The Serial Reaction Time Task: Learning Without Knowing, or Knowing Without Learning?

Maud Boyer
mboyer@ulb.ac.be

Arnaud Destrebecqz
adestre@ulb.ac.be

Axel Cleeremans
axcleer@ulb.ac.be

Séminaire de Recherche en Sciences Cognitives
Université Libre de Bruxelles
Avenue F.-D. Roosevelt, 50 — CP 122
1050 Bruxelles — Belgium

Abstract

Constant interaction with a dynamic environment — from riding a bicycle to segmenting speech — makes sensitivity to the sequential structure of the world a crucial dimension of the cognitive system. Accounts of sequence learning vary widely, with some authors arguing that parsing and segmentation processes are central, and others defending the notion that sequence learning involves mere memorization. In this paper, we argue that sequence knowledge is essentially statistical in nature and that sequence learning involves simple associative prediction mechanisms. We focus on a choice reaction situation introduced by Lee (1997), in which participants were exposed to material that follows an extremely simple rule, namely that stimuli are selected randomly but never appear more than once in a legal sequence. Perhaps surprisingly, people can learn this rule very well. Or do they? We offer a conceptual replication of the original finding, but a very different interpretation of the results, as well as simulation work that makes it clear how highly abstract dimensions of the stimulus material can in fact be learned based on elementary associative mechanisms.

Introduction

Sequence learning is a fundamental process involved in the many different cognitive skills required for successful interaction with an intensely dynamic environment. Among those skills, language is probably the most complex, and the role that elementary associative sequence learning processes may play in its development have recently begun to be explored anew. For instance, Saffran et al. (Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A., & Barrueco, S., 1997) recently showed how incidental exposure to artificial language-like auditory material (e.g., *bupadapatubitutibu...*) was sufficient to enable participants to segment the continuous sequence of sounds they had heard into the artificial words (e.g., *bupada*, *patubi*, etc.) that it consisted of, as evidenced by their performance in a subsequent recognition test. Based on these data, Saffran et al. (1997) suggested that word segmentation abilities develop based on mechanisms that exploit the statistical regularities present in sequences of events, such as for instance the fact that the transitional probabilities of successive syllables are higher within words than between words. Interestingly, Saffran et al. (1997) rooted their interpretation of their findings in the apparently remote literature dedicated to implicit learning. The connection is obvious as soon as one recognizes that language acquisition, like implicit learning (see Berry & Dienes, 1993; Cleeremans, 1993 for reviews), is likely to involve, at least in part, incidental learning of complex information organized at different levels. In particular, research on sequence learning has, over the past decade or so provided a steady stream of relevant evidence suggesting that participants exhibit detailed sensitivity to the sequential structure through differences in their reaction time to stimuli that are or are not predictable based on the temporal context. In typical sequence learning situations, participants are asked to react to each element of sequentially structured and typically visual sequences of events (e.g., Nissen & Bullemer, 1987). Several versions of this basic paradigm can be distinguished. In *rule-based paradigms*, sequences either conform or fail to conform to an abstract rule that describes permissible transitions between successive stimuli. Rule-based paradigms can in turn involve either deterministic (e.g., Lewicki, Hill, & Bizot, 1988) or probabilistic rules, as when the stimulus material is generated based on the output of finite-state grammars (e.g., Cleeremans, 1993). By contrast, in the more common *simple repeating sequence paradigm*, a single sequence containing fixed regularities is repeated many times to produce the training set (e.g., Nissen & Bullemer, 1987).

A perennial question in this context is to determine exactly what people learn about when exposed to sequentially structured stimulus material. Perhaps unsurprisingly, it is often the case that several different accounts are partially or completely consistent with the data. Consider for instance a sequence learning situation in which the stimulus material consists of a simple repeating sequence such as "ABCDBA". When exposed to this material in the context of a choice reaction situation, participants could either (1) learn something about the generation rules, (2) memorize the entire sequence, (3) become sensitive to the frequency of specific repeating fragments of the sequence, (4) learn something about the conditional probability of occurrence of each element in the context of the previous elements, (5) learn about other aspects of the material such as specific movement patterns (e.g., alternations, trills, or more abstract patterns).

Cleeremans and Jiménez (1998) suggested that these different accounts may in fact often turn out to be descriptively equivalent, and concluded that the core processes involved in sequence learning are best thought of as involving elementary associative learning processes that result in a progressively developing sensitivity to the statistical constraints contained in the material (see also Stadler, 1992). Such processes are well instantiated by the Simple Recurrent Network (henceforth, SRN; see Elman, 1990; Cleeremans & McClelland, 1991), which we describe later in this paper.

In this context, Lee (1997) described an interesting sequence learning situation which, at first sight, seems to challenge traditional accounts of sequence learning. Indeed, the stimulus set used by Lee consisted of a random selection of the 720 (6!) sequences of six elements that are consistent with the following simple constraint: Each of six different elements can only appear once in each six-elements sequence. For instance, the sequences "123456" or "236145" are both legal because each stimulus appears only once. The sequence "235451", however, does not follow the rule because element '5' appears twice and element '6' is missing. This rule thus results in a probability gradient across the six positions within each sequence, such that the first element of any legal sequence is always completely unpredictable, and such that the subsequent elements become increasingly predictable based on the context set by the previous elements. The final element of each legal sequence is thus always completely predictable based on the first five elements. Lee's material thus contains almost no structure but for the single highly abstract structural property described by the generation rule. Nevertheless, Lee showed that participants trained on this material tend to respond faster to stimuli that occur in serial position 6 than to stimuli that appear in serial position 1, thereby indicating that they had learned something about the structure of the material. Lee also suggested that learning involved a combination of implicit and explicit learning, to the extent that people were unable to project their knowledge in various direct tests (e.g., recognition or prediction), but nevertheless exhibited better learning when informed of the nature of the rule.

As Lee (1997) indicated, traditional theories of sequence learning may have a hard time accounting for the data. Indeed, theories that rely on the notion that people memorize entire instances would have difficulty in this case because the stimulus material simply does not consist of a few repeating sequences. Fragment-based accounts also appear implausible because even three-elements fragments fail to convey much information about the relevant regularities. For instance, the fragment '123' may end in any of the 6 serial positions and be followed by any of the 6 possible elements but '3' (stimulus repetitions were forbidden). Lee concluded that "both parsing and short-term memory mechanisms must be involved" (p. 428), and that models based on simple associative learning mechanisms, such as the SRN model, were probably incapable of learning this stimulus material.

In the following, we first report on the results of a conceptual replication of Lee (1997)'s experiment. Next, and in contrast to Lee's conclusions, we show that participants' sensitivity to the rule used to generate the stimulus material can actually be understood based on the operation of elementary associative mechanisms that do not involve any parsing of the material. More importantly, we also challenge the idea that any learning is involved in this situation. Finally, we explore how well the SRN model can account for our data.

Experimental Design

Participants

Twelve participants took part in the experiment. They were paid a flat fee of about $14 and could earn an additional bonus of up to $9 based on performance at the task (see below).

Apparatus and Display

The experiment was run on PowerPC Macintosh computers. The display consisted of six dots arranged in a horizontal line on the computer's screen and separated by intervals of 3 cm. Each screen position corresponded to a key on the computer's keyboard. The spatial configuration of the keys was fully compatible with the screen positions. The stimulus was a small black circle 0.35 cm in diameter that appeared on a white screen background, centered 1 cm below one of the six dots. The RSI was 120 msec.

Procedure

The experiment consisted of 24 training blocks during which subjects were exposed to a serial six-choice RT task. Each block consisted of 180 trials, for a total of 4320 trials. On each trial, a stimulus appeared at one of the possible six positions. Participants were instructed to respond as fast and as accurately as possible by pressing on the corresponding key. The target was removed as soon as a key had been pressed, and the next stimulus appeared after a 120 msec interval. Erroneous responses were signaled to participants by means of a tone.

All participants were exposed to two practice blocks of 18 trials each before the onset of the experiment. Short rest breaks occurred between any two experimental blocks. During these breaks, participants were given feedback about their performance during the previous block, and informed about how much bonus money they had earned so far. This amount was computed for each block based on both accuracy and speed. A longer rest break of about 7 minutes occurred after 12 experimental blocks.

All participants were subsequently asked to perform a continuous generation task during which they were required to predict the location at which the next stimulus would appear. This generation task consisted of 540 trials presented over 3 blocks of 180 trials each. For each participant, the stimulus material presented during generation was identical with the material they had been exposed to during blocks 13 to 15 of the RT task, thereby ensuring that the RT and generation tasks were as comparable as possible. No explicit feedback was provided during the generation task. However, participants could obtain feedback merely by comparing their prediction responses with the actual location at which the next stimulus appeared.

Stimulus material

The material used in this experiment was the same as described by Lee (1997). The stimulus set consisted of the 720 (6!) sequences of six elements that were consistent with the following simple constraint: Each element could only appear once in each sequence, as described in the introduction. Each of the 24 training blocks was produced by randomly selecting (without replacement) 30 legal sequences and by concatenating them in random order with the only constraint that the last element of any sequence could not be identical with the first element of the next sequence. This procedure ensured that the material was altogether free of repetitions, which are known to elicit fast reaction times regardless of their probability. Each participant was exposed to a different random order of the 24 training blocks. In contrast to Lee (1997) then, participants in this study were exposed to the entire training set rather than to a subset of all legal sequences.

Results

We first report on overall performance in the RT task.

Choice Reaction Time Performance

To assess whether participants were sensitive to the sequential structure of the material, we first examined whether their reaction times reflected the serial position effect described by Lee (1997). Recall that the stimulus material was such that stimuli appearing as the first element of a sequence were completely random according to the generation rules, and that stimuli appearing at subsequent serial positions were increasingly predictable, up to serial position 6 where the stimulus was completely determined. Figure 1 shows the average reaction times obtained over the entire experiment, plotted separately for each of the 6 serial positions. The figure makes it clear that participants' responses are strongly influenced by the serial position within each sequence: The reaction times indeed decrease linearly from the first to the sixth serial position, with a difference of about 30 msec between the first and last serial positions. These impressions were confirmed by a two-way ANOVA with block [24 levels] and serial position [six levels] as repeated measures factors. This analysis revealed a significant main effect of block, $F(23, 253) = 48.14$, $p<.0001$, Mse = 2997.7 and of serial position, $F(5, 55) = 22.26$, $p<.0001$, Mse = 1554.3. The interaction also reached significance, $F(115, 1265) = 1.26$, $p<.05$, Mse = 754.4, albeit more detailed analyses (see below) do not confirm that is should be taken as an indication of learning. Further, a trend analysis applied to the average reaction times collapsed over serial positions 1 to 6 confirmed their linearity, $F(5, 55) = 22.26$, $p<.0001$, Mse = 1554.3 ($R^2 = 0.96$).

Learning

Figure 2 (left panel) shows how the serial position effect described above changes over training. The figure indicates that the effect is already present early in training, and that the slope of the curves corresponding to different moments during training does not appear to change much. These impressions were confirmed by an ANOVA with block [4 levels] and serial position [6 levels] applied on this

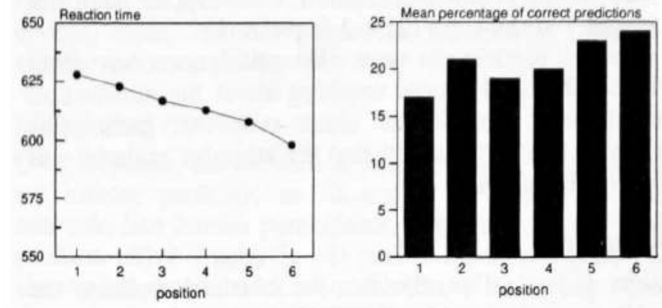

Figure 1: *Left panel*: Mean reaction times as a function of serial position over the entire experiment. *Right panel*: Mean percentage of correct prediction responses produced in the continuous generation task, plotted separately for stimuli associated with serial positions 1 to 6.

aggregate data, and which again produced a significant interaction between blocks and positions, $F(15, 165) = 1.77$, $p<.05$, Mse = 185.3. However, this significant interaction doesn't seem to reflect learning. Indeed, planned comparisons showed that the difference between reaction times to position 1 and position 6 stimuli (30 msec) is already significant over the first two blocks [$F(1,11) = 16.87$, $p<.001$, Mse = 309.5], and that it stays relatively constant up until the last two blocks, for which it averages 41 msec [$F(1,11) = 23.56$, $p<.001$, Mse = 443.74]. In short, there is in fact very little evidence that there is any learning in this situation, short of unspecific practice effects: The serial position effect emerges very early in training and remains quite constant over the entire experiment.

Generation task performance

After the main RT task, participants were asked to perform a continuous generation task during which they were to predict where the stimulus would appear next.

Figure 1 (right panel) shows the average percentages of correct predictions plotted separately for each of the 6 serial positions and averaged over the 3 blocks. Planned comparisons showed that the percentages of correct responses for each position failed to differ from chance (18%) ($p>.10$), thus suggesting that participants were unable to project their knowledge in this direct test. However, further analysis showed that participants were nevertheless more successful in predicting the location at

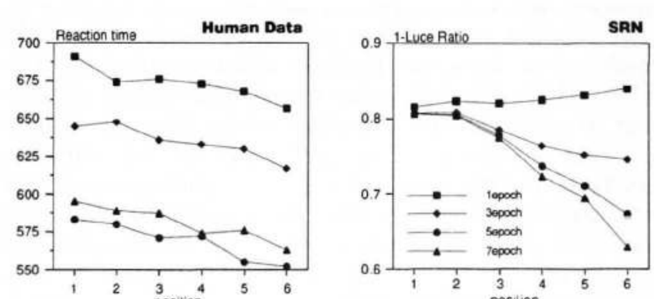

Figure 2: *Left panel*: Mean reaction times as a function of serial position, plotted separately for blocs 1–6, 7–12, 13–18, and 19–24. *Right panel*: Mean SRN responses plotted separately for epochs 1, 3, 5, and 7.

which serial position 6 stimuli could appear than for position 1 stimuli, ($t(12) = 3.8$, $p < .005$).

Finally, participants were also asked questions about whether they had noticed anything about the structure of the stimulus material in either task. All participants indicated that they thought that the stimulus material was completely random.

Discussion

Taken at face value, the data we obtained replicate the results reported by Lee (1997): Participants seem to be able to somehow segment the stimulus material in chunks of 6 elements. This is what prompted Lee to conclude that parsing mechanisms were necessary to understand performance. In addition, and somewhat surprisingly, our data also indicate that no or little sequential learning took place in this situation. A reexamination of Lee (1997)'s results likewise seems to suggest that the serial position effect is already present very early during training. In the following, we will (1) show that parsing mechanisms of any kind turn out to be unnecessary to account for performance, and (2) suggest an account of how participant's sensitivity to the sequential structure may develop before they are first exposed to the task.

The position effect is emergent. Lee (1997)'s analysis rests on the assumption that it is necessary for participants to encode the serial position for them to exhibit faster reaction times to serial position 6 stimuli than to other positions, and therefore to somehow parse the material in successive chunks of 6 elements with the correct boundaries. This, however, needs not be the case: Participants in fact merely need to be sensitive to the lag that separates two occurrences of the same stimulus, and to produce faster responses to stimuli associated with a long lag. To see this, consider the fact that any stimulus that occurs on serial position 6, that is, as the final element of an experimenter-generated sequence, is necessarily associated with a lag of at least length 5, in that, by construction, the same stimulus could not have occurred within the same experimenter-generated sequence. In contrast, stimuli that occur on serial position 1 could have previously occurred as recently as two trials ago (in the previous sequence), and thus be associated with a lag of length 1. This state of affairs is depicted in Figure 3A, which clearly shows that the different serial positions are associated with ranges of lags of increasing length. For instance, position 1 is associated with lags of length 1–5, and position 6 with lags of length 5–10. From this perspective, then, the position effect described by Lee (1997) and replicated in this experiment, merely emerges out of more elementary features of the material, namely (1) that on each trial, the probability of any stimulus increases linearly with the lag that separates the current trial from the stimulus's previous occurrence, and (2) that different serial positions in the experimenter-generated sequences are, by construction, associated with distributions of increasing lags.

If our account is correct, then one should observe a lag effect in the data. Figure 3B shows that such a lag effect is indeed present: Reaction times decrease linearly with the lag, as confirmed by an ANOVA with lag [10 levels],

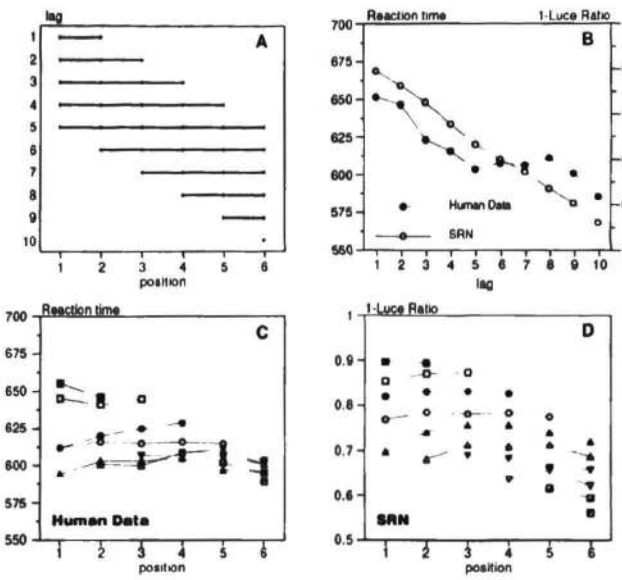

Figure 3: The lag effect. **A:** Distribution of lags for each serial position. **B:** Mean reaction times (filled symbols) and mean SRN responses (open symbols) for each of 10 lag levels. **C:** Mean reaction times, and **D:** Mean SRN responses plotted separately for the 10 lag levels and for each serial position. See text for additional details.

$F(9,99) = 25{,}45$, $p<.0001$. Figure 3C further shows that position, by itself, seems to have little impact on performance: Each curve (corresponding to stimuli with a given lag length as in Figure 3A) is relatively flat across serial positions. An ANOVA with position [six levels] applied to these data and restricted to stimuli with a lag of length 5 (the only case where position and lag are completely crossed) confirmed this impression and showed no significant effect of position ($p=.09$). Hence it should be clear that participants do not need to, and in fact do not, parse the material in order to exhibit the observed serial position effect. Combined with the further fact that we failed to observe learning in this situation, it would appear that sensitivity to the serial position, far from indicating learning of the sequential regularities, may in fact reflect knowledge that participants already possess before being exposed to the task. This knowledge may consist of a tendency to prepare responses that have not been used recently, in a way similar to the well-known fact that spontaneously generated random sequences are in fact much more uniform than true random distributions. How might this knowledge be established? This is the issue we focus on in the rest of this paper.

Simulations

To find out whether simple associative learning mechanisms are in fact sufficient to account for the data, we explored how well the SRN model (Figure 4) could learn Lee (1997)'s material. The network uses back-propagation to learn to predict the next element of a sequence based only on the current element and on a representation of the temporal context that the network has elaborated itself. To do so, it uses information provided by so-called context units which, on every step, contain a copy of the network's hidden unit activation vector at the

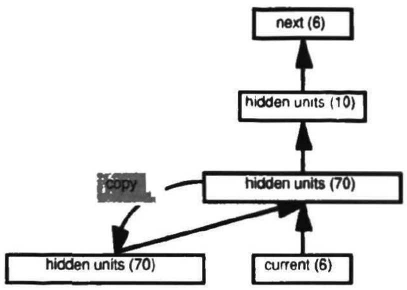

Figure 4: The simple recurrent network (SRN). See text for details.

previous time step. Over training, the relative activation of the output units representing each possible successor come to approximate the optimal conditional probabilities associated with their appearance in the current context, and can thus be interpreted as representing implicit preparation for the next element when the network is used as a model of human sequence learning performance. Previous work (see Cleeremans & McClelland, 1991; Cleeremans, 1993) has shown that the SRN is able to account for about 80% of the variance in sequential choice reaction time data.

Simulation parameters and procedure

To assess how well the SRN could capture RT performance in this experiment, we trained the model on the same material as human subjects. The network consisted of 80 hidden units and local representations on both the input and output pools (i.e., each unit corresponded to one of the 6 stimuli). The network was trained to predict each element of a continuous sequence of stimuli generated in exactly the same conditions as for human subjects. On each step, a stimulus was generated according to the generation rules and presented to the network by setting the activation of the corresponding input unit to 1.0. Activation was then allowed to spread to the other units of the network, and the error between its response and the actual successor of the current stimulus was used to modify the connection weights. During training, the activation of each output unit was recorded on every trial and transformed into Luce ratios to normalize the responses. For the purpose of comparing simulated and observed responses, we assumed (1) that the normalized activations of the output units represent response tendencies, and (2) that there is a linear reduction in RT proportional to the relative strength of the unit corresponding to the correct response. The network's responses were finally subtracted from 1.0 to make increases in response strength compatible with reduction in RT.

We conducted extensive exploration of the parameter space using this and other closely related architectures. Because it appears that human participants do not learn much beyond unspecific practice effects in this experiment, we did not attempt to match the number of experimental and simulated trials. The results presented below provided the best fit we could obtain with the human data, and they involved 7 epochs of training on the entire training set, a learning rate of 0.04 and a momentum of 0.9. Ten networks initialized with different random weights selected in the -0.5, 0.5 range were trained on a total of 30240 (720 sequences x 6 elements x 7 epochs) trials, and their responses, assessed as described above, were averaged together.

Simulation results

Over 7 epochs, the network is able to master the training set almost perfectly, as illustrated in Figure 5. The network, like human participants, exhibits a linear serial position effect (Figure 5, left panel). Further, Figure 3B shows that the network also exhibits a linear lag effect that is not only very similar to the human data, but also remarkably similar to — and indeed almost identical with — the actual distribution of lags over the 6 serial positions within the stimulus material (compare Figures 3A and 3D). A regression analysis using the simulated data depicted in Figure 3D and the corresponding human data shown in Figure 3C indicated that the model explains about 70% of the variance (see Figure 5, right panel) of human reaction times. Thus in all respects, the network provides an excellent descriptive account of the human data. In the next section, we examine how the network learns the stimulus material.

Learning. As shown in Figure 2 (right panel), over the course of training, the network progressively starts exhibiting (after 3 epochs) the serial position effect that human participants already produce at the onset of the experiment. Note that the network does not capture unspecific practice effects — a known limitation of the SRN as a model of choice reaction time performance. Figure 6 provides a more detailed view of the network's performance as it changes over training, and shows that the network becomes progressively able to predict perfectly which elements are possible at each serial position (bottom row). For instance, the network perfectly predicts that '6' is the only possible successor of "12345" As described in Servan-Schreiber, Cleeremans & McClelland, (1991), the development of sequence knowledge in the SRN involves a gradually increasing sensitivity to the sequential constraints contained in an increasingly large and self-developed representation of the temporal context defined by previous elements of the sequence. Initially, the network learns to associate each element with the distribution of its possible successors, and essentially ignores the context information. In this material, each element is associated with a unique

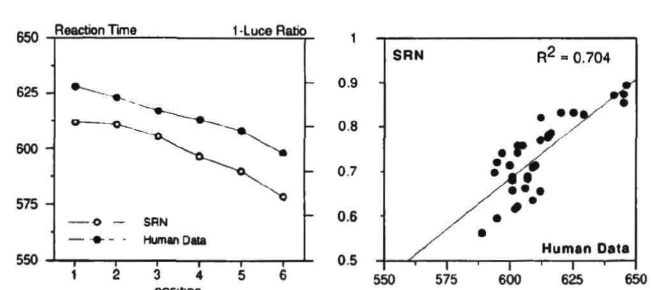

Figure 5: Comparison between simulated and human performance. *Left panel*: Both human participants and the model exhibit a linear serial position effect. *Right Panel*: The model accounts for about 70% of the variance in the distribution of human reaction times.

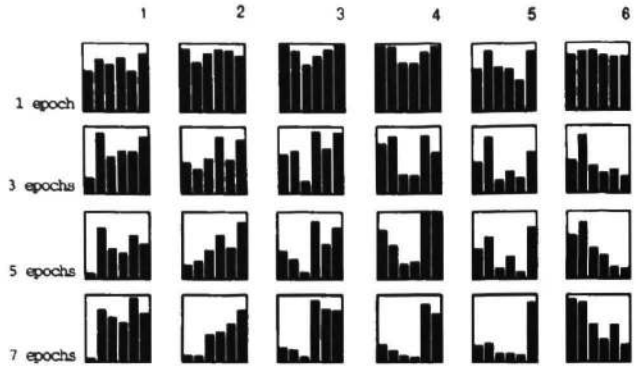

Figure 6: Model prediction responses on the sequence 123456 at different points in training. Each bar represents the strength of the output unit corresponding to each of the 6 possible elements.

distribution of successors because, by construction, an element cannot be followed by itself. Hence after one epoch of training (see Figure 6, top row), the network tends to predict that all the elements but the input element are possible successors, and the patterns of activation over its hidden units now represent these associations. When fed back onto the context units, these patterns can now be used by the network as a representation of the previous element, and it can then start basing its predictions on two elements. It is when the network has become sensitive to three elements that the material's structure starts conveying information about the lag that separates occurrences of the same stimulus. Consider for instance the fragment '123'. It can never be followed by 3 by construction, but it is also more often associated with '1' as a successor than it is with '2', regardless of the serial position at which it ends. This is simply because there are more ways for '1231' to occur in the stimulus set than there are ways for '1232' to occur. Indeed, whereas neither '1231' and '1232' can occur within any legal sequence, '1231' can span two legal sequences in three different ways ('1-231', '12-31', and '123-1'), whereas '1232' can only do so in two different ways ('12-32', and '123-2'). Hence, the lag effect emerges out of the network's prediction-based sensitivity to the statistical structure of the material, and the lag effect is itself the basis for the emerging serial position effect characteristic of human performance. Finally, further simulation work that we cannot present due to lack of space indicated that the network's representations of the material are sufficiently abstract to enable it to generalize flawlessly after training on only a subset of the 720 legal sequences.

General Discussion

In this paper, we suggested that the core mechanism involved in sequence learning is statistical in nature, and rooted in the development of distributed representations of the temporal context acquired through elementary associative learning processes that operate on exemplars. We showed how such mechanisms are in fact sufficient to understand how sensitivity to very abstract features of the material, such the serial position effect described by Lee (1997) can emerge out of a sensitivity to more elementary features of the material, such as the lag that separates successive occurrences of the same stimulus. Parsing mechanisms of any kind are thus clearly unnecessary. More surprisingly, perhaps, our results suggest that Lee (1997)'s findings do not involve learning of the sequential regularities, but merely reflect knowledge that participants already possess before being exposed to the task. This knowledge, perhaps gained through experience in the real world, may consist of a tendency to preferentially prepare responses that have not been produced recently. The structure of Lee (1997)'s material would then simply be congruent with this bias and reinforce it. This bias does appear to be implicit, although further research is necessary to clarify this issue. In conclusion, it appears that performance in this task might be more of matter of knowing without learning than a matter of learning without knowing.

Authors note

Axel Cleeremans is a Research Associate of the National Fund for Scientific Research (Belgium). Arnaud Destrebecqz is a scientific collaborator of the National Fund for Scientific Research (Belgium). Maud Boyer is supported by a grant from the Université Libre de Bruxelles in support of IUAP program #P/4-19 This work was also supported by FRFC grant #2.4605.95 F to Axel Cleeremans. We thank Pierre Perruchet and Bob French for insightful comments on the interpretation of the data.

References

Berry, D.C., and Dienes, Z. (1993). *Implicit Learning: Theoretical and empirical issues*. Hove, UK: Lawrence Erlbaum Associates.

Cleeremans, A. (1993). *Mechanisms of Implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.

Cleeremans, A. & Jiménez, L. (1998). Implicit sequence learning: The truth is in the details. In M.A. Stadler & P.A. Frensch (Eds.), *Handbook of Implicit Learning*. Thousand Oaks, CA: Sage Publications.

Cleeremans, A. & McClelland, J.L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235-253.

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

Lee, Y.S. (1997). Learning and awareness in the serial reaction time. In *Proceedings of the nineteenth. Annual Conference of the Cognitive Science Society* (pp. 119-124). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lewicki, Hill, & Bizot (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, *20*, 24-37.

Nissen, M.J. & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1-32.

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.

Servan-Schreiber, D., Cleeremans, A. & McClelland, J.L. (1991). Graded State Machines: The representation of temporal contingencies in simple recurrent networks. *Machine Learning* , *7*, 161–193.

Stadler, M.A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 318–327.