Memory & Cognition
2007, 35 (6), 1502-1517

Characterizing sequence knowledge using online measures and hidden Markov models

Ingmar Visser, Maartje E. J. Raijmakers, and Peter C. M. Molenaar
University of Amsterdam, Amsterdam, The Netherlands

What knowledge do subjects acquire in sequence-learning experiments? How can they express that knowledge? In two sequence-learning experiments, we studied the acquisition of knowledge of complex probabilistic sequences. Using a novel experimental paradigm, we were able to compare reaction time and generation measures of sequence knowledge online. Hidden Markov models were introduced as a novel way of analyzing generation data that allowed for a characterization of sequence knowledge in terms of the grammar that was used to generate the stimulus material. The results indicated a strong correlation between the decrease in reaction times and an increase in generation performance. This pattern of results is consistent with a common knowledge base for improvement on both measures. On a more detailed level, the results indicate that at the start of training, generation performance and reaction times are uncorrelated and that this correlation increases with training.

What characteristics do learning about causes, natural language acquisition, and understanding social interactions have in common? Learning sequential dependencies between events is at the heart of each of these learning situations. In causal learning, the dependencies to be learnt are between sequences of causes and effects (Shanks, Holyoak, & Medin, 1996). In language learning, dependencies between word categories have to be acquired (McShane, 1991; Pinker, 1994). In the social behavior of both humans and animals, inferences about and applications of sequential strategies have been shown to be important (Colman, 1995). In many instances, these learning processes are incidental or implicit; they proceed without conscious effort and do not necessarily give rise to knowledge that can be easily articulated. Notably, concept formation and natural language acquisition are mostly devoid of directed and conscious efforts (Cleeremans & Jiménez, 1998). The experimental paradigm that combines both of these characteristics is implicit sequence learning, which allows one to explore these essential learning processes. The end product of these learning processes is sequence knowledge, which may be expressed in different ways. Sequence knowledge is best characterized by models for sequential dependencies, such as hidden Markov models (HMMs), and similar models, such as belief networks and Bayes nets. These are used in the analysis of language (Manning & Schütze, 1999; Miller & Chomsky, 1963), causality (Glymour, 2003), and social interactions (Colman, 1995) precisely because of their flexibility in representing sequential dependencies.

Implicit learning has been studied increasingly in different areas of psychological research, ranging from subliminal perception in social psychology (De Houwer, Baeyens, & Hendrickx, 1997) to associative learning of rules in infants (Marcus, Vijayan, Rao, & Vishton, 1999). Implicit learning is usually operationalized as an incidental, rather than an intentional, task: Subjects are not made aware of the contingencies that exist between stimuli, and they are not stimulated in any way to detect these. It is assumed that the effect of such learning conditions is that they produce implicit knowledge—that is, knowledge of which subjects are unaware. The basic results of implicit learning have been established not only in normal populations with a large age range (Marcus et al., 1999; Meulemans, Van der Linden, & Perruchet, 1998), but also in clinical populations, such as patients with Korsakoff's syndrome and amnesia (Nissen & Bullemer, 1987). These results identify implicit learning as a robust and fundamental process in the acquisition of complex knowledge and show that this process is different from explicit learning, which has been central to traditional learning research and modeling (Anderson, 1983).

Sequence learning has become the paradigm of choice in studying implicit learning. In a typical sequence-learning experiment, subjects are presented with a sequence of stimuli that is manipulated so that the order of presentation is not random, albeit unknown to the subjects. The subjects' task is simply to reproduce the current stimulus; that is, they type a unique key for each different stimulus. The behavioral effect of this manipulation is a decrease in reaction times (RTs), as compared with a control condition in which the order of the stimuli is random. Despite a large research effort in the last 15 years (Cleeremans & McClelland, 1991; Frensch, Buchner, & Lin, 1994; Jiménez & Méndez, 2001; Jiménez, Méndez, & Cleeremans, 1996; Lewicki, Czyzewska, & Hoffman, 1987; Lewicki, Hill, & Bizot, 1988; Nis-

I. Visser, i.visser@uva.nl

sen & Bullemer, 1987; Perruchet & Amorim, 1992; Seger, 1997; Shanks & Johnstone, 1999), the precise nature and extent of sequence knowledge resulting from sequence learning have remained hotly debated issues.

To gain a better understanding of the nature of sequence learning and the resulting knowledge, a number of different tasks have been proposed and used, in addition to RT measures. Their main purpose is to verify whether, indeed, implicit knowledge exists and, if so, in which respects it is distinct from explicit knowledge. Reber (1967) used verbal reports to assess subjects' awareness of the structure of memorized sequences. Since the subjects failed to express any knowledge about the presented material, Reber (1967) concluded that the learning process must be implicit and that the resulting knowledge is not open to conscious scrutiny. More recently, researchers have argued that the verbal report task is not sensitive enough to elicit explicit knowledge (Jiménez et al., 1996; Perruchet & Amorim, 1992; Shanks & Johnstone, 1999) and, therefore, have proposed other tests. Among these are the recognition and generation tasks, which are analogical to recognition and recall tasks in implicit versus explicit memory research (cf. Roediger, 1990).

In contrast with the RT task in which subjects *reproduce* the current stimulus, in the generation task subjects are required to *predict* the next stimulus or even a sequence of upcoming stimuli. Nissen and Bullemer (1987) required their subjects to guess the next stimulus at each trial, and the subjects were required to continue guessing until they guessed the correct stimulus. Using a similar generation task, Cleeremans and McClelland (1991) found that subjects score above chance level, but only slightly so. On the other hand, both Perruchet and Amorim (1992) and Shanks and Johnstone (1999) found large associations between RT performance and generation performance in their research. Given these contrasting results, the relationship between different measures of sequence knowledge remains unclear.

In this article, we will address this issue specifically. For sequence-learning research to advance, it is essential to flesh out the precise relationships between these measures and to find out how these relationships develop over the course of the learning process. The present article therefore had two aims. The first was to design an experiment that would allow us to compare generation performance and RT performance in detail throughout the experiment. The second aim was to introduce a new method of analyzing generation data that allows for quantification of knowledge expressed in the generation task. This method is based on Markov models (Manning & Schütze, 1999; Miller & Chomsky, 1963; Wickens, 1982). In the next section, different measures of implicit and explicit knowledge will be reviewed. Then, the HMM will be introduced as a model for analyzing generation data. In two experiments, a novel generation task was presented and the results were analyzed. Finally, the results will be discussed.

Measuring Sequence Knowledge

In the 3 decades since Reber (1967) used verbal reports of subjects to assess their awareness of grammatical structure, many different measures of sequence knowledge have been used. In this section, we will discuss these different measures and their advantages and disadvantages in measuring sequence knowledge.

A frequently used alternative to verbal reporting as a measure of sequence knowledge is the generation task. The main result found by Nissen and Bullemer (1987), who introduced the generation task, is that subjects' performance is significantly above chance level (59% vs. a chance level of 33%) immediately after the RT phase of a sequence-learning experiment. They compared learning under single- and dual-task conditions, but they did not contrast generation performance on structured versus random sequences without a distractor task, making it difficult to draw conclusions about the relationships between these measures.

Cleeremans and McClelland (1991), in their first experiment, used verbal reports to assess awareness of stimulus contingencies. These reports showed that subjects had limited reportable knowledge. In their second experiment, they used a generation task, modeled after the Nissen and Bullemer (1987) task. Subjects were told that the sequence they had seen was characterized by some regularity and were then asked at each trial to predict the location of the next stimulus. Feedback was provided at each error they made, after which the next trial was presented. Qualitatively, the results were similar to those in Nissen and Bullemer, in that the subjects could better predict grammatical than nongrammatical trials. However, even for grammatical trials, the percentage of correctly predicted stimuli was only about 25% versus a chance level of 16.7%. This may, in part, be due to there being six possible stimuli (vs. four in Nissen & Bullemer's study) and the fact that the sequential structure of the stimuli was more complex.

When used to evaluate explicit sequence knowledge, the generation task in Nissen and Bullemer (1987) and Cleeremans and McClelland (1991) has the important drawback that feedback is provided. Feedback has two possibly detrimental effects. First, it may result in learning during the generation task, instead of measuring knowledge that was learned during the RT phase of the experiment. In Cleeremans and McClelland's study, there was a small, although nonsignificant, increase in the percentage of correct predictions during the generation task. Because of this learning effect, only the start of the generation task can be used to assess knowledge that was acquired during the RT phase of the experiment, which limits the reliability of the results. The second important drawback of providing feedback is that it may interfere with subjects' memory when trying to predict imminent stimuli.

Perruchet and Amorim (1992) introduced the so-called *free generation* task to eliminate the undesirable effects of providing feedback and, so, to arrive at a better test of explicit sequence knowledge. Using a 10-trial repeating sequence, they found that substantial sequence knowledge can be expressed in free generation after only 200 trials. They found a close correspondence between free generation performance and improvement in RT as support for "the assumption that there is a common knowledge base for RT improvement and introspective knowledge" (Perruchet & Amorim, 1992, p. 789). In contrast, Destrebecqz and Cleeremans (2001) used a free generation task with the specific instruction that subjects should refrain from typing sequences they had seen in the RT phase of the experiment. Despite this instruction, the subjects produced significant portions of the 12-element sequence to which they had been exposed before. From this result, Destrebecqz and Cleeremans argued that free generation performance is, at least in part, dependent on implicit knowledge. It should be noted here that Wilkinson and Shanks (2004) failed to replicate the results of Destrebecqz and Cleeremans, suggesting that generation performance is fully dependent on explicit knowledge. Given these opposing results, it remains unclear whether generation performance is based on implicit or explicit knowledge. One concern with these studies is that the short repeating sequences that were used as stimulus material may be easy to memorize during both the RT phase and the generation phase. The use of such stimulus material may hence facilitate explicit learning.

Jiménez et al. (1996) used probabilistic sequences generated from a finite state grammar to overcome this problem. Such sequences are much more variable and complex than the repeating sequences that are typically used in sequence learning. It should also be noted that the use of probabilistic sequences is more realistic when the aim is to provide insight into the learning processes required in complex domains, such as, for example, natural language acquisition and causal learning. Jiménez et al. used *continuous* generation, "in which the next stimulus as prescribed by the sequential structure is presented regardless of participants' prediction response" (p. 952). This procedure is identical to that of Cleeremans and McClelland (1991), except that no feedback is provided. This procedure was chosen so as to maximize compatibility with the RT task. Jiménez et al. argued that such compatibility is necessary to ensure maximal sensitivity of the task to explicit knowledge. Even so, they found that subjects scored only slightly (although significantly) above chance level (about 25% correct vs. a chance level of 16.7%). Note that these results are identical to the results found by Cleeremans and McClelland, except that the latter study found a slight (nonsignificant) increase in generation performance over three blocks of generation trials. Even though no direct feedback is given, the continuous generation procedure provides indirect feedback by presenting the next element of the sequence regardless of subjects' responses. This indirect feedback may disrupt memory of previous stimuli. One may think that feedback, be it direct or indirect, should result in better performance. However, remembering previous stimuli is disrupted by feedback to an incorrect response (Shanks & Perruchet, 2002). As a consequence, overall performance may get worse, certainly when higher order dependencies between generated trials are considered. As a result of this procedure, only single-trial predictions can be used for analysis, which does not provide a detailed picture of subjects' higher order knowledge. For this reason, in the present study, neither direct feedback nor indirect feedback was given on generation trials.

One further important problem with the generation task is shared by all of the versions above that have been used hitherto: They have been administered only *after* the RT phase was completed (Cleeremans & McClelland, 1991; Jiménez et al., 1996; Nissen & Bullemer, 1987; Perruchet & Amorim, 1992). This procedure has two important drawbacks. First—for example, in the experiment by Nissen and Bullemer—subjects are told that the accuracy of their responses is more important than the speed of responding. As a result, subjects may adopt a different strategy in responding to the task, as compared with the RT task. Second, as a consequence of having the generation and RT tasks in two different phases of the experiment, there is no means other than the RTs of assessing sequence knowledge in the early phases of training.

Online generation. As can be seen from the discussion above, results obtained with different versions of the generation task have been far from conclusive in assessing the relationships between different measures of expressing sequence knowledge. In the present study, a novel version of the generation task was introduced, which is called *online generation*. In online generation, subjects are required to generate short sequences of trials, which are alternated with sequences of RT trials.

In subliminal semantic-priming research, similar considerations have led to the introduction of online prime identification trials. Usually, in priming studies, the optimal prime threshold is determined by administering a number of prime identification trials prior to the main experiment. However, concerns about whether the prime threshold could change during the experiment—for example, due to fatigue and learning effects—have led some researchers to devise a method of assessing the prime threshold online by administering prime identification trials interspersed with the (target) semantic-priming trials (Durante & Hirshman, 1994; Hirshman & Durante, 1992).

An advantage of the online generation procedure is that the generation and serial reaction time (SRT) tasks are presented concurrently. This prevents problems associated with having two different phases in the experiment—the SRT and the generation tasks—that may be differentially affected by strategic choices, forgetting, and fatigue. Moreover, the possible associations and dissociations between generation performance and RT performance can thus be studied in detail over the course of learning. No feedback is provided so as to avoid the associated problems. Feedback on errors in the RT phase is also suppressed so as to maximize congruence between the RT trials and the generation trials. Complex probabilistic stimulus sequences are used in this research to make the learning more comparable to natural learning situations. Our expectation is that, with this generation task, due to the congruence between RT and generation tasks, there will be a large association between these measures in the absence of verbally reportable knowledge.

Another important and novel aspect of the present research concerns the analysis of generation data. HMMs were introduced here as a means of directly comparing the rules underlying the sequence of stimuli presented to subjects and their generated sequences. Before presenting the experiments, a description of HMMs will be provided.

Assessing Sequence Knowledge From Generation Data

The analysis of (free) generation data, with the aim of measuring sequence or grammatical knowledge, can be done in several ways. Nissen and Bullemer (1987) computed the percentage correct for single generation trials and compared this with chance level, as did Cleeremans and McClelland (1991) and Jiménez et al. (1996). Another possibility is to compare bigrams, trigrams, and so forth of generated sequences between a group that is trained on a repeating sequence of structured stimuli and a group that is trained on random sequences or differently structured sequences (Perruchet & Amorim, 1992; Shanks & Johnstone, 1999). However, the choice for comparing, say, trigrams rather than bigrams is necessarily somewhat arbitrary. In many studies on implicit learning, finite state automata (FSAs) are used to generate sequences of stimuli (Cleeremans & McClelland, 1991; Jiménez & Méndez, 1999, 2001; Jiménez et al., 1996; Reber, 1967, 1976; Seger, 1997). Verbal reports and generation tasks are used to assess explicit knowledge in these studies. Mostly, generation data are compared with chance level[1] or with a control group. To establish an experimental effect, this is sufficient. However, these methods do not directly address the question of whether subjects have learned the rules of a grammar. The subjects' performance on the generation task was not compared directly with the grammar in any of the studies discussed above (except Jiménez et al., 1996; see the discussion below). Hence, it is interesting to learn how much of the structure of the FSA the subjects learned during the experiment.

Jiménez et al. (1996) correlated conditional probabilities of (a selection of) generated sequences up to a length of four with the conditional probabilities of the identical sequences in the grammar, thereby providing a comparison between subjects' generation performance and (aspects of) the grammar. Fitting HMMs directly to generated sequences extends this method by simultaneously modeling all the observed sequential dependencies in subjects' generated sequences, instead of doing so for sequences of different lengths separately. In other words, when HMMs are used, the complete information about the knowledge of subjects expressed in generation trials is used. This results in a model of subjects' knowledge that can be compared both qualitatively and quantitatively with the FSA that was used to generate the stimuli.

HMMs have mostly been applied in speech recognition (Rabiner, 1989; Schmidbauer et al., 1993), biological sequence (DNA, RNA) analysis (Krogh, 1998; Salzberg, Searls, & Kasif, 1998), and machine learning (Ghahramani & Jordan, 1997; Saul & Jordan, 1995). HMMs can be described in two important ways: first, as Markov models with a probabilistic response function, and second, as stochastic FSAs. The latter way of representing HMMs is important when applying them to sequence learning. For an introduction to HMMs, see Rabiner (1989). We first will describe HMMs in the formal manner, in which they are derived from Markov models, and then will compare them with FSAs.

Markov and hidden Markov models. All of the Markov models and HMMs described here are used for categorical data such as binary or polytomous responses. Both Markov models and HMMs consist of a number of states, here denoted as S_i, $i = 1, \ldots, n$. In applications in psychology, these states are commonly interpreted as knowledge states, which correspond to different sets or levels of knowledge. For example, in early models of paired associate learning, there are usually two states, called the *guessing state* and the *learned state*. Subjects start in the guessing state, in which their performance is at chance level, and after a number of trials they jump to the learned state, in which their performance is perfect. This kind of learning is, for the obvious reason, referred to as all-or-none learning (Nicolson, 1982; Wickens, 1982). The process of moving from one state to the next is modeled by transition probabilities $P(S_j \mid S_i)$, denoted here as a_{ij}, $i, j = 1, \ldots, n$.

The states S_i and the transition matrix $\mathbf{A} = \{a_{ij}\}$ together form a Markov model (see Wickens, 1982, for an overview of the use of Markov models in psychology). Markov models have been applied in many areas, including recall and recognition (Kintsch & Morris, 1965), paired associate learning (Nicolson, 1982), and conservation learning (Brainerd, 1979). In a Markov model, the response patterns associated with each state are such that we can determine with certainty in which state a subject is by his or her response. In latent or hidden Markov models, this is not the case. At each point, a subject's state is best described as a probability distribution over all the possible states of the model.

The outputs of the model are denoted by O_j, $j = 1, \ldots, m$. These outputs are also called *observation symbols* or *responses*. In the present application, we will model sequences of trials O^t, where each O^t is one of the O_js. The states and observation symbols are linked by observation probabilities, alternatively called a *response function*, $\mathbf{B} = \{b_{ij}\}$, $i = 1, \ldots, n, j = 1, \ldots, m$. The parameter b_{ij} represents the probability of observing symbol O_j in state S_i. In addition to transition and observation probabilities, there are initial state probabilities π_i that represent the probability of starting in state S_i. For example, in the all-or-none model, π_g equals 1.0 and π_l equals zero, expressing the *assumption* that subjects at the start of the experiment have no knowledge and, hence, start in the guessing state. All parameters together are denoted by $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$.

The λ parameter of the HMM can be estimated on the basis of a sequence (or a number of sequences) of observed responses, using the EM algorithm for optimization of the parameter values from Rabiner (1989). In the analysis of generation data, the sequences are the responses of different subjects on the generation trials of the experiment. We apply the aforementioned optimization procedure in fitting HMMs to the generation data and then compare the resulting models with the grammar that was used to generate the sequences for the experiments. To this end, we first need an HMM representation of the finite state grammar that we used to generate the sequences. This representation is provided below.

Finite state automata and HMMs. In the present experiments, an FSA was used to generate stimuli for sequence learning. This FSA is depicted in Figure 1A. Each state of the grammar has either one or two outgoing arrows. Consequently, when there is only a single outgoing arrow, the corresponding symbol is produced with a probability of 1.0; when there are two outgoing arrows, one of them is chosen with a probability of .5 (in the figure, these probabilities are represented by fat and thin arrows, respectively). Sequences are generated using this FSA by moving from state to state, starting and, eventually, terminating in State 1/7. For example, the sequence ADBD is a grammatical sequence that passes through States 1, 3, 4, 6, and 7. The associated probability of this sequence is .25—that is, the product of the individual arrow probabilities—starting with a probability of .5 for the A (the alternative choice in State 1/7 is a B); then D and B are produced with a probability of 1.0, and finally, from State 6, the D is produced with a probability of 0.5. Similarly, the sequence BCABCD passes through States 1, 2, 5, 4, 6, 5, and 7 and has an associated probability of .0625.

FSAs can also be represented as HMMs (by shifting from a vertex representation to an edge representation; that is, instead of having labeled arcs, HMMs have labeled states (see, e.g., Hopcroft, Motwani, & Ullman, 2001, and Lind & Marcus, 1995, for different representations of FSAs). In Figure 1B, the grammar is represented as an HMM (again, the thin and fat arrows represent probabilities of .5 and 1.0, respectively). The generation of strings in the HMM is very similar to the generation procedure in FSAs: Starting in a given state, which provides the first letter, one leaves the state via one of the present arcs to arrive at the next state, which provides the next letter, and so forth. Again, it can be seen that ADBD is a legal sequence, as it is in the FSA. The sequence BAC is not legal, since there is no sequence of nodes with the labels B, A, and C that are connected in that particular order.

To analyze the generation data, HMMs are used in the following way. HMMs are fitted to the (short) sequences of responses generated by subjects. This results in a model of a grammar that the subjects have learned. A grammar in this context is understood as an FSA or, rather, the HMM representation thereof. Such a grammar encompasses all the information inherent in the sequences generated by subjects—that is, the frequencies of single symbols, bigrams, trigrams, and so forth. The fitted model is then compared with the HMM representation of the grammar. This is done by computing a distance measure between the true grammar and the learned grammar. This distance measure, which is formally defined in the Results section, indicates how much of the grammar has been learned. If sequence learning has an effect on subjects' ability to generate grammatical sequences, the distance between the fitted HMMs and the grammar is expected to decrease.

EXPERIMENTS 1 AND 2
Sequence Learning and Online Generation

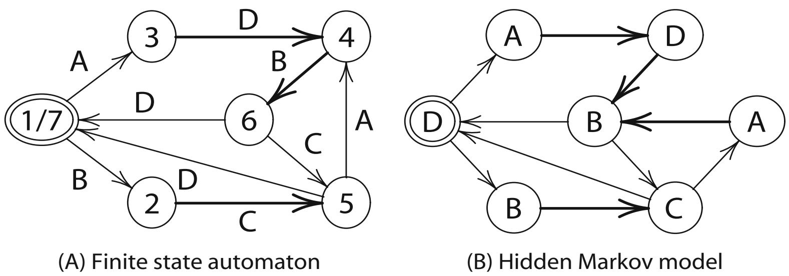
Experiment 1 was a standard sequence-learning task without a generation task, and in Experiment 2, the online generation task was introduced. The RT data from Experiment 1 could thus be used to check whether the online generation task affected RT performance. To assess the relationship between generation and RT performance during the learning process, in Experiment 2, a sequence-learning experiment was carried out, in which series of RT trials were alternated with series of generation trials. The goal of the experiment was to acquire repeated measures of generation performance and RT performance concurrently, so that these could be compared in each phase of the experiment. Moreover, the task was such that the task requirements for the generation and RT trials were as similar as possible. No feedback was given on any of the trials.

Method

In Experiments 1 and 2, subjects were given a four-choice serial RT (SRT) task consisting of a total of 9,480 and 12,000 trials, respectively. Trials were administered in four sessions of approximately 45 min each; the subjects did two sessions a day, on 2 consecutive days. On each day, the subjects had a break of 15 min between sessions. Each session comprised six blocks of 395 and 500 trials in Experiments 1 and 2, respectively. After each block, there was a (subject-controlled) break of at least 2 min. At the end of each block, the subjects' performance and financial rewards were presented on the computer screen.

Subjects

The subjects were 7 (Experiment 1) and 8 (Experiment 2) undergraduate psychology students from the University of Amsterdam. They were given course credits for their participation in the experiment. The subjects also received financial rewards for fast and ac-



(A) Finite state automaton

(B) Hidden Markov model

Figure 1. Finite state automaton (A) and hidden Markov model (B) for the same grammar. In both panels, the arrows are either thin or fat, corresponding to a probability of .5 and 1.0, respectively. See the text for further details.

curate responding. There was no financial reward for generation performance. The maximum reward was about €20.

Procedure and Stimulus Material

During both Experiments 1 and 2, the subjects were seated in front of a computer that was divided into four quadrants, as depicted on the bottom left-hand side of Figure 2. At each RT trial, an "x" was presented in one of the quadrants of the screen. The subjects' task was to press the corresponding key on the numerical keypad of a QWERTY keyboard, using the index finger of their preferred hand. The response keys were the numbers 1, 2, 4, and 5 on the numerical keypad, which have the same layout as the quadrants on the screen. Four mappings of grammatical letters to screen positions were used in a Latin-square design.

On the bottom right-hand side of Figure 2, the display used for the generation trials in Experiment 2 is depicted. At generation trials, an "x" was placed in the quadrant of the previous trial (or in the quadrant of the previous generation response if the previous trial was also a generation trial). In the other quadrants, question marks were shown to indicate that this was a generation trial. The subjects were required to press any of the three keys corresponding to the quadrants in which the question marks were shown. The rationale for indicating the previous response or trial at generation trials is that in the sequence of RT trials, no repeating trials occur. In previous studies, it was found that subjects become aware of this very quickly in the RT task (Visser, Raijmakers, & Molenaar, 2000). In this experiment, the subjects were told in the instructions that this was the case, to prevent the generation of two or more identical responses in succession.

The sequences of trials for the grammatical blocks were produced by generating sequences from the grammar in Figure 1A, as described earlier. In both experiments, each sixth block consisted of random order trials; the others consisted of grammatical trials. The only constraint in the random sequences was that there should be no repetitions of identical stimuli, as is common in sequence-learning experiments. This was done to prevent undesirable repetition priming effects in the RTs (Cleeremans & McClelland, 1991; Nissen & Bullemer, 1987). The last block of each session was used as a control in the assessment of the decrease in RTs, because the decrease of RTs was partially due to nonspecific training at the task. An additional decrease of RTs, in comparison with the random blocks, was expected due to the subjects' growing sensitivity to the contingencies inherent in the grammatical sequences.

In Experiment 2, the blocks of 500 trials were divided into sequences of RT trials and generation trials according to the following scheme (see also Figure 2). Each block consisted of runs of SRT trials with lengths of 17–23, alternated with runs of generation trials with lengths of 3–7. There were 19 such runs, resulting in 105 generation trials per block. Each block began and ended with a run of SRT trials, resulting in a total of 20 runs of RT trials, totaling 395 trials (see Figure 2 for an overview of the relationships between sessions, blocks, and runs of trials).

Instructions

In both experiments, the subjects were told that both accuracy and speed were important in this task. In addition, the subjects in Experiment 2 were instructed that during generation trials, they should "continue pressing at approximately the same rate as during the RT trials." This was done so as to make the generation trials as similar as possible to the RT trials. They were told not to "stop and think" but, instead, just to "type the response that seemed appropriate, or to guess when they felt there was no appropriate response." The subjects were also made aware of the fact that in the RT trials, no repetitions occurred and that at generation trials, they should not produce consecutive identical responses.

Postexperimental Interviews

The subjects were interviewed following the last session to establish the extent to which they could articulate the knowledge they had acquired. They were asked a series of increasingly specific questions. First, they were asked whether they had any idea what the experiment was about. Second, they were asked whether they had

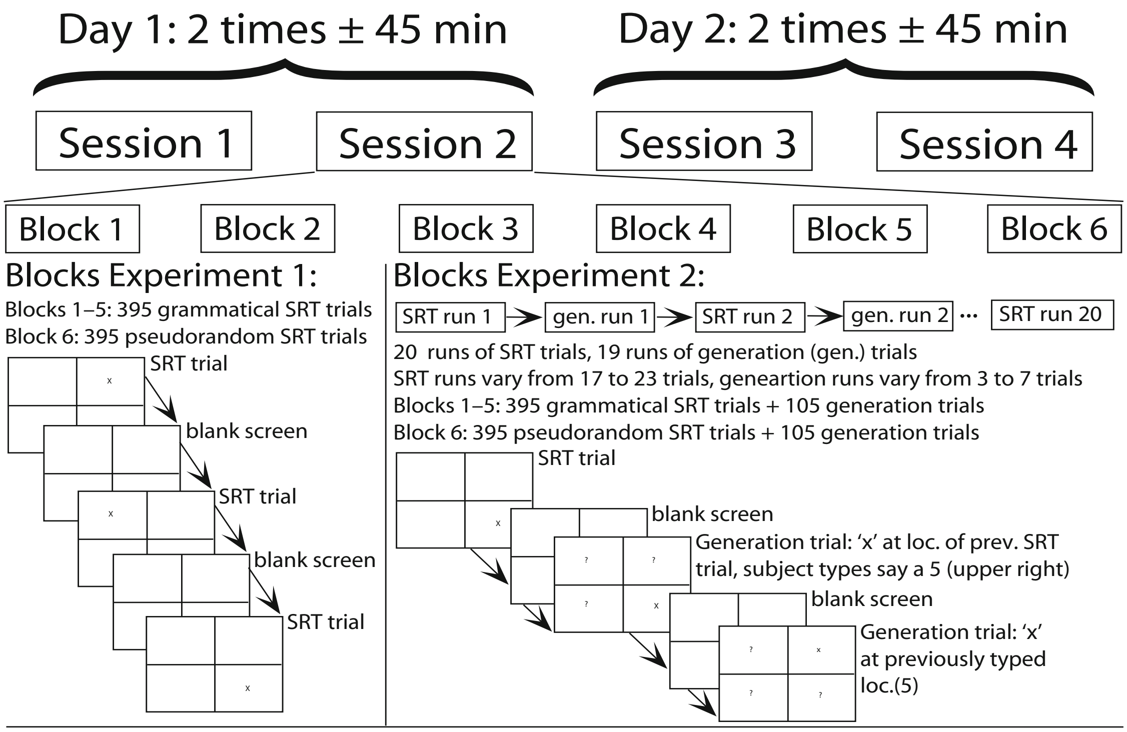

Day 1: 2 times ± 45 min
Day 2: 2 times ± 45 min
Session 1
Session 2
Session 3
Session 4
Block 1
Block 2
Block 3
Block 4
Block 5
Block 6
Blocks Experiment 1:
Blocks 1–5: 395 grammatical SRT trials
Block 6: 395 pseudorandom SRT trials
SRT trial
blank screen
SRT trial
blank screen
SRT trial
Blocks Experiment 2:
SRT run 1
gen. run 1
SRT run 2
gen. run 2
...
SRT run 20
20 runs of SRT trials, 19 runs of generation (gen.) trials
SRT runs vary from 17 to 23 trials, genеartion runs vary from 3 to 7 trials
Blocks 1–5: 395 grammatical SRT trials + 105 generation trials
Block 6: 395 pseudorandom SRT trials + 105 generation trials
SRT trial
blank screen
Generation trial: 'x' at loc. of prev. SRT trial, subject types say a 5 (upper right)
blank screen
Generation trial: 'x' at previously typed loc.(5)

Figure 2. Design of Experiments 1 and 2. SRT, serial reaction time.

noted anything particular about the stimuli. Third, they were asked whether they had found or seen any regularity in the stimuli and, if so, whether they could describe or point out that regularity. Fourth, after the subjects had been told that there was regularity in the sequences, they were asked to reproduce sequences that they thought they had seen if they could. Fifth, if they produced fragments of the sequence, they were asked at which point during the experiment they had first become aware of this. Finally, the subjects were asked whether they had used their knowledge in generating sequences at generation trials.

Results

Accuracy Data and Outliers

RTs were judged to be outliers if the RT deviated more than three standard deviations from the mean. Outlier detection was done for each subject and block separately, after the error responses had been removed. In Experiment 2, the first trial after each run of generation trials proved to be prone to errors, and the subjects responded much more slowly on these trials than on others. Outlier detection was done after removing these trials and other erroneous trials. In Experiment 1, 679 trials were detected as outliers from a total of 66,360 RT trials. The mean accuracy in the RT trials was 96.4%, ranging between 95.5% and 98.9% over the 24 experimental blocks. In Experiment 2, 731 trials from a total of 76,000 RT trials were detected as outliers. Accuracy on the RT trials averaged 97%, ranging between 96% and 98% over 24 blocks. In the following analyses, the data of 1 subject from Experiment 1 were discarded because, in one of the blocks, his mean RT was 100 msec larger than in the preceding and following blocks; using the above criteria between blocks, instead of within, this block as a whole was an outlier. The RT analyses below were conducted with and without this subject, leading to the same results.

Reaction Times

In Figure 3, the mean RTs in both experiments are shown. As can be seen, the pattern of results was very similar for both experiments, and the differences in mean RTs between Experiments 1 and 2 were generally small, with a mean absolute difference of 5.8 msec. The mapping between letters of the grammar and screen positions was varied between subjects. There were no significant differences between these mappings when included in the following analyses of data from either of the experiments (all ps $> .4$). In the following, the multivariate approach to repeated measurement analysis (O'Brien & Kaiser, 1985) was used whenever possible and unless noted otherwise.[2] However, overall, the results were the same when the univariate statistics were considered.

A two-way ANOVA with repeated measures (grammatical vs. random × four levels of practice for each of the four sessions) was conducted on the last two blocks from each session to compare the mean RTs obtained in the grammatical and the random blocks. In Experiment 1, this analysis yielded significant main effects for practice [$F(3,3) = 13.270$, $p < .05$, $\eta_p^2 = .930$] and for grammaticality [$F(1,5) = 26.989$, $p < .005$, $\eta_p^2 = .844$]. The same analyses for Experiment 2 revealed significant main effects for practice [$F(3,5) = 17.685$, $p < .005$, $\eta_p^2 = .914$]

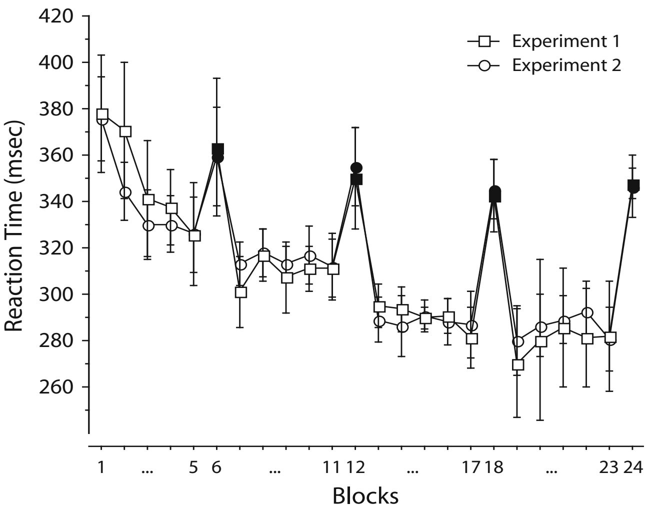


Figure 3. Mean reaction times for each experimental block for both experiments. Bars around the means indicate within-subjects confidence intervals (Loftus & Masson, 1994). Every sixth block is a block with random sequences, indicated by the filled dots/squares.

and for grammaticality [$F(1,7) = 41.056$, $p < .001$, $\eta_p^2 = .854$]. Most important, there were significant interactions between practice and grammaticality, confirming that RTs in the grammatical blocks decreased significantly more than did RTs in the random blocks [$F(3,3) = 43.369$, $p < .01$, $\eta_p^2 = .977$, and $F(3,5) = 42.421$, $p = .001$, $\eta_p^2 = .962$, in Experiments 1 and 2, respectively]. To test whether there were any differences between the experiments, the same analyses as those above were done with data from both experiments combined. The repeated measures ANOVA with grammaticality and practice as within-subjects factors and experiment as between-subjects factor ($2 \times 4 \times 2$) yielded no main effect or interaction (Fs < 1, ps $> .5$). As can be seen from Figure 3, it seemed that the RTs did not decrease anymore in Sessions 3 and 4. This observation was confirmed by a repeated measures ANOVA on the grammatical blocks of Sessions 3 and 4 combined, in which $F(9,4) = 2.297$, $p = .22$. Similarly, *within* Session 2, there was no RT improvement [$F(4,9) = 2.417$, $p = .125$]; however, there was RT improvement between Sessions 2 and 3 [$F(9,4) = 21.868$, $p < .01$].

In both experiments, there was a large decrease in mean RTs in grammatical blocks, but not so in the RTs in the random blocks. This pattern of results replicates the standard findings in implicit-learning experiments, which use probabilistic sequences (Cleeremans & McClelland, 1991; Jiménez et al., 1996; Seger, 1997). There were clear interactions between practice and grammaticality in the ANOVAs above, indicating that the grammatical sequences facilitated larger decreases in RTs than did the random sequences over practice. Most important, the generation task that was used in Experiment 2 did not influence the subjects' performance on the RTs in any significant way, as is clear from the absence of any significant effects when experiment was included as a factor in the analyses above.

Generation Data

All the following analyses concern only the data from the generation task obtained in Experiment 2. To be able to compare the results of the online generation task with those of an earlier generation experiment by Visser et al. (2000), the responses on the first trial of each run of generation trials were analyzed to see whether the subjects were able to predict the next trial in a sequence. Trials were scored as correct if they were predicted according to the sequence used to generate the trials, and as incorrect otherwise.[3] The mean percentages correct were computed for each two consecutive blocks of trials (percentages for single blocks would be not very reliable, due to the small number of data points involved). Prediction accuracy increased from 39% ($SD = 11.0\%$) in the first two grammatical blocks to 61% ($SD = 7.6\%$) in the last two grammatical blocks. Note that the baseline accuracy was 33%, since there were four choices but repetitions did not occur and the subjects were made aware of this in the instructions. In the first block, the subjects had 35% correct predictions on the first trial of each run of generation trials. A repeated measures ANOVA with 10 levels of practice (for two consecutive blocks at a time) revealed that the increase in prediction ability was significant [$F(9,63) = 9.938$, $p < .001$]. The results agree well with earlier findings where the percentage correct on single predictions increased from 36% to 52.2% over a total of 4,800 trials (Visser et al., 2000).

Fitting HMMs: Procedure. To gain insight into the rules that subjects follow during generation trials, HMMs were used to analyze the generated sequences. The procedure was as follows. Generated sequences obtained in each block were analyzed separately. The generated sequences of all the subjects in a single block formed 1 data set, resulting in a total of 24 data sets to be analyzed. For the fitted models to be comparable to the grammar, they had to contain all the transitions between states that were part of the grammar. Therefore, as a basis, the HMM representation for the grammar in Figure 1B was used. In the HMM representation of the grammar, many transition probabilities were zero. For example, there was no direct transition from the leftmost D to the rightmost A, nor vice versa from A to D. Of course, the subjects did not follow all the rules of the grammar, so the model had to be able to accommodate other sequences. This was achieved by setting all the transitions between states to nonzero values at the start of optimization. The observation parameters were fixed at the values from the grammar. Furthermore, for technical reasons, the transitions in the HMMs that occurred in the grammar were constrained so that they could not become zero in the optimization of the models.[4] This was necessary because, for a model in which transition probabilities between states were zero for grammatical transitions, it was impossible to compute the distance between the fitted model and the grammar. The reason for this will be explained below.

In fitting the model to each data set, 300 sets of starting values for the parameters were generated, and the resulting HMMs were optimized. From those 300 models, the best model was selected using an adjusted Bayesian information criterion (BIC). The BIC is defined as BIC $= -2 \log L + \log(T)p$, where L is the likelihood of the model, T is the number of data points used in fitting the model, and p is the number of freely estimated parameters (see Bozdogan, 2000, for an overview of different model selection criteria). The number of parameters p in the BIC is usually the number of freely estimated parameters. In this case, that would be all the parameters in the transition matrix **A** and all the initial state parameters π_i. Since the model had seven states and each row of the transition matrix sums to 1, the transition matrix had $7 \times (7 - 1) = 42$ free parameters. The observation matrix parameters were all fixed, and so they did not contribute to the number of parameters to be estimated. The 7 initial state parameters also summed to 1, and so 6 initial state parameters remained to be estimated, resulting in a total of 48 parameters to be estimated. In general, in fitting HMMs to data generated by FSAs or similar processes, many parameters are expected to be zero. In fact, in the HMM representation of the grammar, only 10 parameters are nonzero. Here, an adjusted BIC was used employing the number of nonzero parameters, instead of the number of free parameters. In simulation studies, it has been found that this criterion works well in selecting the correct model (Visser, Raijmakers, & Molenaar, 2002).

The resulting fitted HMMs were compared with the HMM representation of the grammar by computing a distance measure between the fitted model and the grammar. The expectation was that the distance between the subjects' models and the grammar would decrease due to learning. The distances were computed as follows:

$$D = \left[\log P\left(O_t \mid \lambda_f\right) - \log P\left(O_t \mid \lambda_t\right) \right] / T, \tag{1}$$

where O_t, $t = 1 \ldots T$ is a sequence generated by the true model, T is the length of the sequence, $\log P(O_t \mid \lambda_t)$ is the log-likelihood of the sequence O_t given the parameter values of the true model λ_t (i.e., the HMM representation of the grammar), and $\log P(O_t \mid \lambda_f)$ is the log-likelihood of the sequence O_t given the parameter values λ_f of the fitted model. This distance measure indicates how well the fitted model can describe data that are generated from the grammar, in comparison with how well the grammar itself does so. The distance measure can be interpreted as a cross-entropy between the models (see chap. 4 in Whittaker, 1990, for an introduction to entropy and information distance measures).[5]

Fitting HMMs: Results. HMMs were fitted on the generation trials of the 24 blocks in Experiment 2. In Figure 4, the models are shown for the first grammatical block of the first session and for the last grammatical block of the fourth session. Figure 4 clearly reveals the improvement. Connections that also occur in the grammar have a higher probability (indicated by thicker lines), and connections that do not occur in the grammar are less pronounced or absent in the model from the last block. In the model for Block 1, there are 11 ungrammatical transitions, whereas in the model for Block 23, there are only 8 ungrammatical connections, which have smaller probabilities.

For all models, distances to the grammar were computed using the formula in Equation 1. The distances of

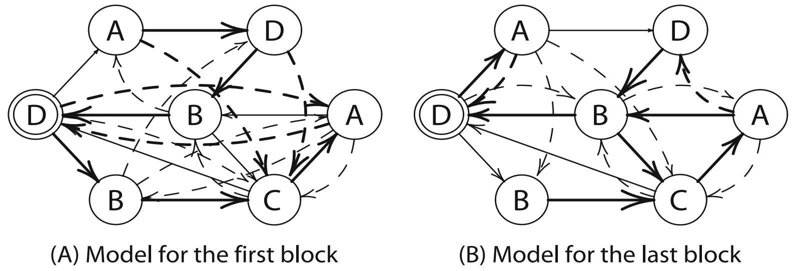


Figure 4. Fitted models for Block 1 (A) and Block 23 (B). The solid lines with arrows are connections that occur in the grammar; the dashed lines are connections that are ungrammatical—that is, those connections have a zero probability in the grammar, but nonzero probability in the models fitted on the subjects' data. The thickness of the lines correspond with the transition probabilities between states. Nongrammatical arrows with a probability of less than .1 are left out for reasons of clarity.

the models shown in Figure 4 are .619 and .371, respectively. The resulting distances are plotted in Figure 5. The pattern of results for the distances is similar to the results for the RTs that are shown in Figure 3. The distances in the random blocks are lower than one would expect on the basis of chance-level performance, which indicates that also in those blocks, the subjects generated sequences that deviated from randomness in the direction consistent with the grammar. The correlation between the mean RTs for each block and the distances for each block is .835 ($p <$ .0001). When the random blocks are left out, this correlation is .814 ($p < .0001$). There was a large effect of learning in the first two sessions: The distance drops from .619 to .485, which is consistent with the drop in RTs during those sessions. The correlation between distances and RTs over the grammatical blocks of Sessions 1 and 2 is .792 ($p < .01$). After that, in Sessions 3 and 4, the RTs hardly decreased; the mean RT in Block 13 was 286 msec, and in Block 23, it was 278 msec (a nonsignificant difference, as reported above in the section on RTs). As a consequence, the correlation between RTs and distances over grammatical blocks in Sessions 3 and 4 equals .065 ($p = .86$). The distances, however, decreased from .499 to .371 in Sessions 3 and 4. A regression analysis of distance on block number confirmed that the drop in distance within Session 3 was significant ($R = -.888, p < .05$). The similar regression for Session 4 did not reach significance ($R = -.750, p = .144$).

It is possible that much simpler models would fit the generation data as well or better. The models that we fitted to the generation data were constrained so as to be comparable with the grammar. As a result, it is possible that this model was overparametrized. To rule out this possibility, we fitted first-order Markov models to the generation data.[6] These are four-state models with a single state for each possible observation: A, B, C, or D. The transition probabilities were then optimized as in the analyses above. The fitted models had BICs in the range of 2,050–2,300 for the 24 experimental blocks. In comparison, all the models reported above, which were used for the distance analysis, had BICs below 1,950, thus showing that these models fit the data much better than did the simpler models.

Postexperimental Interviews

In the exit interviews, the subjects were asked a series of increasingly specific questions to elicit their explicit knowledge about the sequence. At the third or fourth question, 7 of the 8 subjects mentioned some sequences that they thought had occurred in the sequences. These 7 subjects mentioned 17 sequences with a mean length of 3.7, of which 14 were legal sequences according to the grammar and 3 were illegal. Of course, the subjects could only point out sequences on the screen or keyboard, since they did not know the labels from the grammar. The trigrams that were mentioned by more than 1 subject were DAD and ABC. A wrong trigram that was mentioned twice was DAB. The latter trigram corresponds to the loop on the right-hand side of the grammar in Figure 1B. All the subjects said that they had not made use of this knowledge in the generation task. Two subjects said that they had tried to do so at some stage but had found it "easier and less tiring" to trust their "feeling" or "intuition."

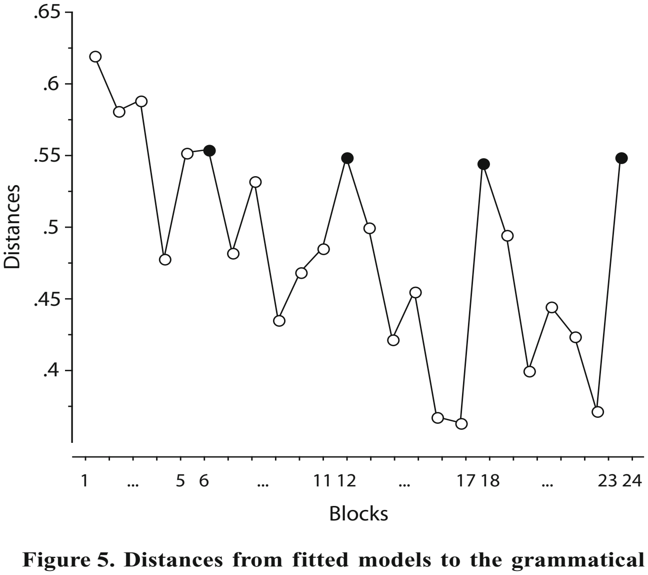


Figure 5. Distances from fitted models to the grammatical model. The random blocks are represented by filled dots.

To establish how much knowledge the sequences that were mentioned represent, an HMM was fitted to those sequences. This was done in the same way as described before with the generation data. The distance from the fitted model to the grammar was computed. The result is a distance of .706, which is larger than the distance of the models fitted to the sequences generated in both the grammatical and the random blocks. It is, in fact, close to the baseline for the distance measure, which equals .72. This baseline distance is computed by generating a random sequence without repetitions, to which an HMM is fitted in the manner described above. Next, the distance from this fitted model to the grammar is computed. The conclusion from this analysis is that the knowledge expressed in the verbal reports is severely limited, in comparison with the knowledge expressed in the generation task.

From the results above, it is clear that there is a strong association between RTs and performance on the generation task. The analyses presented so far, however, show only a global correspondence between the RT and generation tasks. In the next section, we will provide analyses of associations and dissociations between RTs and generation data on a more detailed level.

Relating RTs, Generation Data, and Hidden Markov Models

On the basis of the fitted HMMs above, we showed that there was a close correspondence between learning to express grammatical knowledge in the generation task and the decrease in RTs. Our aim in this section is to relate RTs and generation data in more detail. The important question remained as to whether particular sequences of responses that were generated more often also elicited faster responses on RT trials and vice versa. Following similar analyses by Jiménez et al. (1996) and Perruchet and Amorim (1992), we analyzed RTs to sequences of different lengths. Moreover, we analyzed the relationship between parameters of the fitted HMMs and the RTs to find out whether the HMMs would better capture the variability in RTs than would other statistics derived directly from the generated sequences. We believed that HMMs should do better because, in the fitted HMMs, the information from sequences of different lengths was combined to estimate the parameters. In contrast, when analyzing the relationship between RTs and subsequences, one necessarily has to analyze *either* bigrams *or* trigrams *or* higher order sequences separately.

The following analyses were done only for the grammatical blocks in Experiment 2. First, mean RTs were computed for the last trial of sequences of trials of lengths 2, 3, and 4 or for bigrams, trigrams, and quadruples of trials, respectively. For example, the mean RT on an A trial, which came after CD, was 346.8 msec in Session 1, whereas in the last session, this mean had decreased to 333.0 msec. Second, the conditional probabilities of the final trials of generated bigrams, trigrams, and quadruples were computed for each session. For example, the conditional probability of the subjects generating an A after CD was .26 in Session 1 and .44 in Session 4. Finally, each RT trial was coupled with a specific transition probability from the fitted HMMs.[7] These transition probabilities were then averaged for each bigram, trigram, and quadruple of trials in Sessions 1–4. For example, the mean transition probability corresponding to an A after CD was .36 in Session 1 and .43 in Session 4. The resulting values for quadruples of trials in Session 4 are presented in Table 1. In the first row of the table, the quadruple concerned is ABCD. The mean RT on the last trial of this quadruple is 356.5 msec. The mean transition probability from the fitted HMMs equals .262, and the conditional probability in the generated sequences is .218. This quadruple is not very likely to be produced and has a long associated RT. The quadruples that involve the sequence DAD, on the other hand, have shorter associated RTs (e.g., BDAD, 222.6 msec; CDAD, 227.3 msec) and higher generation and transition probabilities. As can be seen in the table, low HMM transition probabilities correspond with high RTs, and high transition probabilities with low RTs. It can also be seen that the five lowest transition probabilities from the HMMs correspond with RTs well above 300 msec, whereas the five highest transition probabilities correspond with RTs well below 300 msec. Note also that this correspondence does not hold for the generated conditional probabilities.

Table 1
Quadruples (Quad), Transition Probabilities (Trans), Generated Conditional Probabilities (Gen), and Reaction Times (RTs, in Milliseconds)

Quad	Trans	Gen	RT
ABCD	.262	.218	356.5
DBCD	.270	.260	324.9
CABD	.273	.318	376.5
ADBD	.291	.141	350.7
DBDB	.372	.427	360.1
DBDA	.374	.213	282.8
ABDB	.381	.347	349.2
DBCA	.410	.498	310.0
BCDB	.427	.372	301.6
BCDA	.433	.416	333.2
ABDA	.437	.278	312.6
BCAB	.495	.324	202.0
DADB	.522	.374	195.1
CDBC	.531	.447	255.6
BDAD	.540	.247	222.6
CDAD	.560	.354	227.3
BDBC	.594	.167	264.9
CABC	.658	.371	266.7
ADBC	.724	.446	282.8

Note—Entries are sorted from low to high transition probabilities.

The final step in establishing the relationships between RTs, generated sequences, and the fitted HMMs was to compute the correlations between the RTs, the generated conditional probabilities, and the mean transition parameters. For each session (1–4) and for each sequence length (2–4), we computed two correlations: the correlation between RTs and the generated conditional probabilities, and the correlation between RTs and the mean transition parameters. The resulting correlations are plotted in Figure 6 for bigrams, trigrams, and quadruples separately.

A number of things are noteworthy about these results. First, for the bigrams and trigrams in Session 4, the correlations between conditional probabilities and RTs and between RTs and transition parameters are similar to each

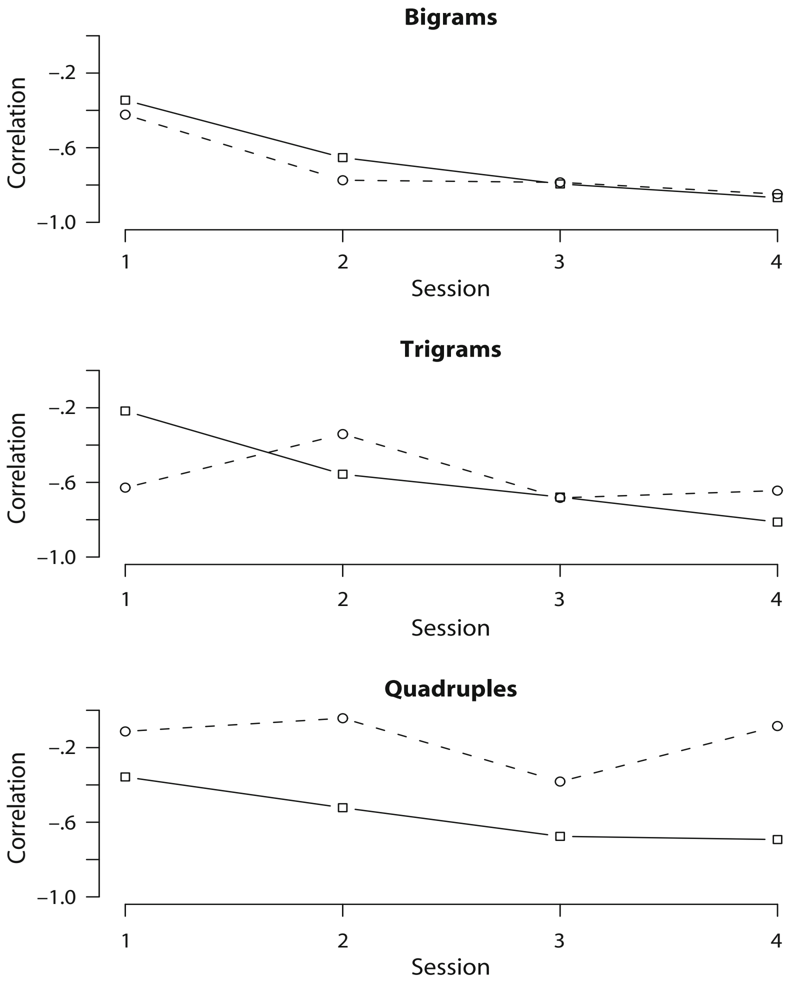


Figure 6. Correlations between reaction times and conditional probabilities of generated sequences (dashed lines) and between reaction times and mean transition probabilities (solid lines) for bigrams (upper panel), trigrams (middle panel), and quadruples of trials (lower panel) for Sessions 1–4. See the text for details.

other and to those found in other research. For the trigrams in Session 4, these correlations are $-.64$ ($p < .05$) for the conditional probabilities and $-.81$ ($p < .005$) for the transition parameters. This is very similar to the results of Perruchet and Amorim (1992), for example, who found a correlation of $-.79$ for trigrams at the end of training. The corresponding correlations for the bigrams are $-.85$ and $-.87$ (both *p*s $< .05$), respectively.

Second, for the quadruples, there is a marked difference between these correlations: The HMM transition parameters are more correlated with the RTs than are the generated conditional probabilities. In Session 4, the correlation between generated conditional probabilities and RTs is $-.084$ ($p = .73$), whereas the corresponding correlation for the HMM parameters is $-.693$ ($p < .005$). The differences between the correlations are $\Delta_1 = .24$, $p = .16$; $\Delta_2 = .48, p = .019$; $\Delta_3 = .29, p = .071$; and $\Delta_4 = .61$, $p = .0047$, for Sessions 1–4, respectively.

Third, it should be noted here that similar analyses, such as those in Jiménez et al. (1996) and Perruchet and Amorim (1992), were done using data from a generation task after the RT phase of the experiment had been completed; that is, subjects' knowledge was measured at the end and then compared with their RT performance from the start to the end of the learning process. At this point, our analyses diverge from theirs. Our analyses reveal that the magnitude of the correlation between HMM transition parameters and RTs *increases* with training. The use of the online generation task made this analysis possible. For example, for the trigrams, the correlation between RTs and HMM transition parameters for Session 1 is only $-.21$ (n.s.), whereas it increases to $-.81$ for Session 4 ($p < .005$). The differences between the correlations for Sessions 1 and 4 for the HMM transition parameters are $\Delta_b = .52$, $p = .086$; $\Delta_t = .59$, $p = .026$; and $\Delta_q = .336$, $p = .087$, for bigrams, trigrams, and quadruples, respectively, with marginal significance for the bigrams and quadruples and significance for the trigrams.

Fourth, the global correlations between HMM distances and RTs show a reverse pattern when compared with the correlations on a detailed level: The correlations between HMM distances and RT were rather high for Sessions 1 and 2, and they were comparatively lower for Sessions 3 and 4. In contrast, the correlations between HMM parameters and RTs based on sequences of trials increase in magnitude with training.

DISCUSSION

In sequence learning, differences in RTs on grammatical and ungrammatical trials are seen as the main indicator of the effect of manipulating the sequential structure of the stimuli. The verbal report task has been used as a measure of explicit knowledge, but it has been disqualified, because of its alleged lack of sensitivity (Jiménez et al., 1996; Perruchet & Amorim, 1992). Other measures of sequence knowledge, such as the normal, free, and continuous generation tasks, have been used, but their interpretation and validity have been the subject of debate. In the present study, we introduced an *online* generation task to overcome two major problems with generation tasks. First, memory and task set requirements are very different in (free) generation from those in the RT task (Shanks & Perruchet, 2002), but less so in online generation. Second, online generation provides the possibility of comparing generation performance and RT performance during the entire learning process, whereas other generation tasks are always administered at the end of training.

HMMs were introduced as a novel way of analyzing generation data. The main advantage of using HMMs is the possibility of directly comparing subjects' generation ability with the grammar underlying the sequences in the RT task. This was done by computing the distance between fitted HMMs and the grammar.

The results indicate that RT and online generation performance show a close global association. The correlation between RTs and HMM distances was found to be very high. This correlation concerns the acquisition process of responding more quickly to grammatical trials and the process of getting better at generating grammatical sequences. More detailed analyses revealed that the correlation between RTs and transition parameters is small (and nonsignificant) at the start of training and increases to $-.81$ at the end of training for the trigrams. These correlations concern subjects' knowledge in each session of the experiment; hence, a low correlation is expected to be found at the start of training, because at that point the generated sequences are basically random or very close to it. The analysis of correlations between transition parameters, RTs, and conditional probabilities revealed that the transition parameters capture the variability in RTs better than do the conditional probabilities.

Sequence Representation

Apparently, the results indicate that the HMMs capture the generation data better than do the conditional probabilities. What is the implication of our results for the validity of using HMMs in analyzing generation data? As we argued in the section on FSAs and HMMs, analyzing generation data using subsequence frequencies or conditional probabilities, as was done by Perruchet and Amorim (1992), Jiménez et al. (1996), and others, has an inherent arbitrariness. The choice to analyze either bigrams or trigrams, say, does not do justice to the generation data. These analyses implicitly assume that subjects learn chunks of trials of a certain length and express these in the generation task. The analyses that we presented suggest, rather, that subjects gradually grow sensitive to increasingly higher order dependencies in the stimuli, and HMMs seem to be able to capture this nicely.

The simple recurrent network (SRN) model for sequence learning (Cleeremans, 1993; Cleeremans & McClelland, 1991) incorporates the following learning mechanism: The SRN learns to become sensitive to ever higher dependencies between consecutive stimuli in a gradual manner. Just as is the SRN, HMMs are naturally suitable for modeling degrees of dependence between consecutive stimuli. In addition to that, HMMs can be optimized for a given data set by using maximum likelihood estimation of the parameters. This is not possible for SRNs, because in typical applications of the SRN in implicit learning, the models that are employed are overparametrized. For example, Cleeremans and McClelland (1991) used an SRN with 15 hidden units, resulting in a total of 270 network weights that had to be trained. In comparison, the HMMs that we fitted to the generated sequences had only about 20 parameters. Moreover, these parameters were directly interpretable as the probability that a subject would generate a certain symbol, given a specific context of previously generated trials. The weights in SRNs do not lend themselves to such direct interpretation. Given that both HMMs and SRNs can represent finite state grammars (Cleeremans, Servan-Schreiber, & McClelland, 1989), the use of HMMs to analyze generation data does not disqualify the SRN as a model of sequence learning but, rather, adds precision and interpretability.

The use of HMMs also enabled us to quantify knowledge expressed in the verbal report task. The results indicated that verbal reporting is at chance level. An explanation for the results in the verbal report task may be found in the fact that it is administered at the end of the experiment. It is possible that subjects are aware of stimulus sequences during the learning phase and that the use of concurrent verbal reports can bring out much more explicit knowledge (Ericsson & Simon, 1980; Perruchet, Vinter, & Gallego, 1997). However, the use of concurrent verbal reporting is not well suited for use in a speeded RT task, and it would be likely, therefore, to interfere with the learning processes that were the focus of the present study. Specifically, probing subjects for concurrent verbal reports during the learning phase of the experiment might lead them to search for patterns in the stimuli.

Alternatively, it may be argued that the HMM is not a robust model in the face of the limited amount of data provided by verbal reports; after all, the subjects had 14 out of 17 reported sequences correct. Therefore, another explanation for the results in the verbal report task may be found in the nature of finite state grammars. The analysis of the verbal report data points to the fact that even with a substantial number of grammatical sequences mentioned by subjects, the knowledge expressed therein is very limited vis-à-vis the finite state grammar. Which sequences of symbols are legal and which are not is only a small part of the contingencies that are encoded in such grammars. In addition to this, the relative frequencies of sequences of symbols are an essential characteristic of such grammars. The generation and RT tasks are able to capture performance differences that are based on these more subtle differences, whereas the verbal report task cannot. Adaptation of the verbal report task in such a way that subjects are pressed to indicate how often they have seen certain sequences of stimulus positions may alleviate this problem.

Even though our results support the notion of a common knowledge base for improvement in RT and generation performance, rather than separate knowledge bases, a number of dissociations were found as well. First, generation performance kept improving after the RTs ceased to decrease. This is shown by the decreasing association between RTs and distances over the four sessions of Experiment 2. Second, within sessions, the correspondence between RT and generation performance was less pronounced than the overall correspondence. In particular, as the correlations between HMM parameters and RTs showed, the correspondence between RTs and generation was very low at the start of training and increased significantly over the course of the experiment. Whereas generation performance improved within Sessions 2–4, RT performance leveled off within these sessions, and there was RT improvement *between* Sessions 2 and 3. The leveling off of the RT improvement in Sessions 3 and 4 could be due to a floor effect. It may be the case that there is a minimal time required for executing keypresses and this minimum was reached after approximately 6,000 trials in our experiment. There is no such limit for the generation task; according to the HMM distances, there is still ample room for improvement on the generation task. Moreover, fatigue may also affect RT performance more than it does generation performance.

Future Research

We showed that there is a large overall association between RT and generation performance. Taken together with the results on the verbal report task, it could be argued that generation performance is, in large part, implicit because of its association with RTs and the dissociation with verbally reported knowledge. There are two (theoretical) arguments in favor of this interpretation of the generation task. First, the SRN model, which has been successfully applied in implicit learning (Cleeremans & McClelland, 1991; Dienes, Altmann, & Gao, 1999; Jiménez et al., 1996; Reber, 1993) and related fields (Chang, Dell, Bock, & Griffin, 2000; Elman, 1993; Kinder & Shanks, 2003), predicts a close correspondence between RT and generation performance, without invoking the notion of explicit knowledge (see the discussion below about the random walk model on how to explain remaining dissociations between generation and RT performance). Second, the relationship between RT and generation performance could be similar to the relationship between language production and generation. Competent language use is not usually accompanied by the ability to express grammatical rules (Pinker, 1994), whereas expressing such rules is certainly the hallmark of explicit knowledge (Reber, 1967; see Berry & Dienes, 1993, for a similar argument). On the other hand, due to lack of sensitivity of the verbal task, no definite conclusions can be reached about this issue. Improved versions of the verbal report task should be able to shed more light on this issue. Regardless of the implicit/explicit dichotomy that is referred to here, different versions of the verbal report task could very well help clarify the extent and nature of sequence knowledge as it is gathered in sequence-learning experiments.

Chunk models, such as that proposed by Servan-Schreiber and Anderson (1990), have possibilities similar to those of HMMs for capturing sequence knowledge. In chunk models, (sub)sequences have discrete individual representations, and the learning process is conceived as a gradual increase in chunk strengths due to repeated presentations. If subjects learn discrete chunks or fragments of sequences, the generation task does not seem to be the best way to get at that knowledge. The reason for this is that in the generation task, subjects produce not only chunks but also combinations of chunks, and the subjects are likely to have either implicit or explicit knowledge about the order in which such chunks should be combined. Chunk models could be and have been adapted to this by including hierarchically organized chunks (Boucher & Dienes, 2003; Servan-Schreiber & Anderson, 1990).

Sequence knowledge possibly consists of a probability distribution over a potentially infinite number of strings or chunks. For a chunk model, this would mean representing an infinite number of chunks and their associated chunk strengths. By using HMMs, it is possible to represent such an infinite probability distribution with a limited number of parameters that can be estimated on the basis of generation data. To the best of our knowledge, such an effort has not been undertaken for chunk models. Such modeling would need to impose specific constraints on the relationships between chunk strengths, in order to be able to estimate parameters on the basis of finite data. Hence, doing that and comparing the results with those of HMMs could show whether there are principled differences between chunk models and HMMs. Comparing these models may also answer the question of whether subjects learn just bigrams or trigrams or higher order dependencies. This could be tested by computing the order of dependence statistics in the generation data (Wickens, 1982; see also Perruchet & Pacton, 2006, for a recent discussion about models for statistical-learning and chunk-based models).

The dissociations that we found may provide interesting topics for future research. When one assumes that both online generation and RTs measure a common knowledge base, the tasks may still differ in sensitivity. In comparing recognition ratings and priming effects, Shanks and Perruchet (2002) have argued that a single underlying parameter may explain the dissociations that they found. Such dissociations come about because different response processes are employed in explicit and implicit tasks. In online and free generation, a subject has to (more or less consciously) decide which response is appropriate at each trial, instead of simply reproducing the current stimulus with an appropriate keypress. It may be argued that the variability in these response processes in RT and generation trials is quite different, whereas the sequence knowledge entering the response process is identical. The sequence knowledge—that is, knowledge about sequential dependencies between stimuli—could be modeled using an SRN or an HMM. The response process could, for example, be implemented by using a diffusion process or a random walk model (Luce, 1986). A random walk model is a process model for generating responses and RTs, which are determined by two parameters: the drift rate and the distance to the boundaries (see, e.g., Lamberts, Brockdorff, & Heit, 2003, for a recent application of the random walk model in recognition memory). At each step, the drift rate determines how far and in which direction the process moves, and a response is given when one of the boundaries is crossed; the RT is related to the number of steps at which this happens. In sequence learning, the drift rate may be interpreted as subjects' anticipation about the next stimulus, which can be derived from the HMM. The boundary corresponding to the correct response could be set to a lower value at an RT trial. This would give rise to comparatively slower responses at generation trials and comparatively fast and accurate responding at RT trials. More generally, the application of latent variable models allows one to flesh out the exact relationships between implicit and explicit measurement while, at the same time, establishing their reliability and sensitivity (see Buchner & Wippich, 2000, and Meier & Perrig, 2000, for discussions of the role of reliability in implicit and explicit memory research; for similar discussions about implicit measures of attitudes, see Fazio & Olson, 2003, and Cunningham, Preacher, & Banaji, 2001).

HMMs were introduced as a means of analyzing generation data, which are especially suitable for modeling sequential dependencies. The results illustrate the usefulness of HMMs in the analysis of sequential data—in this case, generation data. HMMs provide a more detailed account of subjects' responses than do other analyses, such as regression and counting of generated trigrams (cf. Destrebecqz & Cleeremans, 2001; Jiménez et al., 1996; Nissen & Bullemer, 1987; Perruchet & Amorim, 1992). Particularly relevant to the field of sequence learning is the possibility of comparing generation data with the underlying rules of the grammar that was used to generate the stimuli. In addition to extending HMMs to model RTs, HMMs may be useful in other fields of research. In other work, we applied HMMs to the analysis of concept formation (Visser et al., 2002). Another possible application is with data from the random number generation task (Towse, 1998; Wagenaar, 1972), which has become popular in research in connection with executive functioning. HMMs can be used to provide an omnibus test for *deviances* from randomness in generated sequences of numbers. Similarly, in causal learning and evaluative conditioning (De Houwer et al., 1997; Shanks et al., 1996), HMMs may prove useful in future research as a trial-by-trial analysis tool in modeling sequential dependencies.

CONCLUSION

In two experiments, subjects were given a four-choice RT task, in which complex probabilistic sequences were presented as stimuli. The results thus shed light on the acquisition process of complex sequential material, such as natural language. The results indicate that subjects display considerable learning of complex sequential structures. HMMs were shown to capture variability in RTs at least as well as, and in some aspects better than, the use of conditional probabilities derived from generated sequences. At the same time, the HMMs were used to provide an overall statistic for comparing improvement in generation performance and RTs. This combination of characteristics—and the model's equivalence with commonly used FSAs—make the HMM a valuable tool in the field of implicit learning. The strong overall associations that we found point to a common knowledge base for performance on RT and generation trials. Furthermore, we argued that the remaining dissociations between RT and generation performance may spark interesting new research, as may the application of HMMs in implicit-learning research and related fields of inquiry.

AUTHOR NOTE

The research of I.V. was funded by a VENI grant from the Dutch Organization for Scientific Research (NWO). We thank Tessa van Schijndel for invaluable help in collecting the data for this article. We thank Conor Dolan, Han van der Maas, Colleen Kelley, and Luis Jiménez for helpful comments on an earlier draft of the manuscript. Correspondence concerning this article should be addressed to I. Visser, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands (e-mail: i.visser@uva.nl).

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Mahwah, NJ: Erlbaum.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, **27**, 807-842.

Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, **44**, 62-91.

Brainerd, C. J. (1979). Markovian interpretations of conservation learning. *Psychological Review*, **86**, 181-213.

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology*, **40**, 227-259.

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, **29**, 217-229.

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. Frensch (Eds.), *Handbook of implicit learning* (pp. 323-364). Thousand Oaks, CA: Sage.

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

Cleeremans, A., Servan-Schreiber, D., & McClelland, J. L. (1989). Finite state automata and simple recurrent networks. *Neural Computation*, **1**, 372-381.

Colman, A. M. (1995). *Game theory and its applications in the social and biological sciences*. Oxford: Butterworth-Heinemann.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability and convergent validity. *Psychological Science*, **12**, 163-170.

De Houwer, J., Baeyens, F., & Hendrickx, H. (1997). Implicit learning of evaluative associations. *Psychologica Belgica*, **37**, 115-130.

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, **8**, 343-350.

Dienes, Z., Altmann, G. T. M., & Gao, S.-J. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science*, **23**, 53-82.

Durante, R., & Hirshman, E. (1994). Retrospective priming and masked semantic priming: The interfering effects of prime activation. *Journal of Memory & Language*, **33**, 112-127.

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

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, **87**, 215-251.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, **54**, 297-327.

Frensch, P. A., Buchner, A., & Lin, J. (1994). Implicit learning of unique and ambiguous serial transitions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 567-584.

Ghahramani, Z., & Jordan, M. I. (1997). Factorial hidden Markov models. *Machine Learning*, **29**, 245-273.

Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Sciences*, **7**, 43-48.

Hirshman, E., & Durante, R. (1992). Prime identification and semantic priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 255-265.

Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2001). *Introduction to automata theory, languages, and computation* (2nd ed.). Boston: Addison-Wesley.

Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 236-259.

Jiménez, L., & Méndez, C. (2001). Implicit sequence learning with competing implicit cues. *Quarterly Journal of Experimental Psychology*, **55A**, 345-369.

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 948-969.

Kinder, A., & Shanks, D. R. (2003). Neuropsychological dissociations between priming and recognition: A single-system connectionist account. *Psychological Review*, **110**, 728-744.

Kintsch, W., & Morris, C. J. (1965). Application of a Markov model to free recall and recognition. *Journal of Experimental Psychology*, **69**, 200-206.

Krogh, A. (1998). An introduction to hidden Markov models for biological sequences. In S. L. Salzberg, D. B. Searls, & S. Kasif (Eds.), *Computational methods in molecular biology* (pp. 45-63). Amsterdam: Elsevier.

Lamberts, K., Brockdorff, N., & Heit, E. (2003). Feature-sampling and random-walk models of individual-stimulus recognition. *Journal of Experimental Psychology: General*, **132**, 351-378.

Lewicki, P., Czyzewska, M., & Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 523-530.

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

Lind, D., & Marcus, B. (1995). *An introduction to symbolic dynamics and coding*. Cambridge: Cambridge University Press.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1**, 476-490.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford: Oxford University Press.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

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

McShane, J. (1991). *Cognitive development: An information processing approach*. Oxford: Blackwell.

Meier, B., & Perrig, W. J. (2000). Low reliability of perceptual priming: Consequences of the interpretation of functional dissociations between explicit and implicit memory. *Quarterly Journal of Experimental Psychology*, **53A**, 211-233.

Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology*, **69**, 199-221.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 419-491). New York: Wiley.

Nicolson, R. I. (1982). Shades of all-or-none learning: A stimulus sampling model. *British Journal of Mathematical & Statistical Psychology*, **35**, 162-170.

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

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, **97**, 316-333.

Perruchet, P., & Amorim, M. A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 785-800.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, **20**, 233-238.

Perruchet, P., Vinter, A., & Gallego, J. (1997). Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin & Review*, **4**, 43-48.

Pinker, S. (1994). *The language instinct*. New York: Morrow.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**, 267-295.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, **6**, 317-327.

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of the instructional set. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 88-94.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Roediger, H. L., III (1990). Implicit memory: Retention without remembering. *American Psychologist*, **45**, 1043-1056.

Salzberg, S. L., Searls, D. B., & Kasif, S. (Eds.) (1998). *Computational methods in molecular biology*. Amsterdam: Elsevier.

Saul, L. K., & Jordan, M. I. (1995). Boltzmann chains and hidden Markov models. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), *Advances in neural information processing systems* (Vol. 7, pp. 435-442). Cambridge, MA: MIT Press.

Schmidbauer, O., Casacuberta, F., Castro, M. J., Hegerl, G. Höge, H., Sanchez, J. A., & Zlokarnik, I. (1993). Articulatory representation and speech technology. *Language & Speech*, **36**, 331-351.

Seger, C. A. (1997). Two forms of sequential implicit learning. *Consciousness & Cognition*, **6**, 108-131.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 592-608.

Shanks, D. R., Holyoak, K. J., & Medin, D. L. (Eds.) (1996). *Causal learning*. San Diego: Academic Press.

Shanks, D. R., & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1435-1451.

Shanks, D. R., & Perruchet, P. (2002). Dissociation between priming and recognition in the expression of sequential knowledge. *Psychonomic Bulletin & Review*, **9**, 362-367.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Towse, J. N. (1998). On random number generation and the central executive of working memory. *British Journal of Psychology*, **89**, 77-101.

Visser, I., Raijmakers, M. E. J., & Molenaar, P. C. M. (2000). Reaction times and predictions in sequence learning: A comparison. In L. A. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-*

Second Annual Conference of the Cognitive Science Society (pp. 971-976). Mahwah, NJ: Erlbaum.

Visser, I., Raijmakers, M. E. J., & Molenaar, P. C. M. (2002). Fitting hidden Markov models to psychological data. *Scientific Programming*, **10**, 185-199.

Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of the literature. *Psychological Bulletin*, **77**, 65-72.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Chichester, U.K.: Wiley.

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.

Wilkinson, L., & Shanks, D. R. (2004). Intentional control and implicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 354-369.

NOTES

1. Note that comparing performance with chance level may provide results that are hard to interpret. In doing so, one assumes that subjects can perform at this level, which is seldom found to be the case in random number generation experiments (Towse, 1998; Wagenaar, 1972).

2. The multivariate approach does not assume sphericity of the data, which is often violated in repeated measurements (O'Brien & Kaiser, 1985) and is, hence, more general (see Loftus & Masson, 1994, and Stevens, 1996, for discussions of this approach).

3. Note that in some states of the grammar, there are two possible continuations of the sequence. Hence, this way of scoring the correctness of generated trials is rather conservative, but it does make the analysis comparable with those in Visser et al. (2000).

4. In particular, those parameters were constrained to be no smaller than .002. Since each data set contains just under 500 data points, a transition probability smaller than 1/500 means that the transition does not, in fact, occur in the data. This implies that the transition probability, if it reaches this value, is not significantly different from zero. Another way to put this is to say that data sets generated from those models—that is, with or without the constraints—are indistinguishable when the data sets are of the order of magnitude of the data considered here.

5. Note that the distance measure used here is not symmetric. It measures the distance from the grammar to the fitted model, but not vice versa. In this case, that would be impossible, because, in general, the fitted model allows for ungrammatical sequences that cannot be modeled by the grammar and, hence, the log-likelihood that occurs in the equation would be infinite.

6. We thank Luis Jiménez for suggesting this analysis.

7. This coupling is done using the Viterbi algorithm. For a given sequence of symbols, this algorithm provides the states of the HMM that correspond with the sequence at each point (see Rabiner, 1989, for details of this algorithm). In this way, the estimated transition probabilities are provided for each transition between consecutive trials in the RT task.

(Manuscript received January 16, 2006;
revision accepted for publication July 12, 2006.)