In The Proceedings of the 22nd Annual Conference of the Cognitive Science Society (pp. 83-88), 2000. Mahwah, NJ: Lawrence Erlbaum.

A Connectionist Single-Mechanism Account of Rule-Like Behavior in Infancy

Morten H. Christiansen (MORTEN@SIU.EDU)
Christopher M. Conway (CONWAY@SIU.EDU)
Department of Psychology; Southern Illinois University
Carbondale, IL 62901-6502 USA

Suzanne Curtin (CURTIN@GIZMO.USC.EDU)
Department of Linguistics; University of Southern California
Los Angeles, CA 90089-1693 USA

Abstract

One of the most controversial issues in cognitive science pertains to whether rules are necessary to explain complex behavior. And nowhere has the debate over rules been more heated than within the field of language acquisition. Most researchers agree on the need for statistical learning mechanisms in language acquisition, but disagree on whether rule-learning components are also needed. Marcus, Vijayan, Rao, & Vishton (1999) have provided evidence of rule-like behavior which they claim can only be explained by a dual-mechanism account. In this paper, we show that a connectionist single-mechanism approach provides a more parsimonious account of rule-like behavior in infancy than the dual-mechanism approach. Specifically, we present simulation results from an existing connectionist model of infant speech segmentation, fitting the behavioral data under naturalistic circumstances and without invoking rules. We further investigate diverging predictions from the single- and dual-mechanism accounts through additional simulations and artificial language learning experiments. The results support a connectionist single-mechanism account, while undermining the dual-mechanism account.

Introduction

The nature of the learning mechanisms that infants bring to the task of language acquisition is a major focus of research in cognitive science. With the rise of connectionism, much of the scientific debate surrounding this research has focused on whether rules are necessary to explain language acquisition. All parties in the debate acknowledge that statistical learning mechanisms form a necessary part of the language acquisition process (e.g., Christiansen & Curtin, 1999; Marcus, Vijayan, Rao, & Vishton, 1999; Pinker, 1991). However, there is much disagreement over whether a statistical learning mechanism is sufficient to account for complex rule-like behavior, or whether additional rule-learning mechanisms are needed. In the past this debate has primarily taken place within specific areas of language acquisition, such as inflectional morphology (e.g., Pinker, 1991; Plunkett & Marchman, 1993) and visual word recognition (e.g., Coltheart, Curtis, Atkins & Haller, 1993; Seidenberg & McClelland, 1989). More recently, Marcus et al. (1999) have presented results from experiments with 7-month-olds, apparently showing that the infants acquire abstract algebraic rules after two minutes of exposure to habituation stimuli. The algebraic rules are construed as representing an open-ended relationship between variables for which one can substitute arbitrary values, "such as 'the first item X is the same as the third item Y,' or more generally, that 'item I is the same as item J"' (Marcus et al., 1999, p. 79). Marcus et al. further claim that a connectionist single-mechanism approach based on statistical learning is unable to fit their experimental data. In this paper, we build on earlier work (Christiansen & Curtin, 1999) and present a detailed connectionist model of these infant data, and provide new experimental data that support a statistically-based single-mechanism approach while undermining the dual-mechanism account.

In the remainder of this paper, we first show that knowledge acquired in the service of learning to segment the speech stream can be recruited to carry out the kind of classification task used in the experiments by Marcus et al. For this purpose we took an existing model of early infant speech segmentation (Christiansen, Allen & Seidenberg, 1998) and used it to simulate the results obtained by Marcus et al. The simulations demonstrate that no rules are needed to account for the data; rather, statistical knowledge related to word segmentation can explain the rule-like behavior of the infants in the Marcus et al. study. We then explore the issue of timing in stimuli presentation and present additional simulations from which empirical predictions are derived that diverge from those of the rule-based account. These predictions are tested in experiments with adults. Experiment 1 replicated the results from Marcus et al. using adult subjects. Experiment 2 confirmed the predictions from our single mechanism approach. Together, the simulations and experiments suggest that a single mechanism model provides the most parsimoneous account of the empirical data, thus obviating the need for a dual mechanism approach involving a separate rule-learning component.

Simulation 1: Rule-Like Behavior without Rules

Marcus et al. (1999) used an artificial language learning paradigm to test their claim that the infant has two mechanisms for learning language, one that uses statistical information and another which uses algebraic rules. They conducted three experiments which tested infants' ability to generalize to items not presented in the familiarization phase of the experiment. We focus here on their third experiment because it was controlled for possible confounds found in the first two experiments: differences in phonetic features (Experiment 1) and reduplication (Experiment 2). Marcus et al. claim that because none of the test items appeared in the habituation part of the experiment the infants would not be able to use statistical information in this task.

The subjects in Experiment 3 of Marcus et al. (1999) were seven-month old infants randomly placed in an AAB or an ABB condition. During a two-minute long familiarization phase the infants were exposed to three repetitions of each of 16 three-word sentences. Each word in the sentence frame AAB or ABB consisted of a consonant-vowel sequence (e.g., “le le we” or “le we we”). The test phase consisted of 12 sentences made up of words to which the infants had not previously been exposed (e.g., “ko ko ga” vs “ko ga ga”). The test items were broken into two groups for both habituation conditions: consistent (items constructed with the same grammar as the familiarization phase) and inconsistent (constructed from the grammar the infants were not habituated on). The results showed that the infants preferred the inconsistent test items over the consistent ones.

The conclusion drawn by Marcus et al. (1999) was that a single mechanism which relied on statistical information alone could not account for the results. Instead they suggested that a dual mechanism was needed, comprising a statistical learning component and an algebraic rule learning component. In addition, they claimed that a Simple Recurrent Network (SRN; Elman, 1990) would not be able to accommodate their data because of the lack of phonological overlap between habituation and test items. Specifically, they state,

> Such networks can simulate knowledge of grammatical rules only by being trained on all items to which they apply; consequently, such mechanisms cannot account for how humans generalize rules to new items that do not overlap with the items that appeared in training (p. 79).

In the first simulation, we demonstrate that SRNs can indeed fit the data from Marcus et al. Other researchers have constructed neural network models specifically to simulate the Marcus et al. results (Altmann & Dienes, 1999; Elman, 1999; Shastri & Chang, 1999; Shultz, 1999). In constrast, we do *not* build a new model to accommodate the results, but take an existing SRN model of speech segmentation (Christiansen et al., 1998) and show how this model—*without additional modification*—provides an explanation for the results.

Simulation Details

The model by Christiansen et al. (1998) was developed as an account of early word segmentation. An SRN was trained on a *single* pass through a corpus of child directed speech. As input the network was provided with three cues: (a) phonology represented in terms of 11 features on the input and 36 phonemes on the output[1], (b) utterance boundary information represented as an extra feature marking utterance endings, and (c) lexical stress coded over two units as either no stress, secondary or primary stress. Figure 1 provides an illustration of the network.

The network was trained on the task of predicting the next phoneme in a sequence as well as the appropriate values for the utterance boundary and stress units. In learning to perform this task the network also learned to integrate the cues such that it could carry out the task of segmenting the input into words. This involved activating the bundary unit not only

[1]Phonemes were used as output in order to facilitate subsequent analyses of how much knowledge of phonotactics the net had acquired.

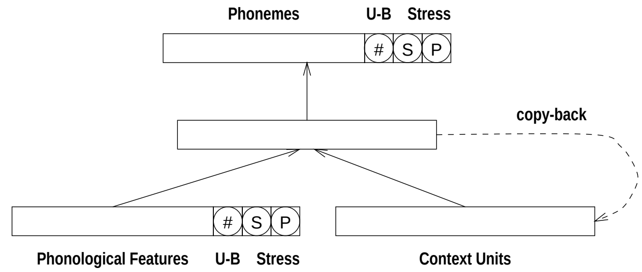

Figure 1: Illustration of the SRN used in Simulations 1 and 2. Solid lines indicate trainable weights, whereas the dashed line denotes the copy-back weights (which are always 1). U-B refers to the unit coding for the presence of an utterance boundary. The presence of lexical stress is represented in terms of two units, S and P, coding for secondary and primary stress, respectively.

at utterance boundaries, but also at word boundaries inside utterances. The logic behind the segmentation task is that the end of an utterance is also the end of a word. If the network is able to integrate the provided cues in order to activate the boundary unit at the ends of words occurring at the end of an utterance, it should also be able to generalize this knowledge so as to activate the boundary unit at the ends of words which occur *inside* an utterance (Aslin, Woodward, LaMendola & Bever, 1996).

The Christiansen et al. (1998) model acquired distributional knowledge about sequences of phonemes and the associated stress patterns. This knowledge allowed it to perform well on the task of segmenting the speech stream into words. We suggest that this knowledge can be put to use in secondary tasks not directly related to speech segmentation—including artificial tasks used in psychological experiments such as Marcus et al. (1999). This suggestion resonates with similar perspectives in the word recognition literature (Seidenberg, 1995) where knowledge acquired for the primary task of learning to read can be used to perform other secondary tasks such as lexical decision.

Method *Networks.* We used 16 SRNs similar to the SRNs used in Christiansen et al. (1998) with the exception that the original phonetic feature representation was replaced by a new representation using 18 features. Each of the 16 SRNs had a different set of initial weights, randomized within the interval [0.25;-0.25]. The learning rate was set to 0.1 and the momentum to 0.95. These training parameters were identical to those used in the original Christiansen et al. model. The networks were trained to predict the correct constellation of cues given the current input segment.

Materials. Prior to being habituated and tested on the stimuli from Marcus et al. the networks were first exposed to the training corpus used by Christiansen et al. This corpus consists of 8181 utterances extracted from the Korman (1984) corpus of British English speech directed at pre-verbal infants aged 6-16 weeks (a part of the CHILDES database, MacWhinney, 1991). The corpus contained a total of 24,648 words distributed over 814 types and had an average utterance length of 3.0 words (see Christiansen et al. for further details). Christiansen et al. transformed each word in the utterances from its orthographic format into a phonological form with accompanying lexical stress using a dictionary compiled from the MRC Psycholinguistic Database available from the Oxford Text Archive[2].

The materials from Experiment 3 in Marcus et al. (1999) were transformed into the phoneme representation used by Christiansen et al. Two habituation sets were created in this manner: one for AAB items and one for ABB items. The habituation sets used here, and in Marcus et al., consisted of 3 blocks of 16 sentences in random order, yielding a total of 48 sentences in each habituation set[3]. As in Marcus et al. there were four different test sentences: “ba ba po”, “ko ko ga” (consistent with AAB), “ba po po” and “ko ga ga” (consistent with ABB). The test set consisted of three blocks of randomly ordered test sentences, totaling 12 test sentences. Both the habituation and test sentences were treated as a single utterance with no explicit word boundaries marked between the individual words. The end of each utterance was marked by activating the utterance boundary unit.

Procedure. The networks were first trained on a single pass through the Korman (1984) corpus as in the original Christiansen et al. model. This corresponds to the fact that the 7-month-olds in the Marcus et al. study already have had a considerable exposure to language, and have begun to develop their speech segmentation abilities. Next, the networks were habituated on a *single* pass through the appropriate habituation corpus—one phoneme at a time—with learning parameters identical to the ones used during the pretraining on the Korman corpus. The networks were then tested on the test set (with the weights “frozen”) and the activation of the utterance boundary unit was recorded for every phoneme input in the test set. Finally, the boundary unit activations for test sentences that were consistent or inconsistent with the habituation pattern were separated into two groups. Furthermore, for the purpose of scoring word segmentation performance on the test items, the activation of the boundary unit was also recorded for each habituation condition across all the habituation items and the mean activation was calculated. The networks were said to have postulated a word boundary whenever the boundary unit activation in a test sentence was above the appropriate habituation mean.

Results and Discussion

To provide a quantitative measure of performance we used completeness scores (Christiansen et al., 1998) to assess segmentation performance.

$$\text{Completeness} = \frac{\text{Hits}}{\text{Hits} + \text{Misses}} \tag{1}$$

[2]Note that these phonological *citation forms* were unreduced (i.e., they did not include the reduced vowel *schwa*). The stress cue therefore provided additional information not available in the phonological input.

[3]The 16 habituation sentences that followed the AAB sentence frame were ”de de di,” ”de de je,” ”de de li,” ”de de we,” ”ji ji di,” ”ji ji je,” ”ji ji li,” ”ji ji we,” ”le le di,” ”le le je,” ”le le li,” ”le le we,” ”wi wi di,” ”wi wi je,” ”wi wi li,” and ”wi wi we”. The 16 habituation sentences that followed the AAB sentence frame were ”de di di,” ”de je je,” ”de li li,” ”de we we,” ”ji di di,” ji je je,” ”ji li li,” ”ji we we,” ”le di di,” ”le je je,” ”le li li,” ”le we we,” ”wi di di,” ”wi je je,” ”wi li li,” and ”wi we we”.

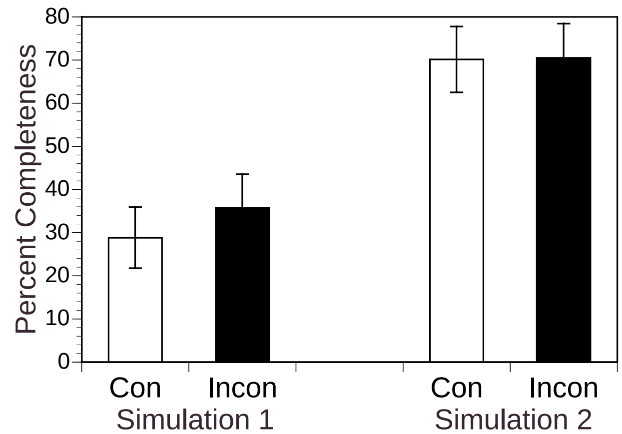

Figure 2: Mean completeness scores for the consistent (con) and inconsistent (incon) test items from Simulations 1 (left) and 2 (right).

Completeness provides a measure of how many of the words in a test set the net is able to discover. With respect to our interpretation of the Marcus et al. data, the completeness score indicates how well networks/infants are at segmenting out the individual words in the test sentences. As an example, consider the following two hypothetical segmentations of two test sentences:

b a b # a # p o # k o # g a g # a

where # corresponds to a predicted word boundary. Here the hypothetical learner correctly segmented out two words, *po* and *ko*, but missed the first and the second *ba* and the first and the second *ga*. This results in a completeness score of 2/(2+4) = 33.3%.

For each of the sixteen networks, completeness scores were computed across all test items, and submitted to the same statistical analyses as used by Marcus et al. for their infant data. The completeness scores were submitted to a repeated measures ANOVA with condition (AAB vs. ABB) as between network factor and test pattern (consistent vs. inconsistent) as within network factor. The left-hand side of Figure 2 shows the completeness scores for the consistent and inconsistent items pooled across conditions. There was a main effect of test pattern ($F(1,14) = 5.76, p < .04$), indicating that the networks were signifantly better at segmenting out the words in the inconsistent items (35.76%) compared with the consistent items (28.82%). Similarly to the infant data, neither the main effect of condition, nor the condition $\times$ test pattern interaction were significant ($F's < 1$). The better segmentation of the inconsistent items suggests that they would stand out more clearly in comparison with the consistent items, and thus explain why the infants looked longer towards the speaker playing the inconsistent items in the Marcus et al. study.

Marcus et al. claim a statistical learning mechanism and a rule-learning mechanism. Simulation 1 shows that a separate rule-learning component is not necessary to account for the data. This simulation shows how an existing SRN model of word segmentation can fit the data from Marcus et al. (1999) without invoking explicit rules. The pretraining allowed the SRNs to learn to integrate the regularities governing the phonological, lexical stress, and utterance boundary information in child-directed speech. This form of statistical learning enabled it to fit the infant data. Importantly, the SRN model—as a statistical learning mechanism—can explain both the distinction between consistent and inconsistent items as well as the preference for the inconsistent items. Note that a rule-learning mechanism by itself only can explain how infants may distinguish between items, but not why they prefer inconsistent over consistent items. Extra machinery is needed in addition to the rule-learning mechanism to explain the preference for inconsistent items. Thus, the most parsimonious explanation is that only a statistical learning device is necessary to account for the infant data. The addition of a rule-learning device does not appear to be necessary.

Simulation 2: It's about Time

Simulation 1 demonstrated that a statistically-based single-mechanism approach can account for the kind of rule-like behavior displayed by the infants in the Marcus et al. study. However, there may be other cases in which a separate rule-learning component would be required. Here we explore one such case in which our model makes a prediction which is different from what would be predicted from a dual-mechanism approach incorporating a rule-learning component.

Recall that algebraic rules were characterized as abstract relationships between variables, such as item X is the same as item Y. Marcus et al. Experiment 3 was designed to demonstrate that rule learning is independent of the physical realization of variables in terms of phonological features. The same rule, AAB, applies to—and can be learned from—"le le we" and "ko ko ga" (with "le" and "ko" filling the same A slot and "we" and "ga" the same B slot). As the abstract relationships that this rule represents only pertains to the value of the three variables, the amount of time between them should not affect the application of the rule. Thus, just as the physical realization of a variable does not matter for the learning or application of a rule, neither should the time between variables. The same rule AAB, applies to—and can be learned from—"le [250ms] le [250ms] we" and "le [1000ms] le [1000ms] we" (the "le"s should still fill the A slots and the "we"s the B slot despite the increased duration of time between the occcurrence of these variables). From this property, one can predict that lengthening the time between variables should not affect the preference for inconsistent items. Indeed, the connectionist implementation of the rule-based approach found in the Shastri & Chang (1999) model would appear to make this prediction.

A lengthening of the pauses between words should, however, have a different effect on our model. In the model, the preference for inconsistent items observed by Marcus et al. is explained in terms of differential segmentation performance. Lengthening the pauses between words would in effect solve the segmentation task for the model, and should result in a disappearance of the preference for inconsistent items. Thus, we predict that the model should show no difference between the segmentation performance on the consistent and inconsistent items if pauses are lengthened as indicated above. To test this prediction, we carried out a new set of simulations.

Method

Networks. Sixteen SRNs as in Simulation 1.

Materials. Same as in Simulation 1 save that utterance boundaries were inserted *between* the words in the habituation and test sentences, simulating a lengthening of pauses between words such that they have the same length as the pauses between utterances.

Procedure. Same as in Simulation 1.

Results and Discussion

Completeness scores were computed as in Simulation 1 and submitted to the same statistical analysis. As illustrated by the right-hand side of Figure 2, the segmentation performance on the test items was improved considerably by the inclusion of utterance boundary-length pauses between words. As predicted, there was no difference between the accuracy scores for consistent (70.14%) and inconsistent items (70.49%) ($F(1,14) = .02$). As before, there was no main effect of condition, neither was there any interaction between condition and test pattern and ($F's < 1$).

Simulation 2 thus confirms the predicted effect of lengthening the pauses between words in stimuli presented to the statistical learning model. This results in diverging predictions derived from the rule-based and the statistical learning model concerning the effect of pause lengthening on human performance on the stimuli. Next, we test these diverging predictions in an artificial language learning experiment using adult subjects.

Experiment 1: Replicating the Marcus et al. Results

Before testing the diverging predictions from the single- and dual-mechanism approaches we need to first establish whether adults in fact exhibit the same pattern of behavior as the infants in the Marcus et al. study. The first experiment therefore seeks to replicate Experiment 3 from Marcus et al. using adult subjects instead of infants.

Method

Subjects. Sixteen undergraduates were recruited from introductory Psychology classes at Southern Illinois University. Subjects earned course credit for their participation.

Materials. For this experiment, we used the original stimuli that Marcus et al. (1999) created for their Experiment 3. Each word in a sentence was separated by 250 msec. The 16 habituation sentences for each condition were created by Marcus et al. using the Bell Labs speech synthesizer. The original habituation stimuli was limited to two predetermined sentence orders. To avoid potential order effects, we used the SoundEdit 16 version 2 software for the MacIntosh to isolate each sentence as a separate sound file. This allowed us to present the habituation sentences in a random order for each subject.

For the test phase, we also used the stimuli from Marcus et al.'s Experiment 3, which consisted of four new sentences that were either consistent or inconsistent with the training grammar. As mentioned earlier, these sentences contained no phonological overlap with the habituation sentences. Like the habituation stimuli, each word in a sentence was separated by a 250 msec interval. As before, we stored the test stimuli as separate SoundEdit 16 version 2 sound files to allow a random presentation order for each subject.

Procedure. Subjects were seated in front of a G3 Power Macintosh computer with a New Micros button box. Subjects were randomly assigned to one of two conditions, AAB or ABB. The experiment was run using the PsyScope presentation software (Cohen, MacWhinney, Flatt, and Provost, 1993) with all stimuli played over stereo loudspeakers at 75dB. The subjects were instructed that they were participating in a pattern recognition experiment. They were told that in the first part of the experiment their task was to listen carefully to sequences of sounds and that their knowledge of these sound sequences would be tested afterwards. Subjects listened to three blocks of the sixteen randomly presented habituation sentences corresponding either to the AAB or the ABB sentence frame. A 1-second interval separated each sentence as was the case in the Marcus et al. experiment.

After habituation, subjects were instructed that they would be presented with new sound patterns that they had not previously heard. They were asked to judge whether a pattern was ”similar” or ”dissimilar” to what they had been exposed to in the previous phase by pressing an appropriately marked button. The instructions emphasized that because the sounds were novel, the subjects should not base their decision on the sounds themselves but instead on the patterns derived from the sounds. Subjects listened to three blocks of the four randomly presented test sentences. After the presentation of each test sentence, subjects were prompted for their response. Subjects were allowed to take as long as they needed to respond. Each test trial was separated by a 1000-msec interval.

Results and Discussion

For the purpose of our analyses, the correct response for consistent items is “similar” while the correct response for inconsistent items is “dissimilar”. The mean overall score for correct classification of test items was 8.81 out of a perfect score of 12. A single-sample t-test showed that this classification performance was significantly better than the chance level performance of 6 ($t(15) = 4.44, p < .0005$). Subjects' responses were then subject to the same statistical analysis as the infant data in Marcus et al. (and Simulation 1 and 2 above). The left-hand side of Figure 3 shows the ratings as dissimilar for the six consistent and six inconsistent test items pooled across condition. As expected, there was a main effect of test pattern ($F(1, 14) = 18.98, p < .001$), such that significantly more inconsistent items were judged as dissimilar (4.5) than consistent items (1.69). Neither the main effect of condition, nor the condition $\times$ test pattern interaction were significant ($F's < 1$).

Experiment 1 shows that adults perform similarly to the infants in Marcus et al.'s Experiment 3, thus demonstrating that it is possible to replicate their findings using adult subjects instead of infants. This result is perhaps not surprising given that Saffran and colleagues were able to replicate statistical learning results obtained using adults subjects (Saffran, Newport & Aslin, 1996) in experiments using 8-month-olds (Saffran, Aslin & Newport, 1996). More generally, these results and ours suggest that despite small differences in the experimental methodology used in infant and adult artificial language learning studies, both methodologies appear to tap into the same learning mechanisms. Also from a dual-mechanism approach, one would expect that the same learning mechanisms—statistical and rule-based—would be involved in both infancy and adulthood, and that similar results should be expected in both infant and adult studies of the kind of material used here.

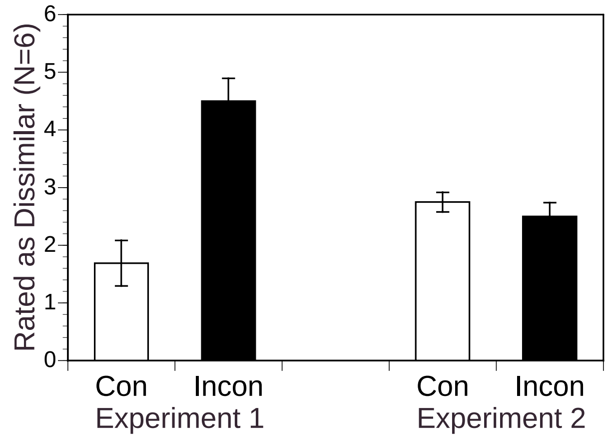

Figure 3: The mean proportion of consistent (con) and inconsistent (incon) test items rated as dissimilar to the habituation pattern in Experiments 1 (left) and 2 (right).

Experiment 2: Testing the Diverging Predictions

Having replicated the Experiment 3 infant data with adult subjects, we now turn our attention to the diverging predictions concerning the effect of pause length on the preference for the inconsistent items.

Method

Subjects. Sixteen additional undergraduates were recruited from introductory Psychology classes at Southern Illinois University. Subjects earned course credit for their participation.

Materials. The training and test stimuli were the same as those used in Experiment 1 except that the 250 msec interval between words in a sentence was replaced by a *1000 msec* interval using the SoundEdit 16 version 2 software. The 1000 msec interval between sentences remained the same as before.

Procedure. The procedure and instructions were identical to that used for Experiment 1.

Results and Discussion

The mean overall classification score was 5.75 out of 12. This was not significantly different from the chance level performance of 6, as indicated by a single-sample t-test ($t < 1$). The responses of the subjects were submitted to the same further analysis as in Experiment 1. The right-hand side of Figure 3 shows the ratings as dissimilar for the consistent and inconsistent test items averaged across condition. As predicted by Simulation 2, there was *no* main effect of test pattern in this experiment ($F(1,14) = .56$), suggesting that subjects were unable to distinguish between consistent and inconsistent items. Similarly to Experiment 1, both the main effect of condition and the interaction between condition and test pattern interaction were not significant ($F's = 0$).

These results show that preference for inconsistent items disappears when the pauses between words are lengthened. This corroborates the prediction from the statistically-based single-mechanism model, but not the prediction from the rule-learning component of the dual-mechanism account. It may be objected that the rules need to work over specific domains, and that by lengthening the pauses between words the input is no longer chunked into sentences at a pre-specified length (three words). Hence, the rule can no longer be expected to apply. Note, however, that this requires additional machinery to pre-process the input prior to the learning or application of a rule. This would require a separate account of how this pre-processing ability was acquired and how it was applied in the specific case of Marcus et al.'s original experiment. Of course, this makes the rule-based account even less parsimonious in comparison with the statistical learning model. The latter model can account for both the preference for inconsistent items in the Marcus et al. Experiment 3 (and our Experiment 1) as well as the lack of preference in our Experiment 2 *without requiring any extra machinery*. Thus, a language learning device that exploits the statistical properties of language and integrates these multiple cues can account for Marcus et al. data thereby removing to need to posit a dual-learning mechanism.

Conclusion

In this paper, we have shown that a single-mechanism approach provides the most parsimonious account of both statistical learning abilities and rule-like behavior in infancy. Simulation 1 showed that an existing connectionist model of early infant word segmentation (Christiansen et al., 1998) could utilize statistical knowledge acquired in the service of speech segmentation to fit the infant data from Marcus et al. (1999; Experiment 3) under very naturalistic circumstances. We then explored whether there were other circumstances that would require a separate rule-learning component. For this purpose, we derived differing predictions from the single- and dual-mechanism accounts concerning the effect of pause lengthening on rule-like behavior. Simulation 2 demonstrated that our connectionist model was unable to distinguish between inconsistent and consistent items when the pauses between words were of the same length as the pauses between utterances. In contrast, pause-lengthening should not affect the application of algebraic rules in a dual-mechanism model because the abstract relationhip between variables remain the same independently of pause duration. We tested these predictions using adult subjects. Experiment 1 replicated the original results from Experiment 3 in Marcus et al. Experiment 2 tested the diverging predictions and found that subjects' performance mirrored the results obtained from Simulation 2; that is, the subjects performed according to the single-mechanism prediction and not the dual-mechanism prediction. Together, the results from both the simulations and the adult experiments show that the statistically-based single-mechanism approach embodied in our connectionist model provides the most simple account of the behavioral data, thus obviating the need for a separate rule-learning component.

Acknowledgments

We would like to thank Gary Marcus for making his stimuli available to us for our experiments.

References

Altmann, G.T.M. & Dienes, Z. (1999). Rule learning by seven-month-old infants and neural networks. *Science*, *284*, 875.

Aslin, R.N., Woodard, J.Z., LaMendola, N.P., & Bever, T.G. (1996). Models of word segmentation in fluent maternal speech to infants. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax*. Mahwah, NJ: Lawrence Erlbaum Associates.

Christiansen, M.H., Allen, J., & Seidenberg, M.S. (1998). Learning to segment using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*, 221-268.

Christiansen, M.H. & Curtin, S.L. (1999). The power of statistical learning: No need for algebraic rules. In *The Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 114-119). Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen J.D., MacWhinney B., Flatt M., & Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257-271.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589-608.

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

Elman, J. (1999). *Generalization, rules, and neural networks: A simulation of Marcus et. al, (1999)*. Ms., University of California, San Diego.

Korman, M. (1984). Adaptive aspects of maternal vocalizations in differing contexts at ten weeks. *First Language*, *5*, 44–45.

MacWhinney, B. (1991). *The CHILDES Project*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Marcus, G.F., Vijayan, S., Rao, S.B., & Vishton, P.M. (1999). Rule learning in seven month-old infants. *Science*, *283*, 77–80.

Pinker, S. (1991). Rules of language. *Science*, *253*, 530-535.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building. *Cognition*, *48*, 21-69.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month olds. *Science*, *274*, 1926–1928.

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

Seidenberg, M.S. (1995). Visual word recognition: An overview. In Peter D. Eimas & Joanne L. Miller (Eds.), *Speech, language, and communication. Handbook of perception and cognition, 2nd ed., Vol. 11*. San Diego: Academic Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568.

Shastri, L., & Chang, S. (1999). *A spatiotemporal connectionist model of algebraic rule-learning* (TR-99-011). Berkeley, California: International Computer Science Institute.

Shultz, T. (1999). Rule learning by habituation can be simulated by neural networks. In *Proceedings of the Twentyfirst Annual Conference of the Cognitive Science Society* (pp. 665-670). Mahwah, NJ: Lawrence Erlbaum Associates.