Models of Sequential Learning

Paul J. Reber, Daniel J. Sanchez, Eric W. Gobel
Department of Psychology
Northwestern University
preber@northwestern.edu

The expression of music, language and physical motor skills share the need to execute well-learned plans of sequential behavior. We can think of each of these as governed by a set of syntactic principles that instantiate organizational rules. From a computational perspective, it has been frequently observed that a fair amount of apparently rule-driven behavior can be captured by simple statistical learning models that identify sequential dependencies within frequently repeated sequences. Across many paradigms, the basal ganglia has been a brain region closely associated with this type of learning, suggesting that a common computational mechanism in this region may be involved in many types of syntax. Computational models that do statistical learning can be based on simple Bayesian statistical principles or simple recurrent connectionist networks. Using set of recent experiments examining skill learning in a novel task (SISL; Serial Interception Sequence Learning), successes and failures of the models to capture key learning, transfer and interference effects will be used to identify candidate mechanisms for neutrally plausible models of sequential learning. These mechanisms may eventually explain key challenges in the learning of syntax in language, music structure or sequential skills.

1 Introduction

Lashley (1951) proposed that the ability to sequence actions is a quintessential human cognitive ability. Across all domains that follow sequential structure, there are constraints on effective ordering that amount to domain-specific versions of syntactic processing. Syntax typically refers to the rule system that governs language processing, but we can think of constraints in other domains such as motor sequence planning or music as syntactic rules in those domains.

Over the past 10-15 years, increasing attention has been paid to statistical learning mechanisms as possibly playing a major role in the early development of language. Saffran et al (2003) reviewed findings that infants can extract the statistical structure of sound sequences and suggested that this ability was critical to language learning. In parallel, a large number of studies have examined statistical learning in perceptual-motor sequences in adults (e.g., Cleeremans, 1998) finding that statistical regularities are extracted incidentally and without awareness. Recently, Perruchet & Pacton (2006) reviewed both research areas and suggested that the commonalities in findings may very well reflect a common learning mechanism. Thus, the ability to implicitly extract statistical regularities from a practiced motor task may depend on the same basic neural learning mechanism that supports the acquisition of language structure from the environment.

From the perspective of the neural systems of the brain, a common mechanism for both language learning and motor learning might be surprising considering that the cortical areas supporting these tasks are distinct. However, language and motor regions both have connections to the basal ganglia in reciprocal loops (Middleton & Strick, 2000) that could potentially serve a similar function for identifying statistical regularities among experienced elements. In support of this idea, damage to the basal ganglia disrupts the ability to learn statistical structure in perceptual-motor sequence learning (Siegert et al., 2006) and also can disrupt rule-governed language use (Longworth et al., 2004). Studying the operating characteristics of the learning mechanism operating within the basal ganglia can therefore potentially indicate how syntax is learned across different cognitive domains.

2 Sequence Learning

We have recently reported a novel task for observing rapid, implicit sequence learning in the laboratory (Sanchez, Gobel & Reber, 2010). During the Serial Interception Sequence Learning (SISL) task, participants attempt to time a motor response to coincide with a cue moving into a target zone on the computer screen. Participants are not told that the cues follow a repeating sequence that requires them to make a regular repeating sequence of motor responses. However, participants exhibit knowledge of the covert repeating sequence by performing the task at higher levels of accuracy than when the cues follow a random order. This improved performance ability occurs even when participants are unaware of the fact that the cues followed a repeating sequence at all.

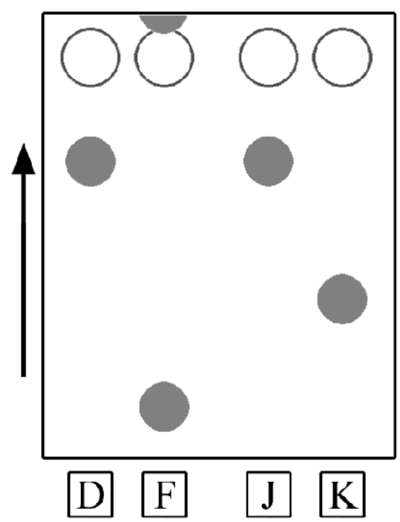

Figure 1. The Serial Interception Sequence Learning (SISL) task. Circular cues scroll vertically across a screen towards one of four target zones marked as rings. Participants press the corresponding key(s) (D, F, J, or K) on the keyboard and attempt to time their responses so that the key is pressed just as the cue moves through the target zone.

The SISL task provides an experimental technique for examining the process of statistical learning and identifying the operating characteristics of this process. In a recent report (Sanchez & Reber, submitted) we examined whether this type of learning could be extended to much longer repeating sequences than had previously been studied. Virtually all previous work on repeating perceptual-motor sequences had looked at repeating sequences with up to 12 elements, but not more. While language is typically analyzed in sentences (which have relatively low numbers of constituent element words), musical structures can be much longer. If a common mechanism supports statistical learning across domains, this mechanism should be capable of learning longer sequences.

Extending the paradigm to look at sequences that were 30, 40, 50 or 60 items long, we found robust learning of all but the 60-item sequences after a single hour-long training session. We hypothesized that one hour was insufficient practice with the longest sequences and in a second experiment examined learning of 60, 70, 80 and 90 item sequences over 2 sessions (2 hours) of training. In this second experiment, we found robust learning of 80-item sequences, although learning of 90-item sequences was only marginal (possibly reflecting insufficient time to practice the longer sequences even in 2 sessions).

While these results suggested that there is no immediate length-based capacity constraint in the human statistical learning mechanism, the underlying statistics that need to be tracked were not extremely complex. Although the training sequences were all balanced across the needed motor operators, participants could learn the sequence by acquiring first and second-order conditional statistics among motor elements that require tracking on the order of 1000 types of occurrences. While this is potentially a lot of information, it is much smaller than would be needed to track statistical relationships between language elements.

In a third experiment, as a challenge to the statistical learning capability of the human basal ganglia, we increased the amount of irrelevant noise present during training. This manipulation should dramatically reduce the learning rate of a simple statistical computational mechanism. For example, when the repeating sequence contains the fragment “DJK” repeatedly, this can be learned statistically that there is a high probability that DJ is followed by K. However, increasing irrelevant noise increases the occurrences of fragments “DJF” and “DJD” making the repeating fragment less detectable. This should lead to a lower learning rate, that is, less should be learned with each repetition of the sequence because as noise increases, there are a lot of irrelevant statistics to overcome (unlearn). Interestingly, we did not find that increasing the amount of irrelevant noise reduced the learning rate. In fact, we found the learning rate to be log-linear with practice across all the conditions in all of our learning experiments (Figure 2) regardless of sequence length or irrelevant noise.

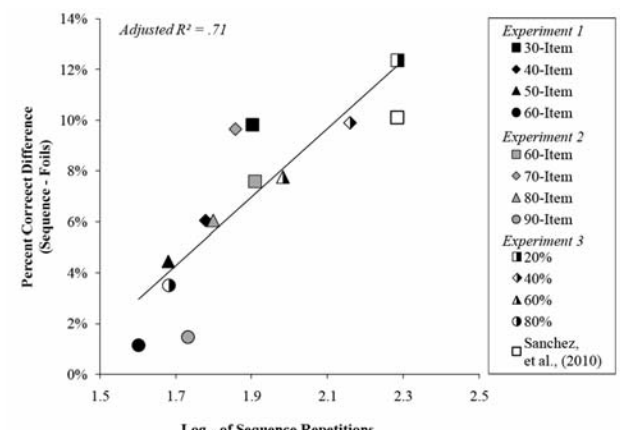

Figure 2. Scatterplot of the sequence-specific learning by $\log_{10}$ of trained sequence repetitions. The three experiments and previous data are distinguished by different marker fills. The amount of sequence knowledge expressed at test is remarkably linear with the log of the number of training repetitions of the sequence experienced.

While most previous sequence learning results suggested that a simple statistical learner could mimic human sequence learning (e.g., Cleeremans & McClelland, 1991, which used a Simple Recurrent Network to model human learning), this type of model should be slowed by experience with irrelevant statistical noise. The fact that humans do not learn more slowly in this condition suggests that the human statistical learning mechanism is augmented beyond basic Bayesian statistics and a more complex computational model of sequence learning is needed to account for learning in the basal ganglia.

A simple statistical learning model would also encounter difficulty with learning difference sequences in succession. In a recent study (Sanchez, Fraser & Reber, in preparation) we examined learning of 3 distinct sequences over a 2 day period. Participants first learned an 'A' sequence in 25 minutes of practice. The repeating sequence was changed to a 'B' sequence without any indication to the participants for the next 25 minutes of practice. Participants returned the next day to train on a 'C' sequence and then took tests of sequence knowledge for all 3 sequences. A simple statistical learning mechanism will fail to learn in this design. As shown in Figure 3, if one is just tracking the necessary statistical probabilities (Panels A & B), learning additional sequences is impaired by prior learning. The previous statistics, e.g. from sequence A, create proactive interference in trying to learn the relevant statistics for sequence B (and C). In Panel B, the effect of recency-weighting for more recent experience is shown. Learning now proceeds more normally for the B and C sequences, but this learning creates retrograde interference for the previously learned sequences (essentially catastrophic interference). Panel C shows the behavioral data and that participants acquire and retain information about all 3

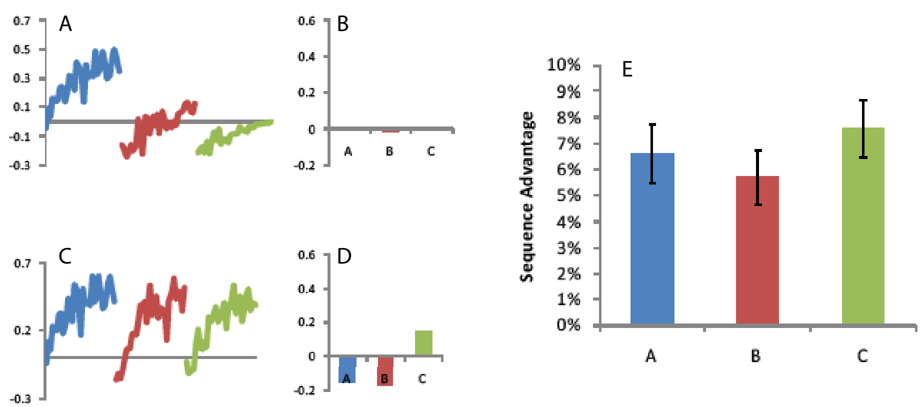

Figure 3. (Panels A, B) Proactive interference in statistical learning. After learning sequence A (blue), sequences B and C are not learned (A) or retained at test (B). (Panels C, D). Retrograde interference in statistical learning. When the learning mechanism favors more recent information, all 3 sequences appear to be learned (C) but sequences A and B are not retained to the test (D). (Panel E) Performance of human participants. Human participants learn all 3 sequences and retain them to test, challenging simple Bayesian statistical learning models.

sequences equally well.

These results suggest that a computational model that mimics learning of statistical structure in the basal ganglia needs to be more complex than just tracking minimal Bayesian statistics such as the SRN model of Cleeremans & McClelland (1991). This type of model could be augmented to better match human performance data in several different ways. The model could be extended to learn higher-orders statistics than are strictly necessary to perform the task. Some chunking models (e.g., Graybiel, 1998) could function this way. Another possibility is to use a hierarchical Bayesian model that allows for the possibility of detecting the switch to a novel sequence and changing the statistical representation to avoid interference. An important question will be which of these mechanisms is incorporated in the basal ganglia and how these mechanisms are related to learning structure in language syntax and music.

3 Abstraction

An additional important challenge to the idea that statistical learning in the basal ganglia is related to rule-learning in language is the question about the flexibility and level of abstraction in the representation of the statistics. Marcus et al. (1999) showed that infant statistical language learning extended to abstract sequences based on relational comparison among items. This type of finding cannot easily be represented in a simple connectionist or Bayesian model of learning and has been proposed to be a flaw in that theoretical approach. In our own work on sequence learning, we have found that the perceptual-motor sequence learning is extremely specific to the practiced sequence. Sequence-specific performance gains did not transfer to sequences of the same actions with slightly modified inter-response timing (Gobel, Sanchez & Reber, in press).

A crucial future direction for this area of research will be to explore the degree to which abstract, relational statistical learning can occur for motor sequences. This type of processing is fundamental to language processing. If abstract sequential information can be acquired by the same mechanisms that learn fixed repeating sequences (or fixed inter-item probabilities), then it is a viable hypothesis that there is a core syntax learning mechanism in the basal ganglia. If motor sequence learning is restricted to statistical information among fixed items, then this element of statistical learning may overlap across the domains of motor, language and music. However, language may be relatively unique in also depending on additional types of learning to support more abstract rules, e.g., about orders of syntactic categories.

4 Conclusion

The examination of learning of perceptual-motor sequences in the laboratory demonstrates some key operating characteristics of the brain's sequence learning mechanism. The mechanism by which the basal ganglia extracts statistical information automatically and implicitly from the environment shows is robust to long sequences, noisy training conditions and proceeds in a manner that is resistant to interference from other learned structures. This system may be sufficiently computationally complex to play a significant role in acquisition of syntactic structures in domains such as language and music. Important remaining questions concern the specificity and ability for this system to learn more abstract relational rules that are required for language learning.

Bibliography

Cleeremans, A. Destrebecqz A., & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, 2, 406-416.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *J Exp Psychol Gen, 120*(3), 235-253.

Gobel, E.W., Sanchez, D.J. & Reber, P.J. (in press). Integration of temporal and ordinal information during serial interception sequence learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*.

Graybiel, A.M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, 70, 119-136.

Lashley, K. S. (1951). The problem of serial order in behavior. *Cerebral mechanisms in behavior; the Hixon Symposium.* (pp. 112-146): Oxford, England: Wiley.

Longworth, C.E., Keenan, S.E., Barker, R.A., Marslen-Wilson, W.D. & Tyler, L.K. (2004). The basal ganglia and rule-governed language use: evidence from vascular and degenerative conditions. *Brain*, 128, 584-596.

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

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233-238.

Sanchez, D.J., Fraser, D. & Reber, P.J. (in preparation). Consolidation and interference effects in implicit skill learning mediated by explicit knowledge.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*(4), 110-114.

Sanchez, D.J., Gobel, E.W. & Reber, P.J. (2010). Performing the unexplainable: Implicit task performance reveals individually reliable sequence learning without explicit knowledge. *Psychonomic Bulletin & Review*, 17, 790-76.

Sanchez, D.J. & Reber, P.J. (under revision). Operating characteristics of the implicit learning system during serial interception sequence learning.

Siegert, R. J., Taylor, K. D., Weatherall, M., & Abernethy, D. A. (2006). Is implicit sequence learning impaired in Parkinson's disease? A meta-analysis. *Neuropsychology, 20*(4), 490-495.