Journal of Experimental Psychology:
Learning, Memory, and Cognition
2001, Vol. 27, No. 2, 483–498

0278-7393/01/$5.00 DOI: 10.1037//0278-7393.27.2.483

Implicit Learning of First-, Second-, and Third-Order Transition Probabilities

Gilbert Remillard and James M. Clark
University of Winnipeg

Most sequence-learning studies have confounded different types of information, making it difficult to know precisely what is learned. Addressing many of the confounds, the current study shows that people can learn 1st-, 2nd-, and 3rd-order transition probabilities. Measures directly assessing awareness of the probabilities show that the knowledge is implicit early in training and becomes explicit with extended training.

Implicit sequence learning has been studied using the serial reaction time (SRT) task in which people respond to the location of a target that follows a structured sequence of locations. People often learn the structure of the sequence, as evidenced by changes in reaction time (RT) and yet show no explicit knowledge of the structure as assessed by free-recall, cued-recall, or recognition tasks (e.g., Cleeremans & McClelland, 1991; Curran & Keele, 1993; Lewicki, Hill, & Bizot, 1988; McDowall, Lustig, & Parkin, 1995; Reed & Johnson, 1994; Stadler, 1989, 1993; 1995; Willingham, Nissen, & Bullemer, 1989). Unfortunately, SRT task studies often fail to identify which of a number of possible sequential constraints have been learned. Determining precisely what is learned is important for accurately assessing explicit knowledge of the information learned (Perruchet, Gallego, & Savy, 1990; Shanks, Green, & Kolodny, 1994; Shanks & St. John, 1994) and for understanding the mechanisms of implicit sequence learning (Cleeremans & Jimenez, 1998; Hoffman & Koch, 1998).

An *n*th-order transition probability, $P(E|A_n\text{-} \ldots \text{-}A_2\text{-}A_1)$, is the probability of an event E occurring on trial t given the occurrence of events $A_1, A_2, \ldots A_n$ on trials $t-1, t-2, \ldots t-n$, respectively, and is defined as the number of times that E follows the run $A_n\text{-} \ldots \text{-}A_2\text{-}A_1$ divided by the total number of times that $A_n\text{-} \ldots \text{-}A_2\text{-}A_1$ occurs. When exposed to a sequence of events, people apparently can learn first-, second-, or third-order probabilities (e.g., Cleeremans, 1993; Cleeremans & Jimenez, 1998; Heuer & Schmidtke, 1996; Howard & Howard, 1997; Jackson, Jackson, Harrison, Henderson, & Kennard, 1995; Jimenez & Mendez, 1999; Reed & Johnson, 1994; Schvaneveldt & Gomez, 1998; Stadler, 1992; 1993). For example, Cleeremans and McClelland (1991) have found that RTs tend to be shorter on transitions with higher than lower first- and second-order probabilities.

The evidence for learning of first-, second-, or third-order probabilities is not conclusive for a number of reasons, however. First, Lag 2 and second-order probabilities are often confounded. A Lag 2 probability, $P(E|A\text{-}x)$, is the probability of an event E occurring on trial t given the occurrence of event A on trial $t-2$. Often the situation is such that, for example, if $P(1|3\text{-}4) > P(2|1\text{-}3)$ then $P(1|3\text{-}x) > P(2|1\text{-}x)$.

Second, Lag 3 and third-order probabilities are, in some cases, confounded. A Lag 3 probability, $P(E|A\text{-}x\text{-}x)$, is the probability of an event E occurring on trial t given the occurrence of event A on trial $t-3$. Cleeremans and McClelland (1991) found evidence for learning of third-order probabilities, whereas Jimenez, Mendez, and Cleeremans (1996) did not. Interestingly, Lag 3 probability was a confound in the former but not the latter study.

Third, first-, second-, and third-order probabilities are often confounded. For example, if $P(1|4) > P(2|3)$, then $P(1|3\text{-}4) > P(2|1\text{-}3)$, or vice versa.

Fourth, transitions are rarely counterbalanced across probabilities. For example, if 1-3 and 1-2 had low and high first-order probabilities, respectively, the assignments would seldom be reversed, possibly confounding transition probability with finger combination and ease of execution (e.g., see footnote 1).

Fifth, number of response alternatives has been confounded with transition probability in some studies. As an example, for sequences of the form 1-2-3-2-4-3, RT on 2 following 1 and on 3 following 4, which have first-order probabilities of 1.0, is shorter than that on other transitions, which have first-order probabilities of .5 (Curran & Keele, 1993, Experiment 2; Frensch, Buchner, & Lin, 1994). However, 1 and 4 are each followed by only one location (i.e., one response alternative), whereas 2 and 3 are each followed by two possible locations (i.e., two response alternatives). Increasing the number of response alternatives tends to increase RT (Hyman, 1953; Kornblum, 1975; Miller & Ulrich, 1998).

Finally, there is the problem of differential run involvement. Responses to events (e.g., 3) and sequential pairings of events

Gilbert Remillard and James M. Clark, Department of Psychology, University of Winnipeg, Winnipeg, Manitoba, Canada.

This research was supported by a Natural Sciences and Engineering Research Council (NSERC) of Canada postgraduate scholarship to Gilbert Remillard and by an NSERC operating grant. We thank Michael Stadler and Axel Cleeremans for their comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Gilbert Remillard, Department of Psychology, University of Winnipeg, 515 Portage Avenue, Winnipeg, Manitoba R3B 2E9, Canada. Electronic mail may be sent to gremilla@uwinnipeg.ca.

(e.g., 1-3) remain primed for a short period of time, and so, RTs to the last element of runs of the form 3-1-3 and 1-3-1-3 are fast (Cleeremans & McClelland, 1991). Because high-probability transitions are more frequent than low-probability transitions, a greater proportion of high- than of low-probability transitions may complete fast runs.

Not only have first-, second-, and third-order probabilities been confounded with other factors, but awareness of the probabilities has been assessed directly in only a handful of studies (Baldwin & Kutas, 1997; Gomez, 1997, Experiment 2; Perruchet, Bigand, & Benoit-Gonin, 1997, Experiment 5; Reed & Johnson, 1994). Moreover, in some of the studies, methodological weaknesses have precluded strong conclusions. For example, Gomez (1997, Experiment 2) had people view two-letter sequences and predict the letter most likely to occur next. Predictions indicated explicit knowledge of second-order probabilities, but explicit knowledge of first-order probabilities could account for performance. Reed and Johnson (1994) assessed awareness of second-order probabilities by having people respond to sequences of two target locations and predict the target's next location. However, each prediction trial was followed by a generation task that could have influenced subsequent predictions. The present study examined implicit learning of first-, second-, and third-order probabilities by addressing the confounding factors outlined earlier and by directly assessing explicit knowledge of the probabilities.

Experiment 1

We used a four-choice SRT task to examine implicit learning of first-order probabilities of .40, .50, and .60. Location frequencies, Lag 2, and Lag 3 probabilities were held constant. Number of response alternatives was also held constant. Given any target location, there were two possible succeeding locations. To discount the possibility that learning was limited to second-, third-, or higher order probabilities that were confounded with first-order probability, discrete blocks of trials were introduced. The trials were presented in discrete sets of two, rather than in a continuous fashion, by introducing a pause after every second trial. We hypothesized that the discrete blocks would hinder the use of second-, third-, and higher order probability information. Thus, shorter RTs on higher than on lower probability transitions would be strong evidence for learning of the first-order probabilities.

The problem of differential run involvement was minimized by calculating RTs on low-, medium-, and high-probability transitions as a function of the preceding context. Eight types of five-element runs were identified on the basis of the first, middle, and last three elements of the run being an alternation (A) or nonalternation (N; see Table 1). For example, 3-4-3-1-3 is an ANA run because the first (3-4-3), middle (4-3-1), and last (3-1-3) three elements constitute an alternation, nonalternation, and alternation, respectively. By averaging across runs, RTs on low-, medium-, and high-probability transitions are equally affected by the different runs.

Short-term priming effects seem to weaken with training (Cleeremans & McClelland, 1991; Soetens, Boer, & Hueting, 1985). If short-term priming effects were present in the current experiment, RT to the last element would be shorter for AAA and NAA runs, where the last element was primed (e.g., 1-3-1 primes 3), than for ANA and NNA runs and longer for AAN and NAN runs, where the last element was different from what was primed

Table 1
Categorization of Runs

Experiments 1 and 2		Experiments 3 and 4	
Run	Example	Run	Example
AAA	3-1-3-1-3	EEE	3-2-1-3-2-1
NAA	2-1-3-1-3	UEE	4-2-1-3-2-1
ANA	3-4-3-1-3	EUE	3-5-1-3-2-1
NNA	2-4-3-1-3	UUE	4-5-1-3-2-1
AAN	2-1-2-1-3	EEU	3-2-6-3-2-1
NAN	3-1-2-1-3	UEU	4-2-6-3-2-1
ANN	2-4-2-1-3	EUU	3-5-6-3-2-1
NNN	3-4-2-1-3	UUU	4-5-6-3-2-1

Note. Five-element runs in Experiments 1 and 2 were categorized as a function of the first, middle, and last three elements constituting an alternation (A) or nonalternation (N). Six-element runs in Experiments 3 and 4 were categorized as a function of the first and fourth, second and fifth, and third and sixth elements being equal (E) or unequal (U).

(e.g., 1-2-1 primes 2 but 3 occurs), than for ANN and NNN runs. If RT differences between runs diminish with training, this would suggest a weakening of short-term priming effects. Finally, explicit knowledge of the first-order probabilities was assessed directly by having participants indicate where the target was most likely to appear next given its current location.

Method

Participants

The participants were 16 university undergraduates (12 women and 4 men) ranging in age from 18 to 35 years ($M = 22$ years).

The SRT Task

Millisecond timing was implemented on a personal computer with standard monitor and keyboard using Bovens and Brysbaert's (1990) routine. The lines marking the four horizontally arranged target locations were 0.2 cm in length and separated by intervals of 1.8 cm. The participants sat approximately 55 cm from the screen. The red-stickered *D*, *F*, *J*, and *K* response keys, on which were placed the left middle, left index, right index, and right middle fingers, corresponded to the first, second, third, and fourth target locations from the left, respectively.

On each trial, the target, a lowercase *o*, appeared above one of the four lines, and participants pressed the corresponding response key. If the correct key was pressed, the target immediately disappeared. Otherwise, the target remained in its location until the correct key was pressed. After the target disappeared, it reappeared 400 ms later. RT was measured as the time between target appearance and the first response.

There was one session on each of 3 consecutive days. Session 1 began with a practice block of 50 trials in which the sequence of target locations was random with the constraint that a target location was not repeated on successive trials. There were 12 blocks of trials in Sessions 1 and 2 and 14 blocks in Session 3. The blocks were 101 trials each, with the exception of Blocks 3, 6, 9, and 12 in Session 3. These were the discrete blocks, each containing 40 discrete sets of 2 trials. The discrete sets were separated by four *x*s that overwrote the four lines after every second trial. After 1,000 ms, the *x*s were replaced by the four lines, and 400 ms later, the target appeared. Session 3 began with a 40-trial discrete block for practice.

A performance history was provided at the end of each block. The numbers *1* to *12* or *1* to *14*, corresponding to the number of blocks in the session, appeared vertically along the side of the screen. An asterisk appeared beside numbers for discrete blocks. Beside the number for a completed block, one of two types of information was displayed. If more than 10% of responses in the block were incorrect, the message *too many errors* and the error rate were displayed. Otherwise, a horizontal line, its length representing the average RT of correct responses, and the average RT were displayed. The next block of trials was initiated by pressing a key in response to a prompt on the screen.

The Sequential Structure

For each participant, the sequence of target locations was randomly generated subject to certain constraints that held across every two blocks. Across every two blocks, first-order probabilities were as shown in Table 2. Numbering the four target locations from left to right, 1, 2, 3, and 4, respectively, $P(2|1) = .60$, $P(3|1) = .40$, and so on. Transitions with probabilities of .40, .60, and .50 were labeled *EL, EH, CL,* or *CH* for experimental low, experimental high, control low, and control high, respectively. CL and CH transitions were matched to EL and EH transitions, respectively, on the basis of being within- or between-hand and requiring a left- or right-hand response. For example, transition 2-4 was labeled *CL* because like EL transition 1-3, it is a between-hand transition requiring a right-hand response.[1] A Class (E, C) × Probability (L, H) interaction with longer RTs on EL than CL transitions and shorter RTs on EH than CH transitions (or with a greater RT difference between EL and EH transitions than between CL and CH transitions) would suggest learning of the first-order probabilities.

Across every two blocks, the target appeared in each location equally often. Lag 2, Lag 3, and probabilities of the form $P(E|A_1\text{-}A_2\text{-}x)$ were .50, for example, $P(2|3\text{-}x) = .50$, $P(2|4\text{-}x\text{-}x) = .50$, and $P(2|1\text{-}3\text{-}x) = .50$, respectively. The number of response alternatives was always two. Given any location (e.g., 1), there were two possible succeeding locations (e.g., 2 or 3). Finally, second- and third-order probabilities were redundant with first-order probabilities, thus adding no information over and above that provided by the first-order probabilities. For example, redundant second- and third-order probabilities for $P(2|1) = .60$ were $P(2|2\text{-}1) = P(2|1\text{-}2\text{-}1) = .60$. Appendix A describes how sequences were generated.

To have each of the eight transitions in Table 2 serve as an EL, EH, CL, and CH transition, four versions of Table 2 were created. One version was Table 2. A second version was created from Table 2 by exchanging EL and EH transitions and CL and CH transitions, a third by exchanging EL and CL transitions and EH and CH transitions, and a fourth by exchanging EL and CH transitions and EH and CL transitions.

Short-term priming effects. The 64 runs of five elements were each classified into one of eight types on the basis of the first, middle, and last three elements being an alternation or nonalternation (see Table 1). If priming effects are present, there should be a Middle (A, N) × Last (A, N) interaction with RT to the last element being shorter for AAA and NAA runs than for ANA and NNA runs and longer for AAN and NAN runs than for ANN and NNN runs. A Middle × Last × Session interaction with RTs on the different runs converging across sessions would suggest that short-term priming effects weakened with training.

Discrete blocks. In a discrete block, there were 40 discrete sets of two trials each. A discrete set was one of the eight runs of two in Table 2. The eight runs were presented in a random order five times. Session 3 began with a 40-trial discrete block for practice. The eight runs were presented in a random order twice followed by four randomly chosen runs.

Table 2
First-Order Probabilities in Experiment 1

Preceding target location	Next target location 1 *p*	Type	2 *p*	Type	3 *p*	Type	4 *p*	Type
1	—	—	.60	EH	.40	EL	—	—
2	.50	CH	—	—	—	—	.50	CL
3	.50	CL	—	—	—	—	.50	CH
4	—	—	.40	EL	.60	EH	—	—

Note. Dashes indicate that first-order probabilities were 0. EL = experimental low transition; EH = experimental high transition; CL = control low transition; CH = control high transition.

The Awareness Questionnaire

Items 1–4 on the questionnaire assessed explicit knowledge of the first-order probabilities associated with rows 1–4 of Table 2, respectively. For example, Item 1 asked whether the letter *o*, after appearing in Position 1, (a) went to Position 2 more often than to Position 3, (b) went to Position 3 more often than to Position 2, or (c) went to Positions 2 and 3 equally often. Two items pertained to EL-EH transitions and two pertained to CL-CH transitions. A Class (E, C) × Probability (L, H) interaction with EL transitions being chosen less often than EH transitions (out of two), and CL and CH transitions being chosen equally often (out of two) would suggest awareness of the first-order probabilities. While participants completed the questionnaire, the lines marking the target locations remained on the screen for reference. The keyboard was removed so keys could not be pressed.

Procedure

Four participants were randomly assigned to each version of Table 2. At the start of Session 1, the SRT task was described, and participants were instructed to try to improve their RT with practice while keeping their error rate below 10%. At the start of Session 3, the discrete blocks were described. After the last block of Session 3, the awareness questionnaire was administered.

Results and Discussion

For each participant, the median RT for correct responses across 10 blocks was determined as a function of transition (EL, EH, CL, CH), type of run completed (see Table 1), and session (1, 2, 3). Blocks 11 and 12 of Sessions 1 and 2 were excluded because Session 3 had only 10 blocks. RTs are displayed in Figures 1 and 2 (left panels).

An analysis of variance (ANOVA) with session (1, 2, 3), class (E, C), probability (L, H), first (A, N), middle (A, N), and last (A, N) as within-subjects factors was performed on the RT data from

[1] In Experiments 1 and 2, the within-hand transitions 1-2, 2-1, 3-4, and 4-3 were quicker than the between-hand transitions 1-3, 2-4, 3-1, and 4-2, and the difference declined with training. For example, in Experiment 2, the difference declined from 36 ms in Session 1 to 20 ms in Session 3 ($p = .014$). The Session 3 difference was significant ($p = .001$). The within-hand advantage was greater in Experiment 3 where a six-choice SRT task was used. Within-hand transitions were 1-3, 2-1, 3-2, 4-5, 5-6, and 6-4, and between-hand transitions were 1-4, 2-6, 3-5, 4-2, 5-1, and 6-3. The advantage declined from 96 ms in Session 1 to 75 ms in Session 3 ($p < .001$). The Session 3 difference was significant ($p < .001$). Thus, type of transition, within- or between-hand, can have a large impact on RTs. Hence, it is important to consider this property when matching CL and CH transitions to EL and EH transitions, respectively. With respect to left- versus right-hand responses, the latter were about 13 ms faster. The difference did not change with training nor with type of task (four- vs. six-choice SRT task).

the nondiscrete blocks. ANOVAs with class (E, C) and probability (L, H) as within-subjects factors were performed on the RT data from the discrete blocks and on the data from the awareness questionnaire. Alpha was .05, two-tailed.

Learning of First-Order Probabilities

The Class × Probability × Session interaction was not significant, $F(2, 30) = 1.03$, $MSE = 655.36$ (see Figure 1). Thus, RT differences between transitions did not change significantly across sessions. Excluding Session 1, the Class × Probability interaction was significant, $F(1, 15) = 30.01$, $MSE = 3{,}967.68$.[2] RT was longer on EL than on CL transitions, $F(1, 15) = 16.17$, $MSE = 4{,}827.16$, and shorter on EH than on CH transitions, $F(1, 15) = 31.84$, $MSE = 1{,}367.02$.

To discount the possibility that RT differences between transitions were due entirely to learning of second-, third-, or higher order probabilities, performance in the discrete blocks, where use of such information was minimized, was examined. The Class × Probability interaction was significant, $F(1, 15) = 4.87$, $MSE = 336.62$. RT was longer on EL than EH transitions, $F(1, 15) = 40.73$, $MSE = 244.91$, whereas the CL-CH difference only approached significance, $F(1, 15) = 3.87$, $MSE = 468.96$, $p = .068$. The EL-CL ($p = .198$) and EH-CH ($p = .059$) differences were not significant. The results from the discrete blocks are strong evidence for learning of the first-order probabilities.

Short-Term Priming Effects

The Middle × Last interaction, $F(1, 15) = 88.63$, $MSE = 3{,}098.54$, and Middle × Last × Session interaction, $F(2, 30) = 14.34$, $MSE = 661.40$, were significant, reflecting the longer average RT on AAN and NAN runs than on ANN and NNN runs and shorter average RT on AAA and NAA runs than on ANA and NNA runs, and their convergence with training (see Figure 2). Thus, priming effects were present and weakened with training.

Awareness of First-Order Probabilities

For each participant, it was determined how many times (out of two) an EL transition, an EH transition, and the equal option were chosen and then how many times (out of 2) a CL transition, a CH transition, and the equal option were chosen. Overall, EL, EH, CL, and CH transitions were chosen .69, .88, .69, and 1.00 times, respectively. The Class × Probability interaction, and the EL-EH and CL-CH differences were not significant, three $Fs(1, 15) < 1$. Thus there was no evidence of explicit knowledge of the first-order probabilities.

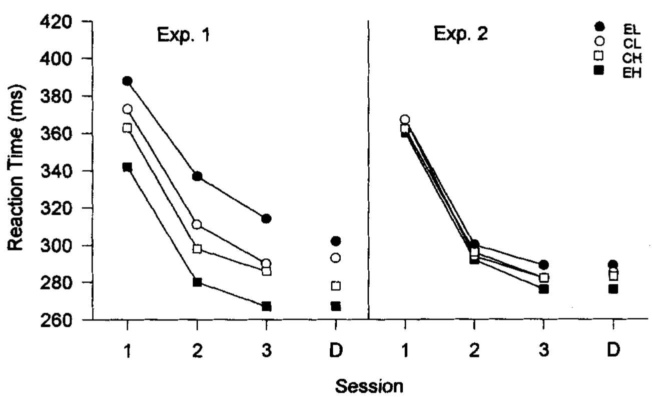


Figure 1. Reaction time, averaged across runs, as a function of experiment, session, and transition. Results from the discrete blocks (D) also are displayed. Exp. = experiment; EL = experimental low transition; CL = control low transition; CH = control high transition; EH = experimental high transition.

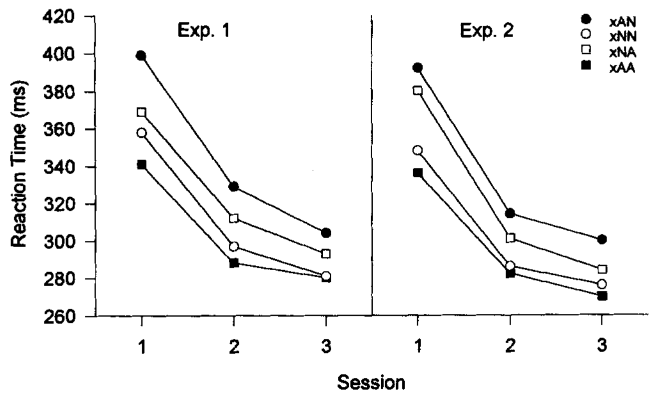


Figure 2. Reaction time (RT), averaged across transitions, as a function of experiment, session, and run. Exp. = experiment; xAN = average RT on AAN and NAN runs; xNN = average RT on ANN and NNN runs; xNA = average RT on ANA and NNA runs; xAA = average RT on AAA and NAA runs.

In conclusion, first-order probabilities of .40, .50, and .60 were learned. To show this unambiguously, the possibility that RT differences between transitions were due entirely to learning of higher order (e.g., second- or third-order) probabilities of .40, .50, and .60 had to be discounted. This was the purpose of the discrete blocks. An alternative approach would be to compare performance in Experiment 1 to that of a group receiving sequences in which higher order probabilities were .40, .50, and .60, but in which first-order probabilities were .50. If RT differences between transitions in Experiment 1 were due entirely to learning of higher order probabilities, then RT differences in the comparison group would be similar to those in Experiment 1. Experiment 2 provided a comparison group and also examined implicit learning of second-order probabilities.

Experiment 2

In Experiment 2, second- and third-order probabilities were .40, .50, and .60, and first-order probabilities were .50.

Method

Participants

The participants were 20 university undergraduates (11 women and 9 men) ranging in age from 17 to 28 years ($M = 19$ years).

---

[2] The critical Class × Probability interaction included only those sessions in which RT differences between transitions were at asymptote. Across experiments, this was Session 2 onward.

The SRT Task

The SRT task was similar to that in Experiment 1 except for the following. There was one session on each of 4 consecutive days. There were 10 blocks of trials in Sessions 1, 2, and 3 and 8 blocks in Session 4. The blocks were 122 trials each, with the exception of Blocks 5 and 8 in Session 3 and Blocks 3 and 6 in Session 4. These were the discrete blocks, each containing 48 discrete sets of 3 trials. The discrete sets were separated by four *x*s overwriting the four lines after every third trial. Session 3 began with a 48-trial discrete block for practice.

The Sequential Structure

For each participant, the sequence of target locations was randomly generated subject to certain constraints that held across every two blocks. Across every two blocks, second-order probabilities were as shown in Table 3. CL and CH transitions were matched to EL and EH transitions, respectively, in a manner analogous to that in Experiment 1. For example, transition 2-1-3 was labeled *CL* because like EL transition 1-2-4, it involved a within-hand transition (i.e., 2-1) and between-hand transition (i.e., 1-3) and required a right-hand response (i.e., 3).

Across every two blocks, the target appeared in each location equally often. First-order, Lag 2, Lag 3, and probabilities of the form $P(E|A_1\text{-}A_2\text{-}x)$ and $P(E|A_1\text{-}x\text{-}A_2)$ were .50. The number of response alternatives was always two. Given any run of two (e.g., 1-2), there were two possible succeeding locations (e.g., 1 or 4). Finally, third-order probabilities were redundant with second-order probabilities. For example, redundant third-order probabilities for $P(1|1\text{-}2) = .60$ were $P(1|2\text{-}1\text{-}2) = P(1|3\text{-}1\text{-}2) = .60$. Appendix A describes how sequences were generated. Four versions of Table 3 were created, as in Experiment 1.

Discrete blocks. In a discrete block, there were 48 discrete sets of three trials each. A discrete set was one of the 16 runs of three in Table 3. The 16 runs were presented in a random order three times. Session 3 began with a 48-trial discrete block for practice. The 16 runs were presented in random order once.

The Awareness Questionnaire

The eight-item questionnaire assessed explicit knowledge of the second-order probabilities associated with rows 1–8 of Table 3. The item for row 1 of Table 3 was a diagram of the target moving from Position 1 to Position 2 and then to Positions 1 and 4 (see Appendix B). Participants had to estimate the percentage of the time the target went to Positions 1 and 4, following its occurrence in Positions 1 and then 2, and to choose the two estimates from a list of eight numbers so that they added to 100. Four items pertained to EL-EH transitions and four to CL-CH transitions. A Class (E, C) × Probability (L, H) interaction with smaller probability estimates for EL than EH transitions (over four items) and similar estimates for CL and CH transitions (over four items) would suggest awareness of the second-order probabilities. While participants completed the questionnaire, the lines marking the target locations remained on the screen for reference. The keyboard was removed so keys could not be pressed.

Procedure

The procedure followed that of Experiment 1 except that 5 participants were randomly assigned to each version of Table 3 and the awareness questionnaire was administered after the last block of Session 4.

Results and Discussion

For each participant, the median RT for correct responses across eight blocks was determined as a function of transition (EL, EH, CL, CH), type of run completed (see Table 1), and session (1, 2, 3). Blocks 9 and 10 of Sessions 1 and 2 were excluded because Session 3 only had eight blocks. The nondiscrete blocks of Session 4 were excluded because, by the end of Session 3, the frequency of runs of three associated with each second-order probability was similar to that in Experiment 1. For example, the EH transition 1-3-4 in Experiment 2 and the run 3-1-2 in Experiment 1 where 1-2 was an EH transition each occurred about 250 times over three sessions. RTs are displayed in Figures 1 and 2 (right panels). ANOVAs were conducted as in Experiment 1.

Learning of Second-Order Probabilities

The Class × Probability × Session interaction approached significance, $F(2, 38) = 2.72$, $MSE = 336.60$, $p = .079$ (see Figure 1). Thus RT differences between transitions increased marginally across sessions, mainly from Session 1 to Session 2. Excluding Session 1, the Class × Probability × Session interaction no longer approached significance, $F(1, 19) < 1$. Importantly, the Class × Probability interaction was significant, $F(1, 19) = 7.95$, $MSE = 1{,}296.63$. RT was longer on EL than CL transitions, *F*(1,

Table 3
Second-Order Probabilities in Experiment 2

Preceding target locations	Next target location: 1		2		3		4	
	p	Type	*p*	Type	*p*	Type	*p*	Type
1-2	.60	EH	—	—	—	—	.40	EL
1-3	.40	EL	—	—	—	—	.60	EH
2-1	—	—	.50	CH	.50	CL	—	—
2-4	—	—	.50	CL	.50	CH	—	—
3-1	—	—	.50	CL	.50	CH	—	—
3-4	—	—	.50	CH	.50	CL	—	—
4-2	.40	EL	—	—	—	—	.60	EH
4-3	.60	EH	—	—	—	—	.40	EL

Note. Dashes indicate that second-order probabilities were 0. EL = experimental low transition; EH = experimental high transition; CL = control low transition; CH = control high transition.

19) = 6.10, MSE = 1,139.84, whereas the EH-CH difference was not significant, $F(1, 19) = 2.10$, MSE = 1,725.34.

To discount the possibility that RT differences between transitions were due solely to learning of third- or higher order probabilities, performance in the discrete blocks, where use of such information was minimized, was examined. The Class × Probability interaction approached significance, $F(1, 19) = 3.44$, MSE = 202.09, $p = .079$. RT was longer on EL than EH transitions, $F(1, 19) = 5.17$, MSE = 347.28, and the CL-CH difference was not significant, $F(1, 19) < 1$. The EL-CL ($p = .309$) and EH-CH ($p = .150$) differences were not significant. The results from the discrete blocks are good evidence for learning of the second-order probabilities.

Short-Term Priming Effects

The Middle × Last interaction, $F(1, 19) = 117.93$, MSE = 3,327.30, and Middle × Last × Session interaction, $F(2, 38) = 30.49$, MSE = 911.01, were significant, indicating, as in Experiment 1, that short-term priming effects were present and weakened with training (see Figure 2).

Awareness of Second-Order Probabilities

For each participant, the probability estimates (over four items) for EL, EH, CL, and CH transitions were determined. Overall, estimates were 49.4%, 50.6%, 50.2%, and 49.8%, respectively. It was also determined how many times (out of 4) EL, EH, CL, and CH transitions were chosen (i.e., given higher probability estimates). Overall, the transitions were chosen 1.50, 1.85, 1.65, and 1.70 times, respectively. For both data sets, the Class × Probability interactions and the EL-EH and CL-CH differences were not significant, six Fs$(1, 19) < 1.16$. Thus, there was no evidence of explicit knowledge of the second-order probabilities.

Comparing Experiments 1 and 2

When we introduced Experiment (1, 2) as a between-subjects factor and excluding Session 1, the Class × Probability × Experiment interaction was significant, $F(2, 34) = 26.14$, MSE = 2,475.04 (see Figure 1). The EL-CL difference was smaller in Experiment 2 than Experiment 1, $F(2, 34) = 15.36$, MSE = 2,766.60, as was the EH-CH difference, $F(2, 34) = 15.05$, MSE = 1,567.26. This suggests that RT differences in Experiment 1 were not entirely the result of learning second-, third-, or higher order probabilities. Otherwise, the differences would have been similar to those in Experiment 2. Consequently, there must have been first-order probability learning in Experiment 1.

The preceding analyses involved comparisons across experiments. Experiment 3 examined implicit learning of first-, second-, and third-order probabilities in a single experiment.

Experiment 3

A six-choice SRT task was used to examine implicit learning of first-, second-, and third-order probabilities of .40, .50, and .60. Explicit knowledge of the probabilities was assessed by having participants observe the target move across one, two, or three locations and then predict where it was most likely to appear next.

Method

Participants

The participants were 24 university undergraduates. Two participants were replaced because of very long RTs in one case and a high error rate in the other case.

The SRT Task

The SRT task was similar to that in Experiment 1 except for the following: There were six target locations. The ring, middle, and index fingers of each hand were placed on the *S, D, F, J, K,* and *L* response keys. There were three sessions, one every second day. Each session consisted of 15 blocks of 101 trials each. There were no discrete blocks. Session 1 began with a practice block of 99 trials in which the sequence of target locations was random, with the constraints that transitions that did not occur in the experimental blocks also did not occur in the practice block and first-, second-, and third-order probabilities were .50. The message *too many errors* was displayed if more than 5% of responses in a block were incorrect.

The Sequential Structure

For each participant, the sequence of target locations was randomly generated, subject to certain constraints that held across every three blocks. Across every three blocks, first-, second-, and third-order probabilities were as shown in Table 4. First-order transitions were like the transitions in Experiment 1. EL1 (e.g., 1-4) and EH1 (e.g., 1-3) transitions had first-order probabilities of .40 and .60, respectively. First-, second-, and third-order probabilities were redundant. For example, if $P(4|1) = .40$, then $P(4|2\text{-}1) = P(4|3\text{-}2\text{-}1) = .40$.

Second-order transitions were like the transitions in Experiment 2. EL2 (e.g., 3-2-1) and EH2 (e.g., 3-2-6) transitions had second-order probabilities of .40 and .60, respectively. First-order probabilities were .50, and second- and third-order probabilities were redundant. For example, if $P(1|3\text{-}2) = .40$, then $P(1|2) = .50$ and $P(1|1\text{-}3\text{-}2) = .40$. Thus second- and first-order probabilities were not confounded.

Third-order transitions EL3 (e.g., 2-1-3-5) and EH3 (e.g., 2-1-3-2) had third-order probabilities of .40 and .60, respectively. First- and second-order probabilities were .50. For example, if $P(5|2\text{-}1\text{-}3) = .40$, then $P(5|3) = P(5|1\text{-}3) = .50$. Thus, third-order probability was not confounded with first- and second-order probabilities.

Because first- and second-order probabilities in the third section of Table 4 were .50, transitions in that section were used to establish CL and CH transitions. CL1, CH1, CL2, and CH2 transitions were matched to EL1, EH1, EL2, and EH2 transitions, respectively, as in Experiments 1 and 2. Across every three blocks, 40% and 60% of all occurrences of CL1 transitions were involved in EL3 and EH3 transitions, respectively (e.g., in Table 4, the CL1 transition 4-2 can be involved in the EL3 transitions 2-1-4-2 or 2-6-4-2, or in the EH3 transitions 5-1-4-2 or 5-6-4-2). This also was the case for CH1, CL2, and CH2 transitions. Thus CL1, CH1, CL2, and CH2 transitions were all equally involved in EL3 and EH3 transitions.

Across every three blocks, the target appeared in each location equally often. Lag 2 and Lag 3 probabilities were .50. Probabilities of the form $P(E|A_1\text{-}A_2\text{-}x)$ were .48, .50, or .52. For second-order transitions, probabilities of the form $P(E|A_1\text{-}x\text{-}A_2)$ were .48 or .52. For third-order transitions, such probabilities were inadvertently confounded with third-order probabilities. For example, $P(2|2\text{-}1\text{-}3) = P(2|2\text{-}6\text{-}3) = .60$ and $P(2|2\text{-}x\text{-}3) = .60$. The confounding could have been avoided. Appendix C describes how sequences were generated.

Six versions of Table 4 were created. Version 1 was Table 4. Version 2 was created from Table 4 by exchanging EL1 and EH1, EL2 and EH2, EL3 and EH3, CL1 and CH1, and CL2 and CH2 transitions. Version 3 was created by having the top, middle, and bottom sections of Table 4 describe

Table 4
First-, Second-, and Third-Order Transitions in Experiment 3

Preceding target locations	Next target location: 1	2	3	4	5	6
			First-order transitions			
3-2-1	—	—	EH1	EL1	—	—
4-2-1	—	—	EH1	EL1	—	—
3-5-1	—	—	EH1	EL1	—	—
4-5-1	—	—	EH1	EL1	—	—
3-2-6	—	—	EL1	EH1	—	—
4-2-6	—	—	EL1	EH1	—	—
3-5-6	—	—	EL1	EH1	—	—
4-5-6	—	—	EL1	EH1	—	—
			Second-order transitions			
1-3-2	EL2	—	—	—	—	EH2
6-3-2	EL2	—	—	—	—	EH2
1-4-2	EH2	—	—	—	—	EL2
6-4-2	EH2	—	—	—	—	EL2
1-3-5	EH2	—	—	—	—	EL2
6-3-5	EH2	—	—	—	—	EL2
1-4-5	EL2	—	—	—	—	EH2
6-4-5	EL2	—	—	—	—	EH2
			Third-order and control transitions			
2-1-3	—	EH3	—	—	EL3	—
		CH1			CL1	
		CL2			CH2	
5-1-3	—	EL3	—	—	EH3	—
		CH1			CL1	
		CL2			CH2	
2-6-3	—	EH3	—	—	EL3	—
		CH1			CL1	
		CH2			CL2	
5-6-3	—	EL3	—	—	EH3	—
		CH1			CL1	
		CH2			CL2	
2-1-4	—	EL3	—	—	EH3	—
		CL1			CH1	
		CH2			CL2	
5-1-4	—	EH3	—	—	EL3	—
		CL1			CH1	
		CH2			CL2	
2-6-4	—	EL3	—	—	EH3	—
		CL1			CH1	
		CL2			CH2	
5-6-4	—	EH3	—	—	EL3	—
		CL1			CH1	
		CL2			CH2	

Note. Dashes indicate that transitions did not occur. EL1 = experimental low first-order transition; EH1 = experimental high first-order transition; EL2 = experimental low second-order transition; EH2 = experimental high second-order transition; EL3 = experimental low third-order transition; EH3 = experimental high third-order transition; CL1 = control low first-order transition; CH1 = control high first-order transition; CL2 = control low second-order transition; CH2 = control high second-order transition.

third-, first-, and second-order transitions, respectively. Version 4 was formed from Version 3 by exchanging EL1 and EH1 transitions, EL2 and EH2 transitions, and so on. Version 5 was created by having the top, middle, and bottom sections of Table 4 describe second-, third-, and first-order transitions, respectively. Version 6 was formed from Version 5 by exchanging EL1 and EH1 transitions and so on.

Short-term priming effects. The 192 six-element runs were each classified into one of eight types on the basis of the first and fourth, second and fifth, and third and sixth elements being equal (E) or unequal (U) (see Table 1). If short-term priming effects associated with bigrams (e.g., 2-1) were present, RT to the last element would be shorter for EEE and UEE runs, where a bigram was repeated and the last element primed, than for EUE and UUE runs and longer for EEU and UEU runs, where the last element was different from what was primed, than for EUU and UUU runs: yielding a Middle (E, U) × Last (E, U) interaction. A Middle × Last × Session interaction with RTs on the different runs converging across sessions would suggest that short-term priming effects weakened with training.

If priming effects associated with trigrams (e.g., 3-2-1) were present, RT to the last element would be shorter for EEE runs, where a trigram was repeated, than for UEE runs and longer for EEU runs, where the last element was different from what was primed, than for UEU runs: producing a First (E, U) × Last (E, U) interaction when we considered only EEE, UEE, EEU, and UEU runs.

The Prediction Task

Prediction Trials 1–2, 3–6, and 7–14 assessed awareness of the first-, second-, and third-order probabilities associated with EL1-EH1, EL2-EH2, and EL3-EH3 transitions, respectively. On Trials 1–2, 3–6, and 7–14, the target moved across one, two, and three locations, respectively, followed by a carat symbol ("∧") that appeared below each of the two locations that could succeed the last target location. Participants indicated at which of the two marked locations the target was most likely to have appeared next by pressing the corresponding key. In a sequence of target movements, target duration was 350 ms, and the interstimulus interval was 400 ms. The order of the two locations (1 and 6 in Table 4), four runs of two (3-2, 4-2, 3-5, and 4-5 in Table 4), and eight runs of three that constituted Trials 1–2, 3–6, and 7–14, respectively, was determined randomly for each participant.

On a prediction trial, the sequence of target locations could be repeated by pressing the space bar. When a prediction was made, the message *Question n,* where *n* indicated the number of the next trial, was displayed for 2,000 ms after which the next trial began. On completion of the prediction task, the percentage of the 14 trials receiving correct predictions was displayed. The prediction task began with three practice trials: one first-, one second-, and one third-order.

Procedure

Four participants were randomly assigned to each version of Table 4. At the start of Session 1, the SRT task was described, and participants were instructed to try to improve their RT with practice while keeping their error rate below 5%. After the last block of Session 3, the prediction task was described, and participants were told that their prediction accuracy would be displayed on completion of the task.

Results and Discussion

For each participant, the median RT for correct responses across 15 blocks was determined as a function of transition (EL1, EH1, CL1, CH1, EL2, EH2, CL2, CH2, EL3, EH3), type of run completed (see Table 1), and session (1, 2, 3). RTs are displayed in Figures 3 and 4. Results for runs ending with second- and third-order transitions were similar, and so the latter are not displayed in Figure 4.

ANOVAs with session (1, 2, 3), class (E, C), probability (L, H), first (E, U), middle (E, U), and last (E, U) as within-subjects factors were performed on the RT data for first-, second-, and third-order transitions (class was not a factor when analyzing third-order transitions). Order (first, second or second, third) was added as a within-subjects factor when comparing RTs on first- versus second- or second- versus third-order transitions, respectively.

Learning of First-, Second-, and Third-Order Probabilities

The Class × Probability × Session interaction was significant for first-order transitions, $F(2, 46) = 4.56$, $MSE = 1{,}199.16$, but not second-order transitions, $F(2, 46) < 1$, and the Probability × Session interaction was not significant for third-order transitions, $F(2, 46) = 1.02$, $MSE = 618.03$ (see Figure 3). Thus, RT differences between second-order and between third-order transitions did not change significantly across sessions, whereas differences between first-order transitions increased, mainly from Session 1 to Session 2. Excluding Session 1, the Class × Probability × Session interaction was no longer significant, $F(1, 23) < 1$.

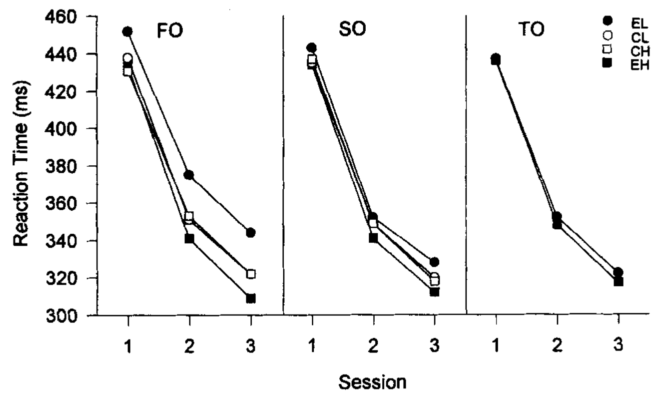


Figure 3. Reaction time, averaged across runs, as a function of order of transition, session, and transition. For third-order transitions, there were no CL and CH transitions in Experiment 3. FO = first-order transition; SO = second-order transition; TO = third-order transition; EL = experimental low transition; CL = control low transition; CH = control high transition; EH = experimental high transition.

Excluding Session 1, the Class × Probability interaction was significant for first-order transitions, $F(1, 23) = 66.74$, $MSE = 1{,}842.40$, and second-order transitions, $F(1, 23) = 13.68$, $MSE = 1{,}136.16$. RT was longer on EL1 than CL1 transitions, $F(1, 23) = 9.63$, $MSE = 10{,}978.99$, whereas the EH1-CH1 difference was not significant, $F(1, 23) = 2.70$, $MSE = 10{,}815.85$. RT was longer on EL2 than EH2 transitions, $F(1, 23) = 14.94$, $MSE = 2{,}314.06$, whereas the CL2-CH2 difference was not significant, $F(1, 23) < 1$. The EL2-CL2 ($p = .307$) and EH2-CH2 ($p = .218$) differences were not significant. RT was longer on EL3 than EH3 transitions, $F(1, 23) = 7.55$, $MSE = 607.92$, suggesting learning of the third-order probabilities.

For first- versus second-order transitions (excluding Session 1), the Class × Probability × Order interaction was significant, $F(1, 23) = 23.75$, $MSE = 1{,}075.26$. The EL2-CL2 difference was smaller than the EL1-CL1 difference, $F(1, 23) = 6.46$, $MSE = 4{,}777.25$, whereas the EH2-CH2 and EH1-CH1 differences were not significantly different, $F(1, 23) < 1$. Like the comparison between Experiments 1 and 2, these results strongly suggest learning of the first-order probabilities associated with the first-order transitions. As a note, the frequency of runs of three associated with each second-order probability was the same for first- and second-order transitions and hence was not a confounding factor. For example, the runs 2-1-3, where 1-3 was an EH1 transition, and 3-2-6, where 3-2-6 was an EH2 transition, each occurred 15 times across every three blocks.

For second- versus third-order transitions (excluding Session 1), the EL3-EH3 difference was smaller than the EL2-EH2 difference, $F(1, 23) = 4.48$, $MSE = 1{,}557.23$. This strongly suggests learning

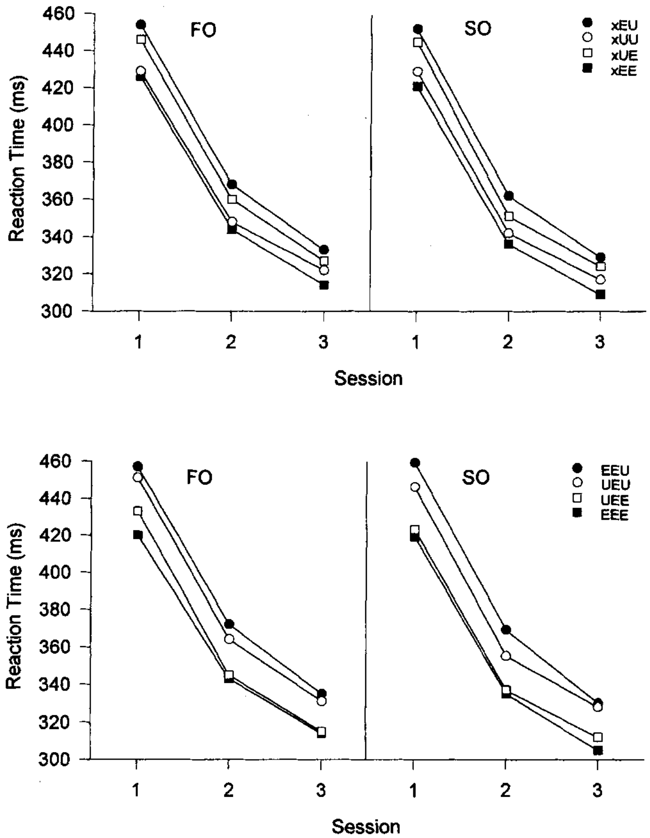


Figure 4. Reaction time (RT), averaged across transitions, as a function of order of transition ending a run, session, and run in Experiment 3. FO = first-order transition; SO = second-order transition; xEU = average RT on EEU and UEU runs; xUU = average RT on EUU and UUU runs; xUE = average RT on EUE and UUE runs; xEE = average RT on EEE and UEE runs.

of the second-order probabilities associated with the second-order transitions. The frequency of runs of four associated with each third-order probability was the same for second- and third-order transitions and so was not a confounding factor. For example, the runs 1-3-2-6, where 3-2-6 was an EH2 transition, and 2-1-3-2, where 2-1-3-2 was an EH3 transition, each occurred nine times across every three blocks.

Short-Term Priming Effects

For runs ending with first-order transitions, the Middle × Last interaction, $F(1, 23) = 120.05$, $MSE = 1{,}442.65$, and the Middle × Last × Session interaction, $F(2, 46) = 6.21$, $MSE = 750.02$, were significant, reflecting the longer average RTs on EEU and UEU runs than on EUU and UUU runs and shorter average RTs on EEE and UEE runs than EUE and UUE runs and their convergence with training (see Figure 4). The results were similar for runs ending with second- and third-order transitions. Thus short-term priming effects associated with bigrams were present and weakened with training.

For EEU, UEU, UEE, and EEE runs ending with first-order transitions, the First × Last interaction was significant, $F(1, 23) = 5.78$, $MSE = 1{,}473.58$, reflecting the longer RTs on EEU than UEU runs and shorter RTs on EEE than UEE runs (see Figure 4). The First × Last × Session interaction was not significant, $F(2, 46) = 2.06$, $MSE = 656.70$. The results were similar for runs ending with second- and third-order transitions. Thus, short-term priming effects associated with trigrams were present and did not diminish significantly with training.

Awareness of First-, Second-, and Third-Order Probabilities

For each participant, the percentages of Trials 1–2, 3–6, and 7–14 receiving a correct prediction were determined. Overall, these were 54.2%, 46.9%, and 50.0%, respectively, and did not differ significantly from chance (50.0%), three $Fs(1, 23) < 1$. Thus, there was no evidence of explicit knowledge of the first-, second-, and third-order probabilities associated with first-, second-, and third-order transitions, respectively.

Experiment 4

The preceding experiments suggested that much of the RT differences between transitions emerged by Session 2. However, differences might have increased with training beyond 3 sessions. Thus, Experiment 4 extended training to 15 sessions.

Method

Participants

The participants were 6 university undergraduates (2 women and 4 men) ranging in age from 18 to 32 years ($M = 24$ years).

The SRT Task

The SRT task was similar to that in Experiment 3 except for the following: Target locations were separated by intervals of 2.8 cm, and the response–stimulus interval was 250 ms. Each of the 15 sessions consisted of 18 blocks of 101 trials each with a 20-s break between blocks.[3] On a given day, there were 0, 1, or 2 sessions (with at least 90 min between sessions). There were never more than 2 consecutive zero-session days. The 15 sessions were completed in 12 to 15 days.

The Sequential Structure

The sequential structure was similar to that in Experiment 3 except there were no third-order transitions (see Table 5). Appendix C describes how sequences were generated. There were six versions of Table 5, as in Experiment 3.

The Awareness Questionnaire

The six-item questionnaire was a paper-and-pencil version of the prediction task in Experiment 3 with numbers representing target locations. Items 1–2 and 3–6 assessed awareness of the first- and second-order

[3] Actually, Sessions 9–14 consisted of 21 blocks, with 3 blocks being transfer blocks involving a reduction in the distance between target locations, a change in the mode of responding from six to two fingers, and no changes, respectively. First- and second-order probabilities in the transfer blocks were .50. The transfer blocks were designed to examine issues not relevant here and so will not be discussed further.

Table 5
First- and Second-Order Transitions in Experiment 4

Preceding target locations	Next target location					
	1	2	3	4	5	6
			First-order transitions			
3-2-1	—	—	EH1	EL1	—	—
4-2-1	—	—	EH1	EL1	—	—
3-5-1	—	—	EH1	EL1	—	—
4-5-1	—	—	EH1	EL1	—	—
3-2-6	—	—	EL1	EH1	—	—
4-2-6	—	—	EL1	EH1	—	—
3-5-6	—	—	EL1	EH1	—	—
4-5-6	—	—	EL1	EH1	—	—
			Second-order transitions			
1-3-2	EH2	—	—	—	—	EL2
6-3-2	EH2	—	—	—	—	EL2
1-4-2	EL2	—	—	—	—	EH2
6-4-2	EL2	—	—	—	—	EH2
1-3-5	EL2	—	—	—	—	EH2
6-3-5	EL2	—	—	—	—	EH2
1-4-5	EH2	—	—	—	—	EL2
6-4-5	EH2	—	—	—	—	EL2
			Control transitions			
2-1-3	—	CH1	—	—	CL1	—
		CH2			CL2	
5-1-3	—	CH1	—	—	CL1	—
		CH2			CL2	
2-6-3	—	CH1	—	—	CL1	—
		CL2			CH2	
5-6-3	—	CH1	—	—	CL1	—
		CL2			CH2	
2-1-4	—	CL1	—	—	CH1	—
		CL2			CH2	
5-1-4	—	CL1	—	—	CH1	—
		CL2			CH2	
2-6-4	—	CL1	—	—	CH1	—
		CH2			CL2	
5-6-4	—	CL1	—	—	CH1	—
		CH2			CL2	

Note. Dashes indicate that transitions did not occur. EL1 = experimental low first-order transition; EH1 = experimental high first-order transition; EL2 = experimental low second-order transition; EH2 = experimental high second-order transition; CL1 = control low first-order transition; CH1 = control high first-order transition; CL2 = control low second-order transition; CH2 = control high second-order transition.

probabilities associated with EL1-EH1 and EL2-EH2 transitions, respectively. For example, Item 3 would be the sequence 3-2 followed by the numbers 1 and 6 (see Table 5), with participants indicating at which of the two positions, 1 or 6, the target was most likely to have appeared next given its appearance at Positions 3 and then 2. Item 1 would be the sequence 1 followed by 3 and 4. For each item, participants were encouraged to press keys on the keyboard. The lines marking the target locations remained on the screen for reference.

Procedure

One participant was randomly assigned to each version of Table 5. After the last block of Session 15, the awareness questionnaire was administered.

Results and Discussion

For each participant, the median RT for correct responses across 18 blocks was determined as a function of transition (EL1, EH1, CL1, CH1, EL2, EH2, CL2, CH2), type of run completed (see Table 1), and session (1–15). RTs are displayed in Figures 5–7. ANOVAs were conducted as in Experiment 3.

Learning of First- and Second-Order Probabilities

The Class × Probability × Session interaction was marginally significant for second-order transitions, $F(14, 70) = 1.62$, $MSE = 16.10$, $p = .095$, but not for first-order transitions, $F(14, 70) < 1$ (see Figure 5). Thus RT differences between first-order transitions did not change significantly across sessions, whereas differences between second-order transitions increased marginally, mainly from Session 1 to Session 2. Excluding Session 1, the Class × Probability × Session interaction no longer approached significance, $F(13, 65) = 1.03$, $MSE = 12.12$.

Excluding Session 1, the Class × Probability interaction was significant for first-order transitions, $F(1, 5) = 13.64$, $MSE =$

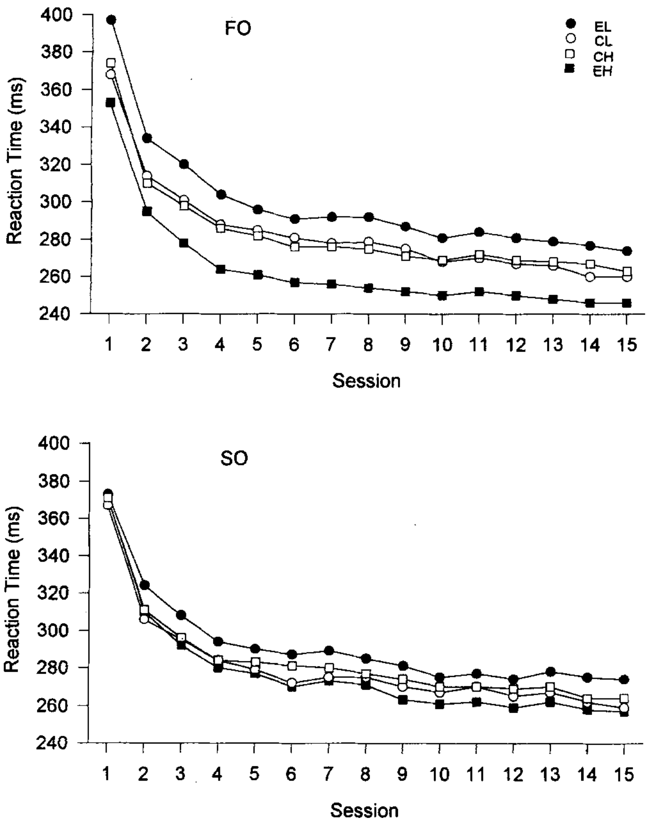


Figure 5. Reaction time, averaged across runs, as a function of order of transition, session, and transition in Experiment 4. FO = first-order transition; SO = second-order transition; EL = experimental low transition; CL = control low transition; CH = control high transition; EH = experimental high transition.

1,730.25, and second-order transitions, $F(1, 5) = 18.73$, $MSE = 393.43$. RT was longer on EL1 than CL1 transitions, $F(1, 5) = 8.13$, $MSE = 1{,}057.58$, and marginally shorter on EH1 than CH1 transitions, $F(1, 5) = 4.68$, $MSE = 3{,}310.09$, $p = .083$. RT was longer on EL2 than EH2 transitions, $F(1, 5) = 44.17$, $MSE = 225.98$, whereas the CL2-CH2 difference was not significant, $F(1, 5) < 1$. The EL2-CL2 ($p = .108$) and EH2-CH2 ($F < 1$) differences were not significant.

For first- versus second-order transitions (excluding Session 1), the Class × Probability × Order interaction was significant, $F(1, 5) = 7.03$, $MSE = 325.63$. The EH2-CH2 difference was smaller than the EH1-CH1 difference, $F(1, 5) = 6.63$, $MSE = 470.19$, whereas the EL2-CL2 and EL1-CL1 differences were not significantly different, $F(1, 5) < 1$. As in Experiment 3, these results strongly suggest learning of the first-order probabilities associated with the first-order transitions. Also, the frequency of runs of three associated with each second-order probability was the same for first- and second-order transitions and hence was not a confounding factor.

Short-Term Priming Effects

For runs ending with first-order transitions, the Middle × Last interaction, $F(1, 5) = 25.83$, $MSE = 80.60$, and Middle × Last × Session interaction, $F(14, 17) = 9.17$, $MSE = 9.22$, were significant (see Figure 6). The results were similar for runs ending with second-order transitions. As in Experiment 3, short-term priming effects associated with bigrams were present and weakened with training.

For EEU, UEU, UEE, and EEE runs ending with first-order transitions, the First × Last interaction was significant, $F(1, 5) = 27.03$, $MSE = 43.13$, whereas the First × Last × Session interaction was not, $F(14, 70) = 1.15$, $MSE = 22.42$ (see Figure 7). The results were similar for runs ending with second-order transitions. As in Experiment 3, short-term priming effects associated with trigrams were present and did not diminish significantly with training.

Awareness of First- and Second-Order Probabilities

For each participant, the percentages of Items 1–2 and 3–6 receiving a correct response were determined. Overall, these were 83.3% and 70.8%, respectively. The former was greater than chance (50.0%), $F(1, 5) = 10.00$, $MSE = 666.67$, whereas the latter was marginally greater, $F(1, 5) = 3.05$, $MSE = 854.17$, $p = .141$. Thus, there was evidence for explicit knowledge of the first-order probabilities associated with the first-order transitions

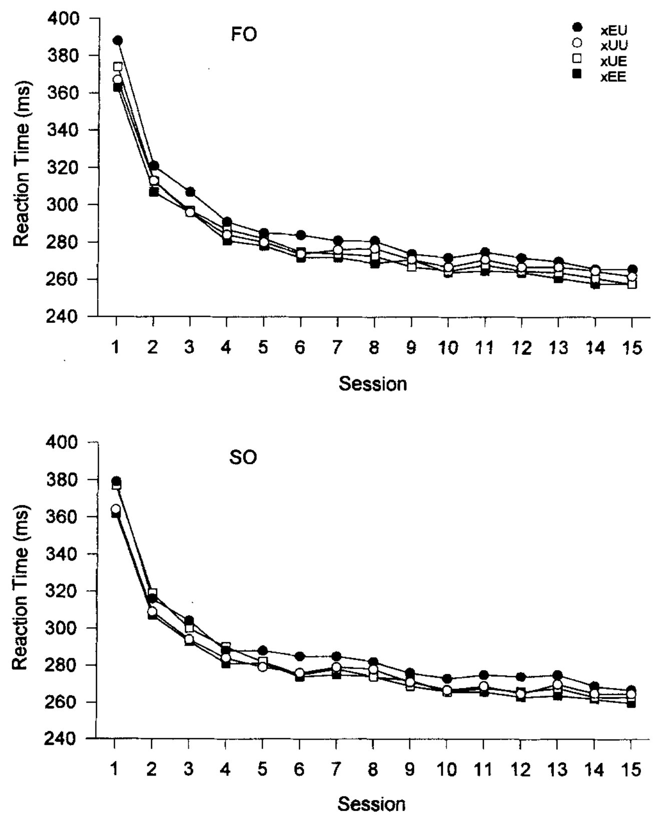


Figure 6. Reaction time (RT), averaged across transitions, as a function of order of transition ending a run, session, and run in Experiment 4. FO = first-order transition; SO = second-order transition; xEU = average RT on EEU and UEU runs; xUU = average RT on EUU and UUU runs; xUE = average RT on EUE and UUE runs; xEE = average RT on EEE and UEE runs.

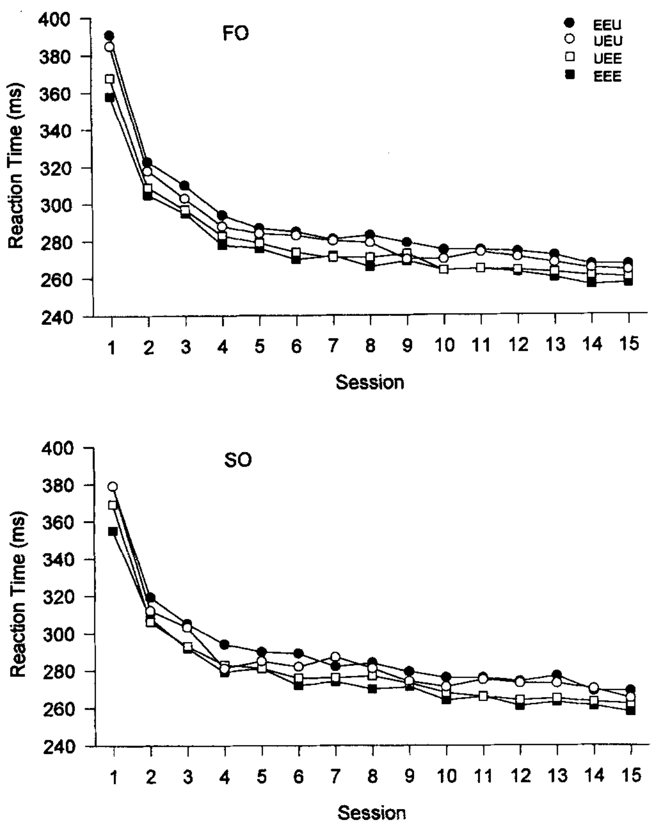


Figure 7. Reaction time, averaged across transitions, as a function of order of transition ending a run, session, and run in Experiment 4. FO = first-order transition, SO = second-order transition.

and marginal evidence for explicit knowledge of the second-order probabilities associated with the second-order transitions.

General Discussion

The current study provides convincing evidence for learning of first-, second-, and third-order probabilities. Using highly controlled sequences with constant Lag 2, Lag 3, and lower order probabilities, RT decreased with increasing first-, second-, and third-order probabilities and was not due entirely to learning of redundant higher order probabilities. In Experiments 1–3, there was no awareness of what was learned. On measures directly assessing explicit knowledge of the probabilities, high-probability transitions were chosen no more often than low-probability transitions. With extended training in Experiment 4, there was awareness of the first- and second-order probabilities. However, such awareness did not seem to affect performance. RT differences between transitions did not change significantly from Sessions 2 to 15 and were similar to those in Experiment 3 where there was no awareness of the probabilities. In addition to the evidence for implicit learning of first-, second-, and third-order probabilities, the current study also produced data patterns that might further characterize the implicit sequence-learning mechanism.

Impaired Use of Distant Information

For first-, second-, and third-order transitions, respectively, target locations on the previous one, two, and three trials were required to determine the most probable location on the current trial. The fact that RT differences between first-order transitions were larger than those between second-order transitions, which in turn were larger than those between third-order transitions (also see Jimenez et al., 1996) suggests that the learning mechanism's ability to use distant information or to conjoin information decreases as the distance or the amount of information increases, respectively.

Independence From Surface Features

The sequences of target locations in Experiments 1–2 and 3–4 consisted, respectively, or four and six events. Moreover, sequences in Experiments 1–2 included runs of the form 1-2-1-2, but those in Experiments 3–4 did not. In spite of the surface differences, learning of the transition probabilities in the two sets of experiments was similar. Session 3 learning, as indexed by (EL − CL) + (CH − EH), was 43 ms, 35 ms, and 39 ms in Experiment 1 and for first-order transitions in Experiments 3 and 4, respectively; and 13 ms, 14 ms, and 17 ms in Experiment 2 and for second-order transitions in Experiments 3 and 4, respectively. Thus, the learning mechanism appears not to be influenced by surface features, and what is important are the transition probabilities.

Linearity in Learning

RTs on CL and CH transitions fell roughly halfway between those on EL and EH transitions, suggesting a linear relationship between transition probability and RT. Averaging the learning scores in the preceding section and dividing by 2 yields, for first- and second-order probabilities, a decrease of 20 ms and 7 ms, respectively, for every probability increase of .10. The learning slopes are consistent with the results of other studies. Schvaneveldt and Gomez (1998) have examined learning of first- and second-order probabilities of .20 versus .80 and .10 versus .90, respectively, and have found RT differences between low- and high-probability transitions of approximately 130 ms and 55 ms. The differences are comparable to the 120 ms and 56 ms predicted by the learning slopes. Howard and Howard (1997, Experiment 1) have used sequences with second-order probabilities of .625 and .125 and have found an RT difference between low- and high-probability transitions of about 35 ms, which is identical to that predicted by the learning slope. Thus, the learning mechanism seems to follow a linearity principle described by the learning slopes (for possibly contradictory evidence, however, see Jimenez & Mendez, 1999).

Quick Learning

In Experiments 1–4, much of the RT differences among EL, EH, CL, and CH transitions emerged in Session 1 or Session 2. Many studies have observed considerable differences in RT between low- and high-probability transitions early in training (e.g., Baldwin & Kutas, 1997; Cleeremans, 1997; Cleeremans & McClelland, 1991; Curran, 1997; Howard & Howard, 1997; Jimenez et al., 1996; Meulemans, Van der Linden, & Perruchet, 1998; Perruchet et al., 1997, Experiment 5; Schvaneveldt & Gomez, 1998; Stadler, 1993). However, in studies involving more than one session, RT differences typically increase across sessions (Baldwin

& Kutas, 1997; Cleeremans, 1997; Cleeremans & McClelland, 1991; Howard & Howard, 1997; Jimenez & Mendez, 1999; Jimenez et al., 1996), although in some cases they do not (Baldwin & Kutas, 1997; Cleeremans, 1997). Aside from the fact that the sequences in the current study were much more controlled than sequences in previous studies, it is not clear what aspects of the sequences promoted quick learning of the transition probabilities.[4] It would be useful to try to empirically identify factors affecting the learning rate.

In conclusion, the current study provides strong evidence for implicit learning of first-, second-, and third-order probabilities and outlines a number of important characteristics that a mechanism of implicit sequence learning may possess (e.g., impaired use of distant information or linearity). Future research using well-controlled sequences such as those in the current study should further clarify these characteristics so as to better understand the learning mechanism(s).

[4] In Experiment 1 and for first- and second-order transitions in Experiments 4 and 3, respectively, RT differences between EL, EH, CL, and CH transitions did not increase significantly across sessions, nor across blocks when performance within Session 1 was examined. This suggested to one reviewer that the differences might be due to an artifact rather than transition-probability learning. However, the artifact would have to be present in all of the experiments in the current study because RT differences between first-order and between second-order transitions were similar across experiments. The artifact would also have to be present in a pilot study we conducted that was similar to Experiment 1, except there were only two sessions, no discrete blocks, and 12 rather than 4 versions of the sequential structure. In the pilot study, RT differences between transitions were similar to those in Experiment 1, and their increase across sessions approached significance. It is difficult to see how an artifact could manifest itself in the various experiments given that different experiments used sequences with different surface features (e.g., four vs. six events or 12 vs. 4 versions) and different orders of transition probabilities (i.e., first vs. second order). Also, RT differences between transitions exhibited significant or marginally significant increases across sessions in some cases but not others. In sum, RT differences between transitions were quite likely due to transition-probability learning, which can occur very quickly.

References

Baldwin, K. B., & Kutas, M. (1997). An ERP analysis of implicit structured sequence learning. *Psychophysiology, 34,* 74–86.

Bovens, N., & Brysbaert, M. (1990). IBM PC/XT/AT and PS/2 Turbo Pascal timing with extended resolution. *Behavior Research Methods, Instruments, & Computers, 22,* 332–334.

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

Cleeremans, A. (1997). Sequence learning in a dual-stimulus setting. *Psychological Research, 60,* 72–86.

Cleeremans, A., & Jimenez, L. (1998). Implicit sequence learning: The truth is in the detail. In M. A. Stadler & P. A. 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.

Curran, T. (1997). Effects of aging on implicit sequence learning: Accounting for sequence structure and explicit knowledge. *Psychological Research, 60,* 24–41.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 189–202.

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, and Cognition, 20,* 567–584.

Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology, 33,* 154–207.

Heuer, H., & Schmidtke, V. (1996). Secondary-task effects on sequence learning. *Psychological Research, 59,* 119–133.

Hoffman, J., & Koch, I. (1998). Implicit learning of loosely defined structures. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 161–200). Thousand Oaks, CA: Sage.

Howard, J. H., Jr., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging, 12,* 634–656.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45,* 188–196.

Jackson, G. M., Jackson, S. R., Harrison, J., Henderson, L., & Kennard, C. (1995). Serial reaction time learning and Parkinson's disease: Evidence for a procedural learning deficit. *Neuropsychologia, 33,* 577–593.

Jimenez, L., & Mendez, C. (1999). Which attention is needed for implicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 236–259.

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

Kornblum, S. (1975). An invariance in choice reaction time with varying numbers of alternatives and constant probability. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (pp. 366–382). New York: Academic Press.

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.

McDowall, J., Lustig, A., & Parkin, G. (1995). Indirect learning of event sequences: The effects of divided attention and stimulus continuity. *Canadian Journal of Experimental Psychology, 49,* 415–435.

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

Miller, J., & Ulrich, R. (1998). Locus of the effect of the number of alternative responses: Evidence from the lateralized readiness potential. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1215–1231.

Perruchet, P., Bigand, E., & Benoit-Gonin, F. (1997). The emergence of explicit knowledge during the early phase of learning in sequential reaction time tasks. *Psychological Research, 60,* 4–13.

Perruchet, P., Gallego, J., & Savy, I. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognitive Psychology, 22,* 493–516.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 585–594.

Remillard, G., & Clark, J. M. (1999). Generating fixed-length sequences satisfying any given nth-order transition probability matrix. *Behavior Research Methods, Instruments, & Computers, 31,* 235–243.

Schvaneveldt, R. W., & Gomez, R. L. (1998). Attention and probabilistic sequence learning. *Psychological Research, 61,* 175–190.

Shanks, D. R., Green, R. E. A., & Kolodny, J. (1994). A critical examination of the evidence for nonconscious (implicit) learning. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 837–860). Cambridge, MA: MIT Press.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17,* 367–447.

Soetens, E., Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 598–616.

Stadler, M. A. (1989). On learning complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1061–1069.

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

Stadler, M. A. (1993). Implicit serial learning: Questions inspired by Hebb (1961). *Memory & Cognition, 21,* 819–827.

Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 674–685.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1047–1060.

Appendix A

Generating Sequences for Experiments 1 and 2

Sequences for Tables 2 and 3 were generated using the matrices in Table A1. The elements in a row determined what followed the three-element run labeling the row. For example, row 1-2-4 in the left matrix consisted of six 2s and nine 3s, so that in a sequence, 2 and 3 followed 1-2-4 six and nine times, respectively. In the left matrix, some numbers appeared in pairs (e.g., 7/8). If the first number was used to generate one sequence, the second number was used to generate the next sequence, and vice versa.

To generate a sequence (i.e., two blocks of trials), the following seven-step algorithm was used:

Table A1
Generation Matrices for Tables 2 and 3

Row label	Table 2				Table 3			
	1	2	3	4	1	2	3	4
1–2–1	—	9	6	—	—	9	9	—
1–2–4	—	6	9	—	—	6	6	—
1–3–1	—	6	4	—	—	6	6	—
1–3–4	—	4	6	—	—	9	9	—
2–1–2	7/8	—	—	8/7	9	—	—	6
2–1–3	5	—	—	5	6	—	—	9
2–4–2	5	—	—	5	6	—	—	9
2–4–3	7/8	—	—	8/7	9	—	—	6
3–1–2	8/7	—	—	7/8	9	—	—	6
3–1–3	5	—	—	5	6	—	—	9
3–4–2	5	—	—	5	6	—	—	9
3–4–3	8/7	—	—	7/8	9	—	—	6
4–2–1	—	6	4	—	—	6	6	—
4–2–4	—	4	6	—	—	9	9	—
4–3–1	—	9	6	—	—	9	9	—
4–3–4	—	6	9	—	—	6	6	—

Note. Dashes indicate frequencies of 0.

1. Permute the elements in each of the 16 rows.

2. Choose a starting row at random.

3. Pick the first unchosen element in the current row.

4. Combine the last two elements of the current row label with the chosen element to establish the new current row. For example, if the current row is 4-2-4 and 3 was chosen from that row, the new current row is 2-4-3.

5. Repeat Steps 3 and 4 until all 200 and 240 elements in the left and right matrices, respectively, have been chosen. The last 3 elements chosen always correspond to the starting row.

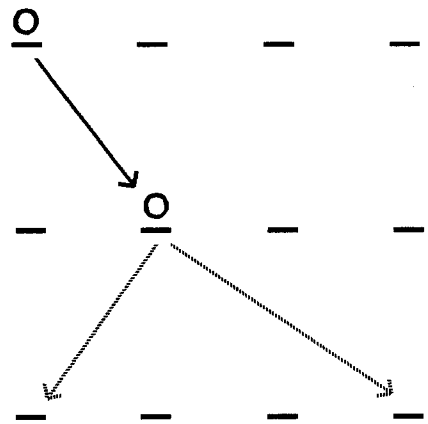
6. To maintain first-order probabilities in the 200-element sequence, copy the first element onto the end, resulting in a 201-element sequence. To maintain second-order probabilities in the 240-element sequence, copy the first 2 elements onto the end, resulting in a 242-element sequence.

7. Let Elements 1–101 form one sequence (i.e., one block of trials) and Elements 101–201 form another (i.e., the next block of trials). The last element of the first sequence and the first element of the second sequence overlap to maintain first-order probabilities. For the 242-element sequence, let Elements 1–122 form one sequence and Elements 121–242 form another. The last 2 elements of the first sequence and the first 2 elements of the second sequence overlap to maintain second-order probabilities.

The algorithm sometimes terminated without choosing every element in the matrix because an element had to be chosen from the starting row and there were none left to choose. If this happened, the row following the starting row was chosen as the new starting row, and the algorithm was restarted at Step 3. If none of the 16 rows were appropriate starting rows, the algorithm was restarted at Step 1.

Appendix B

An Item from the Awareness Questionnaire of Experiment 2



Suppose the o moved from position 1 to position 2. Now from position 2, the o went to position 1 ______% of the time and to position 4 ______% of the time.

35 40 45 50 55 60 65

(*Appendixes continue*)

Appendix C

Generating Sequences for Experiments 3 and 4

Sequences for Tables 4 and 5 were generated using the matrices in Table C1. The elements in a row determined what followed the three-element run labeling the row. For example, row 3-2-1 in the left matrix consisted of six 3s and four 4s, so that in a sequence, 3 and 4 followed 3-2-1 six and four times respectively. In the right matrix, some numbers appeared in pairs (e.g., 8/7). If the first number was used to generate one sequence, the second number was used to generate the next sequence and vice versa.

To generate a 303-element sequence (i.e., three blocks of trials), the matrix of interest in Table C1 was used as input to a sequence-generation algorithm (Remillard & Clark, 1999). Elements 1–101, 102–202, and 203–303 each constituted a block of trials. To generate a 99-element sequence for the practice block at the beginning of Session 1, all numbers in the two matrices in Table C1 were replaced with the number 2.

Table C1
Generation Matrices for Tables 4 and 5

Row label	Table 4						Table 5					
	1	2	3	4	5	6	1	2	3	4	5	6
						First-order transitions						
3–2–1	—	—	6	4	—	—	—	—	9	6	—	—
4–2–1	—	—	9	6	—	—	—	—	6	4	—	—
3–5–1	—	—	9	6	—	—	—	—	6	4	—	—
4–5–1	—	—	6	4	—	—	—	—	9	6	—	—
3–2–6	—	—	6	9	—	—	—	—	4	6	—	—
4–2–6	—	—	4	6	—	—	—	—	6	9	—	—
3–5–6	—	—	4	6	—	—	—	—	6	9	—	—
4–5–6	—	—	6	9	—	—	—	—	4	6	—	—
						Second-order transitions						
1–3–2	6	—	—	—	—	9	9	—	—	—	—	6
6–3–2	4	—	—	—	—	6	6	—	—	—	—	4
1–4–2	6	—	—	—	—	4	4	—	—	—	—	6
6–4–2	9	—	—	—	—	6	6	—	—	—	—	9
1–3–5	9	—	—	—	—	6	6	—	—	—	—	9
6–3–5	6	—	—	—	—	4	4	—	—	—	—	6
1–4–5	4	—	—	—	—	6	6	—	—	—	—	4
6–4–5	6	—	—	—	—	9	9	—	—	—	—	6
						Third-order and control transitions[a]						
2–1–3	—	9	—	—	6	—	—	8/7	—	—	7/8	—
5–1–3	—	6	—	—	9	—	—	7/8	—	—	8/7	—
2–6–3	—	6	—	—	4	—	—	5	—	—	5	—
5–6–3	—	4	—	—	6	—	—	5	—	—	5	—
2–1–4	—	4	—	—	6	—	—	5	—	—	5	—
5–1–4	—	6	—	—	4	—	—	5	—	—	5	—
2–6–4	—	6	—	—	9	—	—	7/8	—	—	8/7	—
5–6–4	—	9	—	—	6	—	—	8/7	—	—	7/8	—

Note. Dashes indicate frequencies of 0.
[a] There were no third-order transitions in Table 5.

Received April 12, 1999
Revision received April 22, 2000
Accepted April 22, 2000 ■