment 1B were identical to those in 1A except that they were converted into a right-branching structure.

Procedure The experiments were run using the E-Prime presentation software with stimuli presented on a computer monitor. Participants in each experiment were randomly assigned to one of the possible conditions: starting small, random, or control. The starting small and random subjects were instructed that they were participating in a memory experiment. They were told that in the first part of the experiment they would see sequences of letters displayed on the screen and that they would be tested later on what they observed. Each sequence in its entirety was presented individually, for a duration of four seconds each. Each of the 50 training items was presented 3 times, for a total of 150 input exposures. The starting small participants received the input staged by level of embedding. The 0-level sequences were presented first; next the 1-level sequences, and last the 2-level sequences. Sequences were randomized within levels. The random group received all the sequences across all levels of embedding intermixed with one another, in random order. Thus, both the starting small and the random groups received the same training input but in different orders of presentation. The control group participants of Experiment 1A did not take part in the training phase.

After the training phase, the starting small and random participants were then told that the items they had just seen had been generated by a complex set of rules which determined the order of the letters. They were instructed that they would now see new letter strings, some of which followed the rules of the grammar, and some of which did not. Their task was to classify whether each letter string followed the same rules as the training sequences or not, by pressing a button marked "YES" or "NO". Both the starting small and random groups received the same test instructions and the same set of 50 test sequences were presented in random order for each participant. The control group participants of Experiment 1A received the test phase only.

Results and Discussion

For Experiment 1A, the mean number of correct endorsements on the 50 test items was 31.5 (63.0%) for the starting small group, 26.4 (52.8%) for the random group, and 26.5 (53.0%) for the no-training control group. We conducted single group t-tests and found that only the starting small group performed significantly above chance levels ($t(7) = 4.08; p < 0.005$). We also compared performance between each of the three groups. The starting small group performed significantly better than both the random group ($t(7) = 2.88; p < 0.05$) and the control group ($t(7) = 3.88; p < 0.05$).

The results of Experiment 1A show that only when the input was presented in a staged fashion were participants able to successfully learn aspects of the center-embedded recursive structure of the artificial grammar. Crucially, the starting small group out-performed the random group, lending empirical support to the starting small hypothesis.

For Experiment 1B, the mean number of correct endorsements on the 50 test items was 35.0 (70.0%) for the starting small group and 27.43 (54.9%) for the random group. Only the starting small group performed significantly above chance levels ($t(6) = 6.99; p < 0.001$). The starting small group also performed significantly better than the random group ($t(6) = 3.86; p < 0.01$).

Like Experiment 1A, these results show that only the starting small group was able to successfully learn aspects of the recursive structure of the artificial grammar. Besides serving as a replication of the general effect of starting small with staged input, it extends its applicability to right-branching structures.

Experiments 2A and 2B: Auditory Learning of Recursive Structure

The first set of experiments reveal that starting small is applicable for visual recursive input. Next we investigate whether starting small also extends to the auditory domain, using the same center-embedded and right-branching sequences from Experiments 1A and 1B.

Method

Subjects For Experiment 2A, eighteen new undergraduate subjects (nine in each condition) were recruited from introductory Psychology classes at Cornell University, earning extra credit for their participation. For Experiment 2B, sixteen new subjects (eight in each condition) were recruited in an identical manner.

Materials The same center-embedded and right-branching input sequences were used from Experiments 1A and 1B except that each letter was mapped onto a consonant-vowel-consonant syllable: C = "biff"; Q = "rud"; M = "sig"; P = "vot"; X = "mib"; S = "jux"; W = "nep"; Z = "dak"; K = "tood"; H = "jic"; T = "cav"; L = "dup". An example of a 0-level center-embedded sequence is "biff-nep", a 1-level sequence is "biff-vot-cav-nep", and a 2-level sequence is "biff-vot-rud-sig-cav-nep". An example of a 0-level right-branching sequence is "biff-nep", a 1-level sequence is "biff-nep-vot-cav", and a 2-level sequence is "biff-nep-vot-cav-rud-sig".

Auditory sequences were generated using the Festival speech synthesizer, which converts written text to synthesized speech (Black, Taylor, & Caley, 1998).

Procedure The procedures for Experiments 2A and 2B were the same as the previous experiments except that the auditory sequences were presented over headphones at a sound level of 70 dB.

Results and Discussion

For Experiment 2A, the mean number of correct endorsements on the 50 test items was 26.4 (52.8%) for the starting small group and 26.1 (52.2%) for the random group. Neither group performed significantly better than chance levels ($p's > 0.1$). In addition, there was no difference in performance between the two groups ($t(8) = -0.23; p = 0.82$). Thus, the learning of auditory, center-embedded structure does not appear to be facilitated with a staged input scheme, at least not with the present stimuli and experimental design.

For Experiment 2B, the mean number of correct endorsements on the 50 test items was 30.0 (60.0%) for the starting small group and 28.8 (57.5%) for the random group. Unlike Experiment 2A, both the starting small group ($t(7) = 3.21; p < 0.05$) and the random group ($t(7) = 3.47; p < 0.05$) performed significantly greater than chance levels. However, there were no performance differences between the two groups ($t(7) = 1.02; p = 0.34$). So, although subjects were capable of learning the right-branching recursive structure, a staged input scheme did not result in a starting small effect.

General Discussion

These four experiments provide insight as to when less is more and when less is less. Experiments 1A and 1B revealed that participants learned visual recursive structure better when the input was staged in an incremental fashion. The situation was more complex for auditory learning. Experiment 2A showed that participants were unable to adequately learn the center-embedded recursive structure of the input sequences, regardless of whether the input was staged or not. Experiment 2B revealed that although the participants were able to learn the right-branching structure, there was no difference in performance between the staged and non-staged conditions. Thus, at least for the present stimuli and procedures, there was no effect of starting small for auditory, recursive structure. The data are summarized in Figure 1.

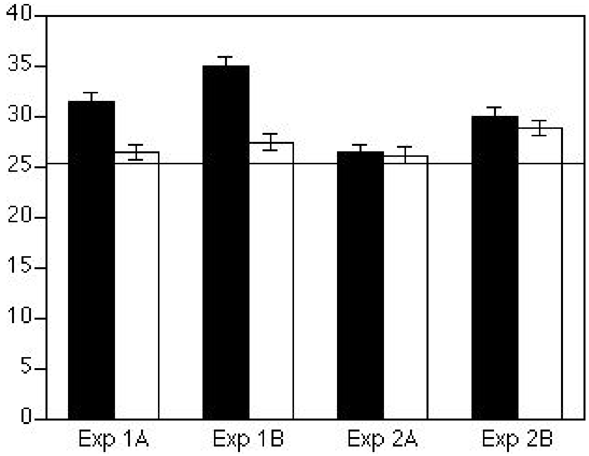

Figure 1: Test performance for starting small (shaded) and random (unshaded) conditions in Experiments 1A, 1B, 2A, and 2B.

Under what conditions is less more? We suggest that there may be multiple factors that determine whether there will be a learning advantage, including: whether starting small is implemented in terms of external or internal limitations; which sensory modality receives the input; and the input's level of complexity. Our results reveal that starting small may be most advantageous when the input is staged incrementally, is presented visually, and is relatively complex (i.e., recursive).

One might wonder why an effect of starting small would be found for visual but not auditory input. Certainly, if starting small aids in language acquisition, as most proponents of the theory have suggested, the lack of an effect for auditory learning is surprising. One possibility is that the observed differences actually reflect differences in processing serially-presented (Experiment 2) versus simultaneous (Experiment 1) input. This may explain why it was difficult for participants to learn the center-embedded recursive structure of Experiment 2A: memory constraints prohibited the learning of long-distance, non-adjacent relationships. Accordingly, participants could learn the right-branching structure of Experiment 2B because it consisted only of adjacent dependencies.

However, despite the general learning effect in Experiment 2B, there was *no effect of starting small.*

Thus, it may be that the lack of an auditory starting small effect arose from some constraint intrinsic to the auditory modality itself, rather than from serial input processing limitations. Modality differences in sequential learning tasks have been previously observed (Conway & Christiansen, 2002; Saffran, 2002), which suggest that under some conditions, humans are better at encoding and processing auditory compared to visual sequences. This might explain why under "normal", non-staged input conditions, auditory performance (random condition, Experiment 2B) was numerically greater than visual performance (random conditions, Experiments 1A and 1B). However, the novel, unexpected result here is that when the input is staged incrementally, visual learning improves while auditory learning does not. Besides replicating this result, future work must attempt to explain why vision and audition might be differentially affected by staged input.

Of course, perhaps starting small does help auditory learning but the current experimental conditions were insensitive to the effect. There was a small but statistically non-significant effect of starting small in Experiment 2B. It is possible that with more participants and increased statistical power, different stimuli, or more training, a stronger effect could be revealed. If so, this would fit with the observation that whereas center-embedded constructions are infrequent in the world's languages, right-branching structure is fairly common. This in turn may suggest a direct role for starting small in language acquisition.

Conclusion

Whether less is more or less is less appears to depend on a number of factors. We have found that an incrementally-staged training scheme improves the learning of visual, recursive input. At least with the present procedures and stimuli, a staged input scheme does not appear to aid auditory learning. It remains to be seen whether this is also the case for other types of auditory stimuli, such as tone sequences. Although the lack of an auditory starting small effect suggests starting small may not play a major role in spoken language acquisition, it is also likely that starting small can be considered as one cue out of many used in the service of learning language. If so, then starting small may have a more noticeable effect when it is combined and integrated with other cues (Christiansen & Dale, 2001).

The present results also may point to differences in the way that spoken vs. visual-based languages (such as ASL) are acquired. If starting small aids visual learning, as Experiments 1A and 1B show, then the learning of complex structure in sign languages may also benefit from a staged input training regiment. Future experiments may help verify this hypothesis, as well as uncover what role starting small plays in spoken language acquisition and other complex learning domains.

Acknowledgments

This research has been supported in part by the Human Frontiers Science Program.

References

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes*, *1*, 249-262.

Black, A., Taylor, P., & Caley, R. (1998). *The Festival speech synthesis system* [On-line]. Available: http://festvox.org/festival

Christiansen, M.H. & Dale, R.A.C. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 220-225). Mahwah, NJ: Lawrence Erlbaum.

Cochran, B.P., McDonald, J.L., & Parault, S.J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language*, *41*, 30-58.

Conway, C.M. & Christiansen, M.H. (2002). Sequential learning by touch, vision, and audition. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 220-225). Mahwah, N.J.; Lawrence Erlbaum.

Ellefson, M. R. (2002). *The difficulty of learning complex structure: A comparative study of knowledge acquisition.* Unpublished doctoral dissertation, Southern Illinois University, Carbondale.

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

Fletcher, J., Maybery, M.T., & Bennett, S. (2000). Implicit learning differences: A question of developmental level? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 246-252.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General*, *126*, 278-287.

Kersten, A.W. & Earles, J.L. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language*, *44*, 25-273.

Ludden, D. & Gupta, P. (2000). Zen in the art of language acquisition: Statistical learning and the less is more hypothesis. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 812-817), Mahwah, NJ: Lawrence Erlbaum Associates.

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

Pine, J.M. (1994). The language of primary caregivers. In C. Gallaway & B.J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 109-149)

Rohde, D.L.T. & Plaut, D.C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*, 67-109.

Rohde, D.L.T. & Plaut, D.C. (in press). Less is less in language acquisition. To appear in Quinlin, P. (Ed.), *Connectionist modeling of cognitive development.* Hove, UK: Psychology Press.

Saffran, J.R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*, 172-196.