THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (9), 1757–1783

Psychology Press
Taylor & Francis Group

Applying an exemplar model to the serial reaction-time task: Anticipating from experience

Randall K. Jamieson
University of Manitoba, Winnipeg, Manitoba, Canada

D. J. K. Mewhort
Queen's University at Kingston, Kingston, Ontario, Canada

We present a serial reaction time (SRT) task in which participants identified the location of a target by pressing a key mapped to the location. The location of successive targets was determined by the rules of a grammar, and we varied the redundancy of the grammar. Increasing both practice and the redundancy of the grammar reduced response time, but the participants were unable to describe the grammar. Such results are usually discussed as examples of implicit learning. Instead, we treat performance in terms of retrieval from a multitrace memory. In our account, after each trial, participants store a trace comprising the current stimulus, the response associated with it, and the context provided by the immediately preceding response. When a target is presented, it is used as a prompt to retrieve the response mapped to it. As participants practise the task, the redundancy of the series helps point to the correct response and, thereby, speeds retrieval of the response. The model captured performance in the experiment and in classic SRT studies from the literature. Its success shows that the SRT task can be understood in terms of retrieval from memory without implying implicit learning.

Keywords: Artificial grammar; Serial response-time task; Global memory model; Implicit learning; Hick–Hyman Law.

Learning a set of words is difficult and effortful; it involves deliberate organization of the material (e.g., Tulving, 1962). Yet, people become sensitive to regularities in their environment without deliberate effort and without explicit awareness. Contrasting the two examples, several theorists have argued for two learning systems: an explicit system that handles deliberate learning, and an automatic system that abstracts contingencies and guides behaviour adaptively.

Reber (1967) was first to discuss the idea of automatic learning under the name *implicit learning*. In his now classic paper, people in a control condition studied strings of letters ordered at random; those in an experimental condition studied strings ordered by the rules of an artificial grammar. After they had studied the strings, the participants were told that the stimuli had been constructed using rules and were invited to sort new grammatical test strings from new

Correspondence should be addressed to Randall K. Jamieson, Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2. E-mail: randy_jamieson@umanitoba.ca

The research was supported by grants from the Natural Sciences and Engineering Research Council of Canada to both authors. We thank Joaquin Vaquero, Beth Johns, Bill Hockley, Scott Brown, Denis Cousineau, Matthew Kelly, and Ben Murdock for comments on a draft of the paper.

http://www.psypress.com/qjep
DOI:10.1080/17470210802557637

ungrammatical ones. Those who had studied grammatical exemplars could distinguish the two classes of stimuli (achieving a score of about 69% correct), but they could not characterize the rules that made strings grammatical. Those in the control condition could not classify test items better than chance. Reber argued that the ability to sort the stimuli must reflect knowledge of the rules, but, because the participants could not characterize the rules, the knowledge must be implicit. Following Reber, the idea that people abstract rules implicitly has been endorsed widely (e.g., Dienes, Broadbent, & Berry, 1991; Knowlton, Ramus, & Squire, 1992; Knowlton & Squire, 1993, 1994, 1996; Kuhn & Dienes, 2005; Manza & Reber, 1997; Mathews et al., 1989; McAndrews & Moscovitch, 1985; Reber, 1969, 1989, 1993; Rossnagel, 2001).[1]

The position that people abstract rules using an implicit-learning system is known as the two-systems view. It treats the judgement-of-grammaticality task as a laboratory model of how people acquire rules without explicit awareness or effort. In Reber's (1967) words, implicit abstraction is "a rudimentary inductive process that is intrinsic in such phenomena as language learning and pattern perception" (p. 863).

Despite its popularity, the two-systems view is controversial. Three kinds of argument have been marshalled against it. (a) The evidence that participants abstract a grammar is almost always confounded with other possibilities (see St. John & Shanks, 1997). Rather than using the grammar, for example, people may use information correlated with it, such as the similarity of the test exemplars to the studied exemplars (e.g., Jamieson & Mewhort, 2005, 2009). (b) The evidence that learning is implicit—the fact that the participants cannot articulate the rules—is almost always incomplete. People might be able to describe the rules if they are questioned more adroitly (e.g., Shanks & St. John, 1994; St. John & Shanks, 1997). (c) Current evidence favouring the two-systems view does not compel that position: The dissociation between performance and awareness can be explained without invoking two learning systems (Dunn & Kirsner, 1988; Hintzman, 1990; Jamieson & Mewhort, 2009; Vokey & Higham, 2005).

Other accounts have been advanced to replace the implicit-abstraction view. One alternative, called the *exemplar view*, proposes that people infer the grammatical status of test strings by comparing them to memory of the training strings (see Brooks, 1978; Brooks & Vokey, 1991; Jamieson & Mewhort, 2009; Vokey & Brooks, 1992, 1994; Whittlesea & Dorken, 1993; Whittlesea & Wright, 1997; Wright & Whittlesea, 1998).

To implement the exemplar position, Jamieson and Mewhort (2009) adapted Hintzman's exemplar model of memory, Minerva 2, to the judgement-of-grammaticality task (see Hintzman, 1984, 1986, 1988). In the task, grammatical exemplars are constructed by ordering strings of symbols according to the grammar's rules. To simulate performance, the model was provided with the same series of symbols as those shown to the participants. A unique vector of random values was used to represent each of the separate symbols within the model, and a grammatical string was constructed by concatenating the appropriate vectors to form a separate trace in memory. The grammaticality of a novel exemplar was assessed by its similarity to an aggregate of the stored traces.

Simulations with Minerva 2 matched performance in three new experiments and in three

[1] Although the judgement-of-grammaticality task has been used widely to investigate implicit learning, it lacks precision. First, it is unclear exactly what information a participant uses to make the choice. Second, a binary judgement truncates information into two categories: unlikely to do justice to subtle details in whatever has been learned. Third, decisions about test strings' grammaticality are collected after a study phase; it is not clear how loss of information between acquisition and test will affect those decisions. Finally, the grammars explored in many experiments conflate stimulus properties including positional dependencies (e.g., strings can begin with only one or two symbols), sequential dependencies (e.g., each letter can be followed by only one or two others), string length, the frequency of repeated letters, and so forth. Conflating so many factors makes a clear analysis difficult (see Johnstone & Shanks, 1999, 2001, for a full discussion).

experiments from the literature (Dienes et al., 1991; Reber, 1967; Vokey & Brooks, 1992). In all cases, increasing the number of rule violations in the probe increased the probability that the participants would reject it. Jamieson and Mewhort (2009) concluded that, although the evidence from standard artificial-grammar-learning tasks remains consistent with an account based on a separate implicit-learning system, it does not compel the two-process view. Instead of abstracting the rules of the grammar, the simulations confirm that participants could judge grammaticality on the basis of similarity. The demonstration shifts the burden of proof back to proponents of the two-process view.

There is evidence for the two-process view, however, that finesses the exemplar position and Jamieson and Mewhort's (2009) simulations, in particular. Nissen and Bullemer's (1987) influential serial response time (SRT) experiment illustrates the situation. They required participants to identify the location of a target (an asterisk) on a computer screen; each location at which the target could appear was mapped to a unique response key. When the target appeared, the participant noted its location by pressing the response key mapped to that position. In the experimental condition, the target occurred in a repeating sequence; in a control condition, the target appeared in a random sequence. Response time decreased at a faster rate across trials in the experimental condition than in the control condition. The faster responding for the repeating sequence indicates that people could exploit the repetition-induced structure in the series. Nevertheless, the participants were unable to describe how the structure was built into the series. The combination of results illustrates a dissociation between performance and awareness similar to the one that Reber (1967) obtained using the judgement-of-grammaticality task. In the SRT task, however, learning was assessed using changes in response time; the assessment did not involve test exemplars. Hence, the exemplar view, and Jamieson and Mewhort's simulation model in particular, does not apply.

Difficulty in applying the categorization model to the SRT task, however, does not falsify Jamieson and Mewhort's (2009) main claim: Performance reflects established principles of retrieval from memory rather than an implicit-learning mechanism. To apply the argument to the SRT task, however, their account of retrieval would have to be adapted to suit the difference in the way learning is measured; that is, the model must be modified to predict reaction time. Here, we adapt Minerva 2 to estimate reaction time and ask whether performance in SRT tasks can be understood in terms of retrieval from memory. A demonstration that the same principles apply both to the judgement of grammaticality and to the SRT task would unify our understanding of the two paradigms and, again, call the two-systems view into question. Before we develop our account of retrieval in the SRT task, however, we present an experiment to serve as a target for the simulations.

In Jamieson and Mewhort's (2009) third experiment, participants studied strings of digits arranged according to one of three grammars. The grammars varied the redundancy of the strings. After participants had studied the training strings, they were told that the materials were constructed using rules and were required to judge the grammaticality of novel test strings in a two-alternative forced-choice (2-AFC) task. Participants' ability to select the grammatical item was an increasing linear function of the redundancy of the strings used in the training materials. Experiment 1 examines whether or not the redundancy of a grammar—the same manipulation as that used by Jamieson and Mewhort—yields similar results in the SRT task.

EXPERIMENT 1

On each trial in the experiment, a target (i.e., a white disc) appeared in one of six locations on a black computer screen. Each location was mapped to a key on the computer's keyboard. The participant identified the location of the target, as quickly as possible, by pressing the corresponding key. The location of successive targets was determined by the rules of a grammar.

We used three grammars to introduce three degrees of predictability, or redundancy, into the series of targets. If the effect of redundancy in the grammar on performance in the SRT task parallels its effect on performance in the judgement-of-grammaticality task, the time to identify the location of the target should decrease across trials, with the rate of the decrease an increasing function of redundancy.

Method

Participants

A total of 36 students from the Queen's University undergraduate participant pool took part in the study. Each participant was assigned randomly to one of three conditions defined by an artificial grammar. All participants reported normal or corrected-to-normal vision.

Apparatus

The experiment was administered on a personal computer equipped with a 17-in. monitor and QWERTY keyboard. The experiment was administered using software written in Turbo Pascal 7.0. To ensure accurate measurement of response time, we used a modified version of Heathcote's (1988) timing and screen-control routines and cleared the keyboard's memory buffer before each trial. To apply Heathcote's routines, we conducted the experiment under MS-DOS 6.1.

Materials

The experiment included 837 trials. The first 37 trials were practice trials designed so that the target moved once from each of the six locations to each of the other six locations (including itself). For the remaining 800 trials, the location of successive targets was defined by the transition probabilities described in the three grammars shown in Table 1.

Each grammar in Table 1 shows the probabilities with which the target moved from one location to another on successive trials. In Table 1, the six locations are numbered 1 through 6. In the leftmost grammar in the table, if the target appeared in Positions 1, 2, or 3 on trial *n*, it appeared in Locations 2, 3, and 4, respectively, on the succeeding trial. If the target appeared in Positions 4, 5, or 6, on trial *n*, it appeared at one of two potential locations on the succeeding trial (with equal probability). In the middle grammar in the table, the target could move from any of the locations on trial *n* to one of two allowable locations on the succeeding trial (with equal probability). In the rightmost grammar in the table, the target could move from any location on trial *n* to one of three others on the succeeding trial (with equal probability). In all of the conditions, a target could not reappear at the same location on consecutive trials.

We created sequences of 800 experimental trials by picking the first position at random and

Table 1. *Three transition grammars used in Experiment 1*

	Redundancy																	
	.81						*.61*						*.39*					
	Successor						*Successor*						*Successor*					
Position	*1*	*2*	*3*	*4*	*5*	*6*	*1*	*2*	*3*	*4*	*5*	*6*	*1*	*2*	*3*	*4*	*5*	*6*
1		1						1/2	1/2					1/3	1/3	1/3		
2			1						1/2	1/2					1/3	1/3	1/3	
3				1						1/2	1/2					1/3	1/3	1/3
4					1/2	1/2					1/2	1/2	1/3				1/3	1/3
5	1/2					1/2	1/2					1/2	1/3	1/3				1/3
6	1/2	1/2					1/2	1/2					1/3	1/3	1/3			

then selecting successive digits according to the probabilities in the relevant grammar.

We quantified the amount of structure in each grammar using the redundancy statistic G (e.g., Jamieson & Mewhort, 2005, 2009). G is computed using Shannon and Weaver's (1949) expression for first-order sequential uncertainty:[2]

$$U = -\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}\log_2 p_{ij},$$

where p_{ij} denotes the probability of symbol j following symbol i in a sequence, and n is the number of symbols in the grammar. The grammatical redundancy of a target grammar was computed by comparing its first-order sequential uncertainty against the first-order sequential uncertainty of an equivalent, but unconstrained, grammar:

$$G = 1 - \frac{U(Grammar)}{U(Unconstrained)}$$

Using a six-by-six transition matrix, an unconstrained grammar would have all transition probabilities equal (i.e., each entry in the table would be 1/6).

G measures bigram structure associated with a grammar of interest. For the three grammars reading from left to right, $G = .81$, .61, and .39, respectively.

Procedure

Participants were seated in front of a computer with a monitor and a keyboard attached to it. The phrase "Press any key to begin ..." was printed in white at the centre of the monitor.

On each trial, a white disc appeared at one of six locations on the computer's screen. As illustrated in Figure 1, six squares, each marked by a white border, identified the six locations at which the target could appear. The task was to identify the location of the target by pushing one of six response keys. The six keys *s, d, f, j, k*, and *l* on the computer's keyboard were mapped to the six successive positions on the screen (moving from left to right, respectively). Each participant used the middle fingers (i.e., the ring, middle, and index) of the left hand for responses *s*, *d*, and *f* and used the corresponding fingers of the right hand for responses *j*, *k*, and *l*.

Before starting the eight blocks of experimental trials, the participant pressed a key on the keyboard to initiate a block of practice trials. On the participants' keystroke, the screen was cleared, and, after a 250-ms pause, the six square borders were displayed on the computer screen. The borders were arrayed horizontally and were centred both horizontally and vertically on the screen (see Figure 1). After a pause of 500 ms, the target appeared inside one of the borders. The target remained on the screen until the participant pressed one of the six response keys. When the participant responded, the target was cleared but the six borders remained. The target reappeared 500 ms later at another location. If the participant's response was wrong, a 150-Hz tone was sounded for the first 100 ms of the 500-ms intertrial interval.

[2] For an example with a number of equally likely stimuli, uncertainty, U, is computed as,

$$U = \log_2(N) = -\log_2(p)$$

where N is the number of stimulus alternatives, and $p = 1/N$ is the probability that each alternative will occur on a given trial. U is expressed in bits. To calculate mean uncertainty when the probabilities of events are unequal, the uncertainty of each event is weighted by its probability, and the product is summed across all events—that is,

$$U = \sum_{i}^{N} -p_i \log_2 p_i,$$

where N refers to the number of events, and p_i is the probability of event i (see Shannon & Weaver, 1949).

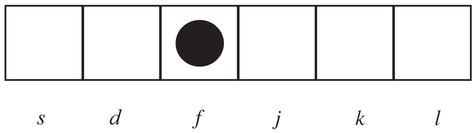

Figure 1. *A diagram of the display presented to participants. The target is shown on the screen in the third position from the left. The response keys* [s], [d], [f], [j], [k], [l]*) were mapped to the corresponding locations, but did not appear on the screen in the experiment.*

Testing continued until the participant had completed the 37 practice trials. Following the practice, the screen was cleared, and after a 1,000-ms pause, "Press any key to begin . . ." was displayed at the centre of the computer's screen.

The participants were informed that the experiment would be administered in eight blocks of 100 trials and that they would have an opportunity to rest between blocks. The instructions asked the participants to respond as quickly as possible without sacrificing accuracy. The procedure for the eight blocks of experimental trials was identical to that in the practice session, except, of course, that the sequence of target locations was determined by the probabilities in the relevant grammar.

After the experiment, each participant was asked if he or she had noticed anything about the sequence. The experimenter prompted the participant again by noting that the sequence had been constructed according to rules.

Results

The top panel in Figure 2 shows mean response time (RT) as a function of practice and of grammatical redundancy. As is shown in the figure, RT decreased linearly across practice in all three groups with the pattern of results yielding a reliable practice (linear) by redundancy (linear) interaction, $F(1, 33) = 14.94$, $\eta_p^2 = .15$, $p < .05$.[3] Mean accuracy across participants was uniform at 96% ($SD = 2.04\%$) and was independent of both practice, $F(7, 231) = 1.76$, $\eta_p^2 = .05$, $p > .05$, and redundancy, $F(2, 33) = 2.21$, $\eta_p^2 = .12$, $p > .05$ (both tests were assessed after applying the Greenhouse–Geisser correction).

We compared performance for low- and high-probability transitions in the $G = .81$ condition where some transitions were determined (i.e., 0-bit transitions in the first three rows in Table 1) and others were not (i.e., 1-bit transitions in the second three rows in Table 1). We computed each participant's mean RT for the two types of transitions. Mean RT for the 0-bit transitions ($M = 357$ ms, $SE = 21$ ms) was faster than that for the 1-bit transitions ($M = 390$ ms, $SE = 13$ ms), $t(11) = 3.09$, $\eta^2 = .46$, $p < .05$. The greater the uncertainty, the slower participants were to identify the target's location.

As is clear in the top panel in Figure 2, the participants learned to use the sequential redundancy of successive positions, and the rate of their learning was an increasing linear function of the grammar's redundancy. When questioned, however, only 2 participants in the most highly constrained condition ($G = .81$) reported that there were contingencies to be learned. Neither of the 2 participants was explicit about the probability structure. Hence, the present results confirm a pattern of results associated with the implicit-learning view: The participants capitalized on redundancy in successive stimuli but they could not describe the redundancy, and, indeed, most were effectively unaware of it. The SRT procedure appears to provide a strong case for automatic learning of the structure across trials and, therefore, challenges the claim that performance in implicit-learning tasks can be understood in terms of retrieval.

UNDERSTANDING PERFORMANCE IN THE SRT TASK

According to the implicit-learning view, people abstract whatever rules underlie order in the

[3] Orthogonal trend coefficients for unequal intervals were computed using an algorithm suggested by Gaito (1965). Software to compute the coefficients is available from the authors for Linux, Solaris, Macintosh-Intel, or 32-bit Windows machines.

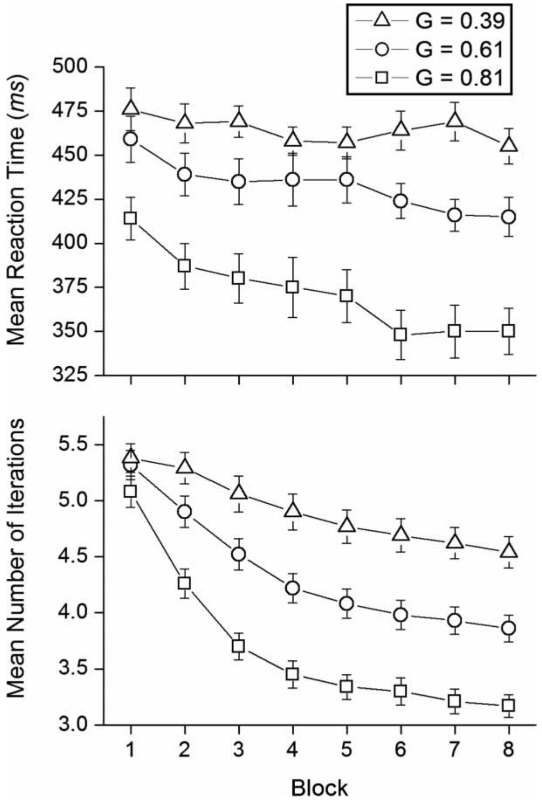

Figure 2. *The top panel shows reaction time as a function of practice and sequential redundancy. Practice is expressed in blocks of 100 trials. The bottom panel shows simulation results: iterations as a function of practice and sequential redundancy. Error bars represent standard errors. For the simulated results, error bars are computed for* n = *25.*

chain of stimuli. Moreover, it is thought that a separate implicit-learning system does so automatically. Our position differs both on what is learned and on how it is learned.

At the outset of the experiment, the participants are told of the mapping between target position and the response. To respond correctly when the target appears, they must remember that mapping and press the appropriate button; that is, they must remember the response associated with the current stimulus. Early in practice, they have no alternative but to depend on the stimulus-to-response mapping alone to guide their response. As they develop experience in the task, however, a new and reliable source of evidence about the response becomes available—namely, the participant's own memory of his or her performance in the task.

What kind of information do participants store about their own performance? The implicit-learning position claims that they learn rules that describe the probability structure of the sequence but cannot report the rules because the knowledge is implicit. Stadler (1992, 1993, 1995), however, has noted that learning in the SRT task is specific to short runs of trials in the same way that priming in word identification is specific to previously studied stimuli (Jacoby, 1983), and learning in an alphabet-arithmetic task is specific to the studied problems (e.g., Logan, 1988). Our position is an implementation of the same idea: We suggest that people record an incomplete history of local information about the task. The history includes the participant's response to each stimulus; it also includes the context of that response—namely, the response made on the previous trial—a position consistent with Willingham's (1999) demonstration that performance in the SRT task reflects learning of trial-to-trial response contingencies. Such a history—even if it were very incomplete—provides reliable information about the current response because the sequence of stimuli is predictable at the first order (see Stadler, 1992, p. 319). Yet, because the history records a stream of local events, it should be no surprise that it provides little global information about the rules needed to generate the stimuli. Rather than postulating an implicit system that abstracts the rules underlying a chain of responses, then, we propose that people use local information about their own responses from their history of responding. To show that our position is competent to accommodate data from the SRT task, we implement the theory by adapting an established model of human memory to it: Minerva 2, the same model as that used by Jamieson and Mewhort (2009) to understand performance in the judgement-of-grammaticality task.

Minerva 2 (Hintzman, 1984, 1986, 1988, 1990) was developed to understand the explicit-recognition and judgement-of-frequency tasks (Hintzman, 1984, 1988). It has since been applied to a wide range of phenomena including categorization (Hintzman, 1986, 1988), confidence–accuracy inversions in recognition memory (Clark, 1997), recognition failure of recallable words (Hintzman, 1987), false recognition in the Deese–Roediger–McDermott paradigm (Arndt & Hirshman, 1998), likelihood judgement (Dougherty, Gettys, & Ogden, 1999), extrapolation in function learning (Kwantes & Neal, 2006), speech perception (Goldinger, 1998), word naming (Kwantes & Mewhort, 1999), and access to semantic memory (Kwantes, 2005). Finally, as noted earlier, Jamieson and Mewhort (2009) showed that it handles judgement-of-grammaticality data. In the next section, we adapt Minerva 2 to the SRT task.

Minerva 2

Minerva 2 is a multitrace theory of memory. When a participant studies an item, or a pair of items, the event is encoded to memory as a unique trace. In the model, a unique vector of n elements is used to represent each item. Each vector element takes one of two values: $+1$ or -1 with equal probability—that is, $p(+1) = p(-1) = .5$. An association between two items (or between a cue and a response, or an exemplar and a category label) is represented by concatenating the constituent item vectors to form a new vector of larger dimensionality.

Memory is a matrix with one row (vector) for each studied event or association. Encoding an event involves copying its vector to a new row in the memory matrix. Encoding is sometimes imperfect. The model accommodates imperfect encoding by setting some vector elements to zero (indicating that the element is indeterminate or unknown). A parameter L controls the probability with which an element is stored correctly. As L increases, the quality of the encoded stimulus improves. Minerva 2 treats forgetting as the inverse of correct encoding; hence, reducing L can be used to accommodate memory loss.

All retrieval is cued. A cue may refer to a complete trace, as in item recognition, or to part of a trace, as in cued recall. We consider Minerva 2's retrieval mechanisms for the complete-trace example first because those mechanisms are generalized easily to the cued-recall example.

When a retrieval cue is presented, it activates each trace in memory in proportion to its similarity to the cue. The activation from all traces is aggregated into a composite trace (called the *echo*). The similarity of trace, i, to the probe, P, is computed using a vector cosine. Because the stimulus vectors are composed of $+1/-1$ elements, the vectors are normalized to 1; hence the cosine can be computed:

$$S_i = \frac{\sum_{j=1}^{n} P_j \times M_{ij}}{n},$$

where P_j is the value of the jth feature in the probe, M_{ij} is the value of jth feature of the ith row in memory, and n is the number of features in the comparison. Like the Pearson r, the similarity of the ith item to the probe, S_i, is scaled to the interval $\{-1, +1\}$. Similarity equals $+1$ when the trace is identical to the probe, 0 when the trace is orthogonal to the probe, and -1 when the trace is opposite to the probe.

The ith trace's activation, A_i, is the cube of its similarity to the probe—that is,

$$A_i = S_i^3. \qquad (1)$$

The activation function exaggerates the differences in similarity between the probe and the items in memory by attenuating the contribution of exemplars to the echo that are dissimilar to more so than those that are similar to the probe. Note that using an odd-numbered exponent in the activation function preserves the sign of the argument, S_i.

The echo, C, is a vector obtained by weighting each of the $i = 1 \ldots m$ traces in memory by its activation and summing all m traces into a composite,

$$C_j = \sum_{i=1}^{m} A_i \times M_{ij}.$$

The echo represents the composite of the information in memory activated by the probe; it is used to model cued-recall and categorization experiments.

The echo's intensity is computed by summing activation across the $i = 1 \ldots m$ traces in memory,

$$I = \sum_{i=1}^{m} A_i. \qquad 2$$

The echo's intensity, I, ranges between $-m$ and m, and it is a function of both the number of traces in memory and the degree to which the traces match the probe: If all traces in memory were identical to the probe, $I = m$. The echo's intensity measures the activation triggered by the probe, and it is often treated as an index of the probe's familiarity with respect to the studied material. Because the intensity measure is summative, however, its range of scores varies with the number of items stored in memory and so is inappropriate for simulations in which the number of traces varies—for example, an experiment that involves learning with practice (e.g., the SRT task). An alternative measure that solves this problem is obtained by normalizing the echo and computing the similarity between it and the probe,

$$I = \frac{\sum_{j=1}^{n} P_j \times C_j}{n}, \qquad 3$$

where P_j is the value of the jth feature in the probe, C_j is the value of jth feature in the echo, and n is the number of features in the vectors under comparison. Using this measure of echo intensity, $I = 1$ when the probe and the echo are identical, $I = 0$ when the probe and echo are orthogonal, and $I = -1$ when the probe and the echo are opposite. We use the measure of echo intensity in Equation 3 throughout this paper.

To extend the model to the cued-recall situation, consider an example in which each trace is composed of associated items so that features $j = 1 \ldots k$ represent features of the probe, and features $j = (k + 1) \ldots n$ represent features of the response. To retrieve a response to the probe, the probe is submitted to memory with features $1 \ldots k$ filled in and features $(k + 1) \ldots n$ empty. Because activation is applied to the entire trace, however, the response information is activated and, if the probe is paired sufficiently often with a particular response, retrieves a clear representation of that response into the echo. This is how Minerva 2 accomplishes cued recall. We model response selection in the SRT task using the operations that are used in Minerva 2 to model cued recall.

Adapting Minerva to the SRT task

In Experiment 1, a white disc was shown in one of six positions, and the participant was asked to identify the disc's position by pressing one of six response keys. As participants practised the task, their response time decreased, indicating that they were able to speed the correct response.

To simulate the SRT task, we started by constructing 12 vectors, 1 to stand for each of the six stimulus positions and for each of the six possible responses. Each vector was of dimensionality 20 with values of +1 or −1 selected at random with $p(+1) = p(-1) = .5$.

When participants made a response, we assume that they updated memory with a record of the current stimulus, the response required by that stimulus, and the context provided by their own response on the previous trial. Hence, each trial was represented by a vector of dimensionality 60 constructed by concatenating the vectors for the current stimulus (S_i), the response associated with that stimulus (R_i), and the response on the previous trial ($R_{(i-1)}$)—that is, $S_i//R_{(i-1)}//R_i$, where // indicates concatenation, and i identifies the ith trial in the series.[4]

[4] In our simulations, we have limited the context information to the immediately preceding response. The decision to limit the context was one of convenience; we are open to the possibility that a larger context might be required if second-, third-, or higher order predictability were introduced into a sequence of stimuli (cf. Reed & Johnson, 1994; Remillard, 2008).

To acknowledge the instructions given to the participants before starting the experiment proper, we stocked memory with one example of each $S_{(i)}//R_{(i-1)}//R_{(i)}$ contingency with $L = 1$. Then, prior to the simulation we stored trials corresponding to four practice blocks. Thereafter, we presented the model with a sequence of stimuli generated as it was in the target experiment. As the experiment proceeded, we added a new row to the memory matrix to record the events from each trial.

Although each stimulus was visually distinct and clear, a task of 800 trials is, we were told, quite boring, and so it is unlikely that participants encoded each event in complete detail. Accordingly, we set the learning parameter to a moderate value, $L = .7$. As a result, about 30% of the elements in each string were indeterminate, representing incomplete encoding of events.

When a stimulus is presented, we assume that the participant attempts to retrieve the correct response given a probe composed of the current stimulus and the participant's own response from the previous trial. To do so, we applied Minerva 2's mechanism for cued recall; that is, we used S_i // $R_{(i-1)}$ as a prompt to recover R_i .

Minerva 2 does not have a mechanism with which to simulate response time.[5] To estimate response time, we borrowed an idea from the iterative resonance model (IRM) and applied it to Minerva's retrieval mechanism for cued recall (see Mewhort & Johns, 2005).

Both Minerva 2 and the IRM treat retrieval in terms of a resonance metaphor (see Ratcliff, 1978; Semon, 1909/1923). When a tuning fork is sounded near an undamped piano, energy is transferred to the piano's strings. The transfer is proportional to the match in frequency between the tuning fork and the strings of the piano. In Minerva 2, activation, described by Equation 1, describes resonance in memory produced by a probe, and the echo is a summary of the activation.

As the energy in the strings dissipates, the mix of sounds changes. In particular, the tone closest to the frequency of the tuning fork increasingly dominates the mix of tones. The IRM extends the resonance metaphor across time. Initially, when the retrieval probe is applied to memory, all items are activated, but as retrieval proceeds, the contribution of the item(s) most similar to the probe increasingly dominates the echo. To implement the changes in the contribution of the studied items to the echo across time, the IRM computes a fresh echo at each time step. On the first time step, the exponent in the activation function (see Equation 1) was set to 1; on successive iterations, the exponent in the activation function was increased by 1. Increasing the exponent exaggerated differences in similarity of the probe to each trace. For that reason, retrieval exaggerated contributions from traces most similar to the probe. In terms of the IRM, Minerva's activation function (Equation 1) matches the IRM's activation function at the third time step.

One can think of the changes in the echo in terms of a ratio of signal to noise. The signal is the contribution from the traces from the item(s) most similar to the probe; contributions from other traces constitute noise. Initially, the echo contains information from many traces. The dominance of one trace over the rest is determined by the similarity of the stored traces to the cue. In an extreme example, if the cue were orthogonal to all traces but one, only the single nonorthogonal trace would contribute to the echo. In this case, the signal-to-noise ratio for information in the echo would be 1.0. If the cue is similar to several different traces, however, they will all contribute to the echo (in proportion to their similarity to the probe), and the signal-to-noise ratio for information in the echo will be less than 1.0. Increasing the exponent in the activation function exaggerates differences in the similarity of the probe to the traces in memory and, thereby, increases the signal-to-noise ratio.

[5] Minerva 2 has a mechanism, called deblurring, in which the echo retrieved is fed iteratively back into the system until the response subtrace is clear enough to support a response. We tried to model performance with this mechanism but failed. Indeed, as we increased the number of traces studied, when deblurring, the response subtrace became increasingly noisy rather than becoming increasingly clear. Dienes (1992) reported the same problem.

In the simulations that follow, we used the IRM's retrieval mechanism to home in on the item(s) most similar to the probe.[6] On the first iteration, the exponent in the activation function (see Equation 1) was set to 1; for each time step, the exponent in the activation function was increased by 1. In all cases, the sign of a similarity score was preserved in the output of the activation function (i.e., a negative similarity score produced a negative activation score, and a positive similarity score produced a positive activation score). Increases to the exponent across iterations increased the relative contribution of traces that most closely matched the $S_i//R_{(i-1)}$ values of the probe; that is, increases to the exponent increased the signal-to-noise ratio in the echo by forcing retrieval to home in on those traces that best matched the corresponding elements of the probe.

As the echo homed in on the traces that most closely matched the $S_i//R_{(i-1)}$ values of the probe, it also homed in on the response information associated with the probe. When the response portion of the echo was similar enough to one of the possible responses, it was selected for report. We set a similarity criterion of $k = .99$.

The model's retrieval time is based on the number of iterations needed to select a response; we used the number of iterations to predict reaction time. To obtain stable estimates of performance, we averaged across 25 independent simulations of the experiment.

The results of the simulation for Experiment 1 are shown in the bottom panel of Figure 2. Simulated RT (expressed as mean number of iterations) decreased with practice. The rate of learning increased the greater the sequential constraint in the grammar that was used to construct the test sequence. Accuracy was at ceiling throughout.

As is clear from the two panels of Figure 2, the simulation captured the main features from the empirical results, $r^2 = .89$. In addition, when we examined performance on 0-bit versus 1-bit transitions within the series of trials, we found that the 0-bit transitions were faster than the 1-bit transitions. The model anticipated the same pattern: The mean number of iterations to retrieve the correct response was smaller for 0-bit ($M = 2.83$) than for 1-bit transitions ($M = 4.11$). Thus, looking at the data overall (as in Figure 2) or at a transition-to-transition perspective, the model correctly anticipated performance.[7]

We conclude that performance in the target SRT task can be understood in terms of extra information about the response that becomes available from the participant's noisy history of responding to particular stimuli. The history includes the local context of each response (i.e., the immediately preceding response), and it provides a reliable source of information to the extent that the sequence of successive stimuli is predictable. In short, learning in our SRT task does not require an implicit system to abstract the rules. Instead, learning can be understood in terms of the participant's ability to retrieve useful information from his, or her, history of responding.

Despite clear evidence of learning in the simulation, our account explains why participants were

[6] The IRM, described by Mewhort and Johns (2005), set the exponent on the initial time step to 0 and increased it by 1 on each iteration; this differs from our practice of setting the initial time step to 1 and increasing it by 1 on each iteration. Their vectors were composed of zeros and ones; ours were composed of −1s and +1s.

[7] The model has only two free parameters—namely, L, the encoding quality parameter, and k, the decision criterion. The encoding parameter, L, controls the quality of the representation in memory. If it is set to 1, learning is speeded: In Experiment 1, for example, learning would be all but complete within the first block of trials. If it is set to a low value, the rate of learning is strongly attenuated. Its influence is greatest for the low-redundancy sequences. The decision criterion, k, controls the quality of information retrieved needed to issue a decision. Decreasing the decision criterion speeds response time but increases the probability of an error response. If the decision criterion is set to a very low value, accuracy suffers accordingly. In addition to the free parameters, the model has a structural parameter—namely, the dimensionality of the vectors that represent each symbol. Decreasing the dimensionality both reduces the rate of learning and increases the number of errors. The dimensionality parameter trades off with L, and all three parameters interact. We selected parameter values so that the same values work in all the simulations. We could have modestly better fits to each task, but we have little justification for changing the parameters across tasks.

unable to articulate the series when prompted to do so. The simulation is based on the idea that people retrieve information about the current response given a probe—that is, they seek R_i given $S_i//R_{(i-1)}$ as a prompt. Suppose, after 800 trials, that an experimenter were to invite the participant to provide R_i with only, $R_{(i-1)}$ as a cue—that is, with only half the usual cue. Performance with half the cue would be better than chance. For an example based on the probabilities used by the middle redundancy ($G = .61$), a participant should be able to narrow the next response to two possible positions; the probability of guessing the correct response is 1 in 2—that is, $p(\text{correct}_1) = .5$. If we assume a correct first response, the participant should be able to narrow the next response to two possible positions, and the probability of guessing correctly is again 1 in 2—that is, $p(\text{correct}_2) = .5^2 = .25$. The probability of guessing the third position is $.5^3 = .125$. Generalizing the example, the probability of guessing successive items is a decreasing geometric function. Although it is better than chance, performance would appear dismal to anyone who thought it possible that the participant might be able to recite the sequence, even though performance on the last few of the 800 trials provided ample evidence of learning. In short, as the calculation of chance during half-prompted recall illustrates, the model anticipates a participant's inability to recall the series even as it demonstrates clear evidence of learning. The combination is, of course, the pattern underlying the argument for the two-systems position.

Applying Minerva 2 to other SRT experiments

In our experiment, the sequences were generated according to an artificial grammar (i.e., probabilistically). In contrast, many experiments with the SRT procedure present a short sequence that is repeated many times over the experiment (e.g., Nissen & Bullemer, 1987; Stadler, 1992; Stadler & Neely, 1997). The fact that we gave participants probabilistic rather than repeated sequences opens the possibility that our explanation does not apply to those standard studies from the literature. In particular, if, as the implicit-learning view holds, performance in the SRT task reflects participants' learning a short repeated sequence (either explicitly or implicitly), then our model is ill equipped to predict performance. To evaluate our account against standard results, we test it against a set of experiments that use repeated sequences.

Nissen and Bullemer (1987; Experiment 1)

Nissen and Bullemer's (1987) paper is the classic application of the SRT task to implicit learning. Because it is the classic demonstration it provides a rational point of departure.

On each trial in Nissen and Bullemer's (1987) experiment, a target (an asterisk) appeared at one of four locations on a computer screen, and the participant pressed a corresponding key to identify its location. The experiment included eight blocks of 100 trials each. In a control condition, the target appeared in a random sequence. In an experimental condition, the target occurred in a repeated sequence of 10 locations. Designating the 4 locations as A, B, C, and D from left to right, the sequence followed the pattern D−B−C−A−C−B−D−C−B−A.

Mean performance of Nissen and Bullemer's (1987) participants is shown in the top panel of Figure 3. Reaction time to locate a target decreased at a faster rate when participants were presented with the structured sequence than when participants were presented with the random sequence. The results have been taken as evidence that participants learned the short repeated sequence; because they could not report the sequence in its entirety, some have argued that the participants learned the sequence implicitly. By our account, however, the advantage for the repeated sequence does not reflect implicit knowledge of the sequence but, rather, a benefit in retrieval due to a growing and cumulative history of trial-based responding.

To test our account, we applied our version of Minerva 2 to Nissen and Bullemer's (1987) sequences. The details of the simulation were unchanged from before, except that, of course,

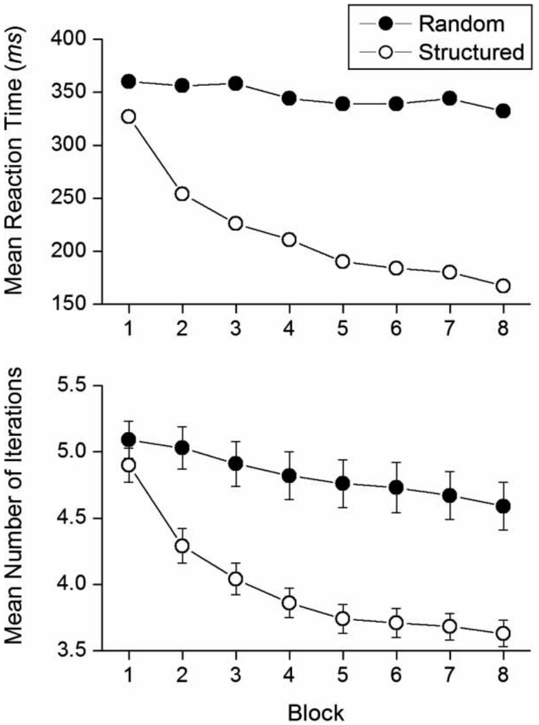

Figure 3. *The top panel shows reaction time as a function of practice and structure. Practice is expressed in blocks of 100 trials. The bottom panel shows the simulation of the experiment. Error bars for the simulated results are standard errors, n = 25. Redrawn from Cognitive Psychology, 19, M. J. Nissen and P. Bullemer, "Attentional requirements of learning: Evidence from performance measures", p. 8, © 1987, with permission from Elsevier.*

the target sequence was taken from Nissen and Bullemer (1987).

Simulated performance is shown in the bottom panel of Figure 3. The simulation reproduced the main empirical trend from Nissen and Bullemer's (1987) study: The mean number of iterations to retrieve the correct response to the probe decreased at a faster rate when the model was presented with the repeated sequence than when it was presented with the random sequence. To quantify the match between the simulated and observed performance, we computed a correlation between mean RT in the observed data and mean number of iterations in the simulation, $r^2 = .98$.

The model handled learning of a repeated sequence as well as it handled learning of a probabilistic one. Memory for the events from individual trials was sufficient to reproduce the learning advantage for a repeated sequence, over a random sequence. The simulation reinforces our conclusion that performance in the SRT task can be understood as a growing reduction of the signal-to-noise ratio in the echo with practice, without invoking a separate implicit learning mechanism.

Although Nissen and Bullemer's (1987) experiment is the classic demonstration of learning in the SRT task, it presents a general distinction: Learning is better for a redundant sequence than for a random one. Ideally, we would like to test our model against data that compare learning of repeated sequences that vary in redundancy.

Stadler (1992; Experiment 1)

Stadler (1992) tested learning of sequences that varied in redundancy. In his experiment, the target (an asterisk) appeared at one of four locations and remained until the participant pressed a corresponding response key. The screen was cleared for a short time following each response, after which the target appeared at another location. The experiment included eight blocks of 100 trials.

Stadler (1992) varied redundancy of sequences in four conditions. In a random condition, the target followed a random pattern across the 800 trials, subject to the constraint that the target could not appear at the same location on consecutive trials. The other three conditions introduced redundancy into the series by repeating a specific sequence of locations. Table 2 shows Stadler's (1992) low-, medium-, and high-redundancy sequences. For completeness, the redundancies from the zero through third orders are shown for each of the four sequences; in this context, order refers to the number of successive trials involved in the prediction. Zero-order redundancy measures predictability of the target on a single trial independently of events from preceding trials. First-order redundancy measures predictability depending on the immediately preceding trial (i.e., the same order of redundancy

Table 2. *Stadler's (1992; Experiment 1) sequences and their associated redundancies*

		Order of redundancy			
Condition	*Sequence*	*Zero*	*First*	*Second*	*Third*
Random		.00	.10	.14	.16
Low redundancy	*BDBCAB*ADAC	.01	.06	.44	.58
Medium redundancy	*BDBCAB*CDBC	.07	.33	.47	.68
High redundancy	*BDBCAB*DBCD	.07	.38	.51	.60

Note: The random sequence is not shown because it varied from simulation to simulation. Notice that the first six events in the low-, medium-, and high-redundancy sequences are identical (BDBCAB . . .). Table adapted from "Statistical structure and implicit serial learning", by M. A. Stadler, 1992, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, p. 327. © 1992 by the American Psychological Association. Adapted with permission.

as that used in Experiment 1). Second-order redundancy measures predictability of the target depending on the preceding two trials. Third-order redundancy measures predictability depending on the preceding three trials (see Attneave, 1959, for a worked example). For each of the four orders, if redundancy is equal to 0, the target is unpredictable at that order; if redundancy is equal to 1, the target is perfectly predictable at that order.

The top panel of Figure 4 shows performance of Stadler's (1992) participants (redrawn from the top panel of his Figure 1). The open circles show performance with the random sequence; closed shapes show performance with the redundant sequences. Participants' RT decreased systematically across the first seven blocks of trials, and the rate of learning increased systematically with redundancy. In Block 8, when the sequence was made random, RT increased sharply—the "negative transfer effect"—with the magnitude of the cost an increasing function of redundancy: 210 ms, 152 ms, 72 ms, and −16 ms for the high-redundancy, medium-redundancy, low-redundancy, and random conditions, respectively.

Stadler's (1992) results provide two challenges for our account. First, although we showed that our model is sensitive to differences in the redundancy of a probabilistic sequence, it remains to be seen whether it is sensitive to differences in redundancy of repeated sequences. Secondly, we have not yet tested whether the model handles a cost to performance when transferred to a random sequence; hence, it remains to be seen whether the model predicts the negative-transfer effect shown in Figure 4.

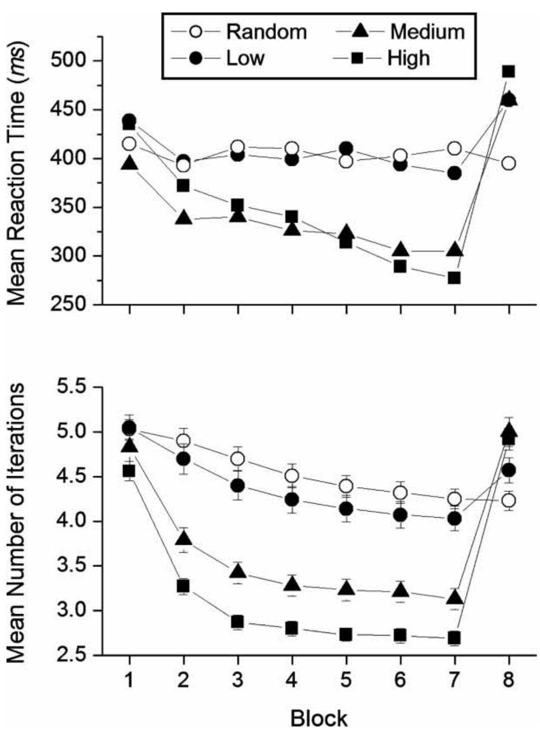

Figure 4. *The top panel shows reaction time as a function of practice and redundancy. The bottom panel shows the simulation of the experiment. Error bars for the simulation are standard errors, n = 25. Redrawn from the Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, M. A. Stadler, "Statistical structure and implicit serial learning", p. 322, © 1992 by the American Psychological Association. Adapted with permission.*

We applied our version of Minerva 2 to Stadler's (1992) task. The details of the simulation were unchanged from the earlier simulations except that the sequences presented to the model were taken from Stadler.

The results of the simulation are shown in the bottom panel of Figure 4. As shown, the model reproduced the main empirical trends: Both the rate of learning and the magnitude of the negative transfer effect increased systematically as a function of redundancy, $r^2 = .89$. The fit is very good given Stadler's concerns about the variability in baseline RT that his participants exhibited across the four conditions.

In his paper, Stadler (1992, p. 320) noted that the redundancy of the repeated sequence used by Nissen and Bullemer (1987) fell between that of the low- and medium- redundancy sequences that he used. Consistent with Stadler's (1992) note, simulated learning with the Nissen and Bullemer sequence (open circles in the bottom panel of Figure 3) does, in fact, fall intermediate to simulated learning with Stadler's low- and medium-redundancy sequences (closed circles and closed triangles, respectively, in the bottom panel of Figure 4).

Our version of Minerva 2 reproduced Stadler's (1992) results with good fidelity. Memory for the events from individual trials is sufficient to reproduce differences in the rate of learning and the size of the negative transfer effect, both as a function of sequence redundancy. Negative transfer occurred in the simulation because switching to a random sequence of trials, after having practised a repeated sequence, introduced novel trial-to-trial response contingencies into memory. Introducing those novel contingencies had the effect of increasing noise in the echo forcing a larger number of iterations to reach the decision criterion. In the next section we test the model against an experiment that provides two additional challenges to our model.

Stadler and Neely (1997; Experiment 2)

The limit on peoples' ability to recall sequences of items is recognized widely; indeed, the digit span is a standard component of IQ tests. Taking this limit as a hallmark of explicit memory, Stadler and Neely (1997) asked whether implicit learning is free of such limits by manipulating sequence length and sequence structure in a standard SRT task.

In their experiment, the stimuli were white rectangles presented on a black background. On each trial, a rectangle appeared at one of four locations. The participant's task was to identify the location of the rectangle by pressing a corresponding key on a computer keyboard.

Table 3 shows the sequences that Stadler and Neely (1997) used to define their six conditions (as well as corresponding measures of redundancy). The sequences differ in two ways: sequence redundancy (low or high) and sequence length (8-, 12-, and 16-trial patterns). The classification of a

Table 3. *Stadler and Neely's (1997; Experiment 2) sequences and their associated redundancies*

			Order of redundancy			
Sequence length	*Redundancy*	*Sequence*	*Zero*	*First*	*Second*	*Third*
8-trial	Low	*ABCDCADB*	.00	.25	.50	.63
	High	*ABCDCBCD*	.05	.38	.54	.63
12-trial	Low	*ABCDBDCABCDB*	.02	.27	.49	.59
	High	*ABCDBABCBCDB*	.06	.37	.54	.64
16-trial	Low	*ABCDACABDCABCDAC*	.02	.27	.49	.59
	High	*ABCDABCACDABCACD*	.02	.36	.55	.64

Note: The random sequence interpolated into the sequence is not shown. Table adapted from "Effects of sequence length and structure on implicit serial learning", by M. A. Stadler and C. B. Neely, 1997, *Psychological Research / Psychologische Forschung*, *60*, p. 17, © 1997. Adapted with kind permission of Springer Science + Business Media.

sequence as high or low redundancy was made relative only to the redundancy of the alternate sequence of the same length.

In the experiment, participants received a total of nine blocks of trials. The first five blocks of trials involved a repeated sequence, the sixth block involved a random sequence, the seventh and eighth blocks reintroduced the repeated sequence, and a final block involved a new random sequence. Each block included 10 consecutive presentations of the relevant sequence; thus, for sequences with 8-, 12-, and 16-trials, a block included 80, 120, and 160 trials, respectively.

Figure 5 presents a facsimile of the mean number of milliseconds that participants took to locate the target as a function of practice, sequence length, and redundancy (redrawn from Stadler & Neely's, 1997, Figure 2). Performance in the top, middle, and bottom panels of Figure 5 corresponds to the 8-, 12-, and 16-trial patterns in Table 3, respectively. As shown, RT decreased in blocks with the repeated sequence, increased sharply in blocks with the random sequence (Block 6 and again in Block 9), and was systematically better for the high- than for the low-redundancy patterns.

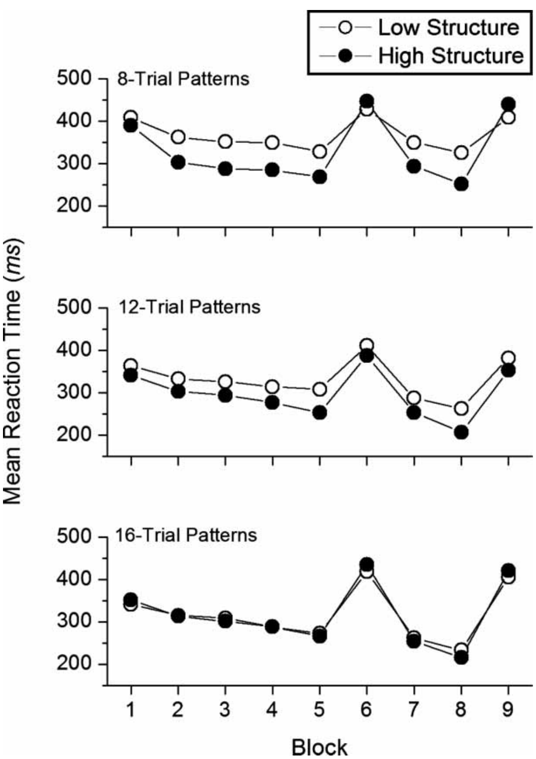

Figure 5. *Reaction time as a function of practice, repetition pattern, and structure (high/low). Standard errors were not reported in the original manuscript. Redrawn from Psychological Research / Psychologische Forschung, 60, 1997, p. 19, "Effects of sequence length and structure on implicit serial learning", M. A. Stadler and C. B. Neely, © 1997, with kind permission of Springer Science + Business Media.*

Stadler and Neely's (1997) results pose two challenges to our model. First, when learning is measured in terms of the difference between performance on the first block of random trials (Block 6) and on the preceding block of repeated trials (Block 5), it appears as if "learning actually increased as [sequence] length increased" (p. 20). Hence, as anticipated by the implicit-learning view, performance in the SRT task does not appear to obey the same capacity limitation as that for explicit recall (e.g., digit span): If recall is limited to 7 $\pm$ 2 items (Miller, 1956), there should be a cost rather than a benefit to learning as sequence length is increased. Stadler and Neely's results, then, appear to provide strong evidence for a fundamental difference between implicit and explicit learning. From our perspective, however, there is no conflict. Rather, speeded responding in the SRT task reflects facilitation in retrieval as the trial-based history of responding cumulates in memory.

Second, following a block of random trials (Block 6) people show a rapid and full recovery when the repeated sequence is reintroduced (Block 7). Rapid and full recovery is problematic for some models of learning: a phenomenon that is known as catastrophic interference (McClosky & Cohen, 1989; Mewhort, 1990; Ratcliff, 1990). Because we did not test recovery from negative transfer in our previous simulations, it remains an open question whether the model recovers when the repeated sequence is reintroduced.

To test whether our account handles the challenges posed by Stadler and Neely's (1997) study, we applied our version of Minerva 2 to their design. The details of the simulation were the same as before except that the sequences were based on the ones from Table 3.

The results of the simulation are presented in Figure 6. The simulation captured the three main trends from Stadler and Neely's (1997) experiment: RT decreased across trials in blocks with the repeated sequence, increased sharply in blocks with the random trials (Block 6 and again in Block 9), and was systematically better for the high- than for the low-redundancy patterns, $r^2 = .92$. We conclude that retrieval from memory for the events from individual trials was sufficient to predict learning in Stadler and Neely's experiment.

As we noted earlier, Stadler and Neely's (1997) experiment provided two challenges to our position. The first challenge concerned an apparent discrepancy in the capacity limitation of explicit recall (e.g., a digit span task) and an apparent reversal from that limitation in an implicit-learning task (e.g., SRT performance). We reproduced the apparent reversal using a cued-recall mechanism, without invoking an implicit-learning system. We did not need to include an additional mechanism to capture Stadler and Neely's data. The second challenge to our position was the demonstration of rapid and full recovery of learning when the repeated sequence was reintroduced. Our version of Minerva 2 recovered rapidly and fully from negative transfer without any modification or extension. We conclude that the model does not suffer from catastrophic interference.

The simulations reinforce our previous conclusion: Performance in the SRT task can be understood as facilitation at retrieval that emerges from the participant's growing history of his, or her, own responses to particular stimuli (cf. Willingham, 1999).

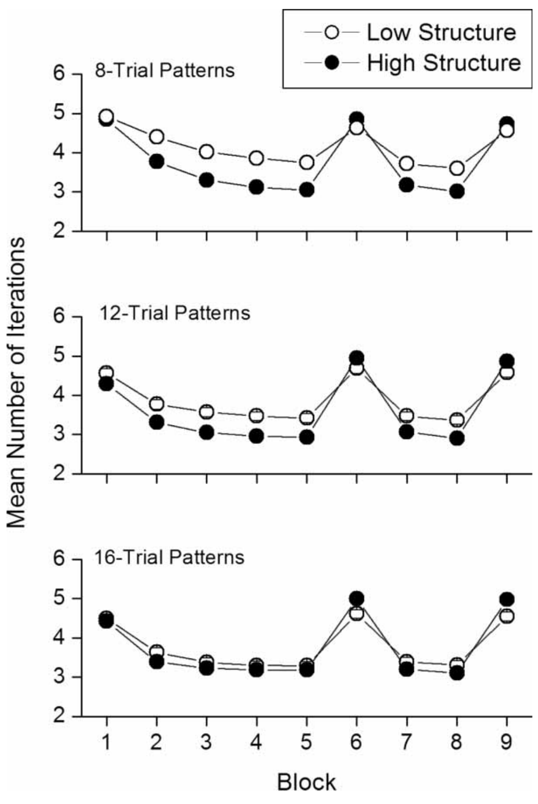

Figure 6. *Simulated reaction time in Stadler and Neely's (1997) Experiment 2 as a function of practice, repetition pattern, and structure (high/low). Error bars are standard errors,* n = *25.*

Hyman (1953)

So far, we have discussed the SRT task in terms of the implicit-learning issue, a frame introduced by Nissen and Bullemer (1987). The same task was used three decades earlier, however, to address a different issue: the time required to execute a mental process (e.g., see Fitts & Posner, 1967, chap. 6). In this context, the participants were given extensive practice, the probability structure was described, and the task was learned to a criterion of perfect performance. From the implicit-learning perspective, learning such a regime could hardly be called implicit. However, our claim to understand performance in tasks thought to tap an implicit-learning system rests on the assumption that the mechanisms of the model work when applied to an explicit-learning task. If that assumption is sound, the model should capture examples of SRT performance drawn from explicit-learning tasks, such as those that examined the time to execute a deliberate mental process.

In an early experiment on speeded choice, Merkel (1885) reported a logarithmic increase in response latency as function of the number of stimuli to be identified. Not much was made of the result, however, until Hick (1952) interpreted the relation in terms of communication theory. In his study, Hick used an array of lights; each light was associated with a unique response. On each trial, one of the lights was brightened, and the participant identified the light as quickly as possible (see Reynolds, 2004, for a photograph of Hick's equipment). Hick systematically varied the number of lights and replicated Merkel's findings. Instead of plotting response latency as a function of the number of alternatives, however, Hick plotted response latency as a function of the uncertainty associated with the number of stimulus alternatives.

Hick (1952) interpreted the linear latency–uncertainty relation in terms of a series of binary decisions (see also Swanson & Briggs, 1969). When uncertainty is measured in bits, it specifies the minimum number of optimal binary questions—questions that each eliminate half the options and that can be answered by a yes or no—required to specify a stimulus uniquely from a set of alternatives. With that idea in mind, Hick proposed that the latency–uncertainty relation reflected the time that participants needed to compute the decisions.

Uncertainty associated with a fixed number of alternatives can be reduced by using some alternatives more frequently than others or, as illustrated in the SRT tasks examined earlier, by introducing sequential dependencies among the alternatives. For example, if two of four alternatives each occur 40% of the time, and the other two each occur 10% of the time, uncertainty is reduced from the 2.0 bits associated with four equally likely alternatives to 1.722 bits. If response latency is an increasing function of uncertainty, reducing uncertainty should yield faster responding.

Hyman (1953) tested Hick's (1952) claim that response latency is related to uncertainty by manipulating the number of equally probable stimuli (Experiment 1), the relative frequencies with which the stimuli occurred (Experiment 2), and the sequential dependencies among occurrences of successive stimuli (Experiment 3). He reasoned that if uncertainty predicts response latency, the data should fall on the same function regardless of how the predictability was introduced.

In Hyman's (1953) task, the stimuli were eight lights taken from the corners of a six-by-six matrix: four lights from the outer corners of the full matrix and four from the inner four-by-four matrix. He taught the participants to associate each light with a unique syllable: *Bun*, *Boo*, *Bee*, *Bore*, *By*, *Bix*, *Bev*, and *Bate*. Once participants had learned to associate each light with its response, they named the lights in a speeded task under 24 uncertainty conditions (8 conditions for each of three conceptually separate experiments that were run concurrently). The procedure on each trial was common to all conditions: A light was brightened, and the participant named it as quickly as possible by saying the associated syllable.

The 24 uncertainty conditions were administered to 4 participants who each completed 40 experimental sessions run over a 3-month period. To avoid the problem of interpreting response latencies contaminated by a speed–accuracy tradeoff, the data were taken from error-free blocks of approximately 125 trials (the exact number depended on the uncertainty condition). To obtain error-free performance and to avoid a shift in performance reflecting sudden insight into the structure of a condition—a clear departure from the procedure in SRT studies of implicit learning—Hyman (1953) described the probability structure for each condition to the participant, and he gave them extensive practice with each condition (1 participant practised naming the lights for 1 month before beginning the experiment). The details of the sequences that Hyman presented to participants can be found in his Tables 2 and 3.

The results of Hyman's (1953) experiment are reproduced in Figure 7.[8] Each panel shows response latency as a function of stimulus uncertainty for 1 of his 4 participants. As is clear in

[8] There is an error in Hyman's (1953) figure. The panel for participant L.S. shows nine data points for Experiment 2 and seven for Experiment 1. The data point at $U = 2.81$ bits for Experiment 2 should be attributed to Experiment 1. We fixed the error in Figure 7.

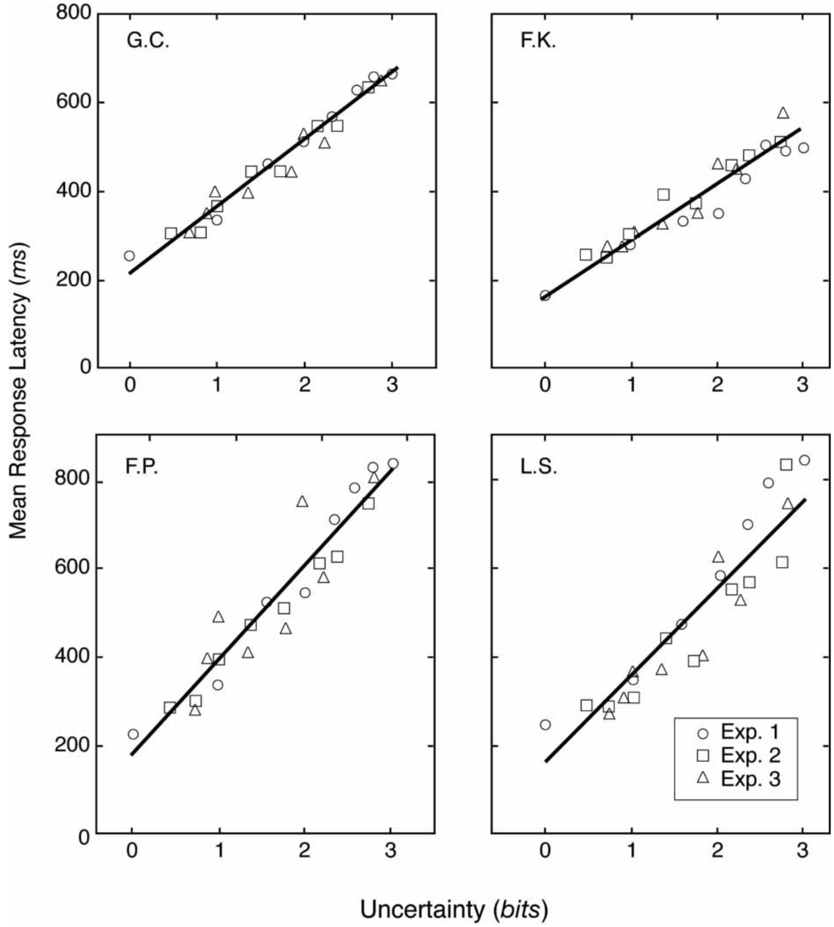

Figure 7. *Mean reaction time as a function of stimulus uncertainty. Redrawn from "Stimulus information as a determinant of reaction time", R. Hyman, 1953, Journal of Experimental Psychology, 45, p. 192, © 1953 by the American Psychological Association. Adapted with permission.*

Figure 7, response latency was a linear function of uncertainty, regardless of how predictability was introduced. The relation plotted in Figure 7 is widely known as the Hick–Hyman law. An implication of the Hick–Hyman law (and of communication theory generally) is that a stimulus's role in guiding performance depends on how people treat it in relation to its alternatives, not on its characteristics as a stand-alone event (Garner, 1962, 1974; Pomerantz & Lockhead, 1991).[9]

Hyman's (1953) classic experiment provides an interesting test for our explanation of performance

[9] Others have argued that Hick's results reflect a strategy based on a speed–accuracy trade-off (Fitts, 1966; see also Fitts & Posner, 1967). Because the chance of a response error increases with the number of response alternatives and because the participants were given opportunity to note the number of alternatives, they may have delayed their response conditional on the number of alternatives to ensure that they respond correctly. Usher, Olami, and McLelland (2002) have worked out the details of the trade-off idea in the context of a diffusion-based information accumulation model (see Brown & Heathcote, 2005, for a review of information accumulation models). Although their account handles Hick's data successfully, the trade-off idea has no way to anticipate

in the SRT task. If the explanation we have developed is consistent with performance when participants are informed of contingencies in the stimulus series, our adaptation of Minerva 2 should be capable of fitting performance in the classic experiment.

To simulate Hyman's (1953) experiment, we set $L = 1$. We assumed perfect encoding to acknowledge the large amount of practice given to the observers. In addition, for the 24 separate simulations (1 for each of the 24 contingency conditions), we stocked memory with all eight possible stimulus–response pairs and with the stimulus–response pairs for each of the 24 uncertainty conditions (but with an empty subfield for the previous response). We stocked memory in this way so that each of the 24 separate contingency conditions would start with the same basic knowledge of the task. Before conducting a simulation, we gave the model practice with a sequence of 500 trials in which the eight stimulus alternatives were presented and followed one another randomly. The practice was intended to acknowledge the several months of practice naming the lights before data were taken in Hyman's experiments. For each of the 500 practice trials, we recorded all three subfields in each trace in memory. After practice, we ran a full block of 500 trials under the conditions appropriate to the uncertainty condition.

We ran each of the 24 simulations four times as if they represented 4 different participants. On each of the four replications, we generated fresh random vectors for the stimuli and responses. Hence, variability among the four pseudoparticipants illustrates the variability in the model inherent in our use of random vectors to represent the events in the experiment.

Figure 8 shows the results of the simulations. Points corresponding to conditions in Hyman's Experiments 1, 2, and 3 are shown by circles, squares, and triangles, respectively. Simulated performance was error free (as were the data in Hyman's task). Each panel in the figure shows the data for one pseudoparticipant. As is clear, each pseudoparticipant showed the same pattern of results: Simulated response time was an increasing linear function of stimulus uncertainty, the same pattern as that demonstrated by Hyman's (1953) participants. We obtained RT estimates of both observed and simulated performance in each of the 24 conditions in Hyman's study. The fit for the simulated and observed means computed across the 24 means was high, $r^2 = .89$.

As before, performance in the SRT task can be understood in terms of information about the correct response that becomes available from the participant's history of responding to particular stimuli. The history includes the local context of each response (i.e., the immediately preceding response), and it provides a reliable source of information to the extent that the sequence of successive stimuli is predictable. Learning in the SRT task does not require an implicit system to abstract the rules. Instead, it can be understood in terms of the participant's ability to retrieve useful information from his or her history of responding. The same principles extend to an explanation of speeded choice when participants are fully aware of contingencies.

Reflecting on 50 years of research in psychology, Luce (2003) noted that the early successes of information theory led many to conclude that a very deep truth had been discovered. He argued that information theory has not lived up to its promise and remained pessimistic for its future. The Hick–Hyman law illustrates the dilemma nicely. Uncertainty predicts response latency, but prediction, by itself, does not explain anything.

Unlike Luce (2003), we think that information theory has an important future in psychology

changes in RT based on sequential dependencies, as in Hyman's data. To make the trade-off idea work for Hyman's task, one would have to suppose that participants adjust response times conditional on sequential dependencies. The problem is that there are no cues to the dependencies equivalent the number of alternatives in Hick's task. In short, when applied to the Hick–Hyman law, the trade-off idea begs the question. Hence, we see no way that the speed–accuracy trade-off idea could handle Hyman's (1953) demonstration that stimulus uncertainty, not number of stimuli, predicts response latency.

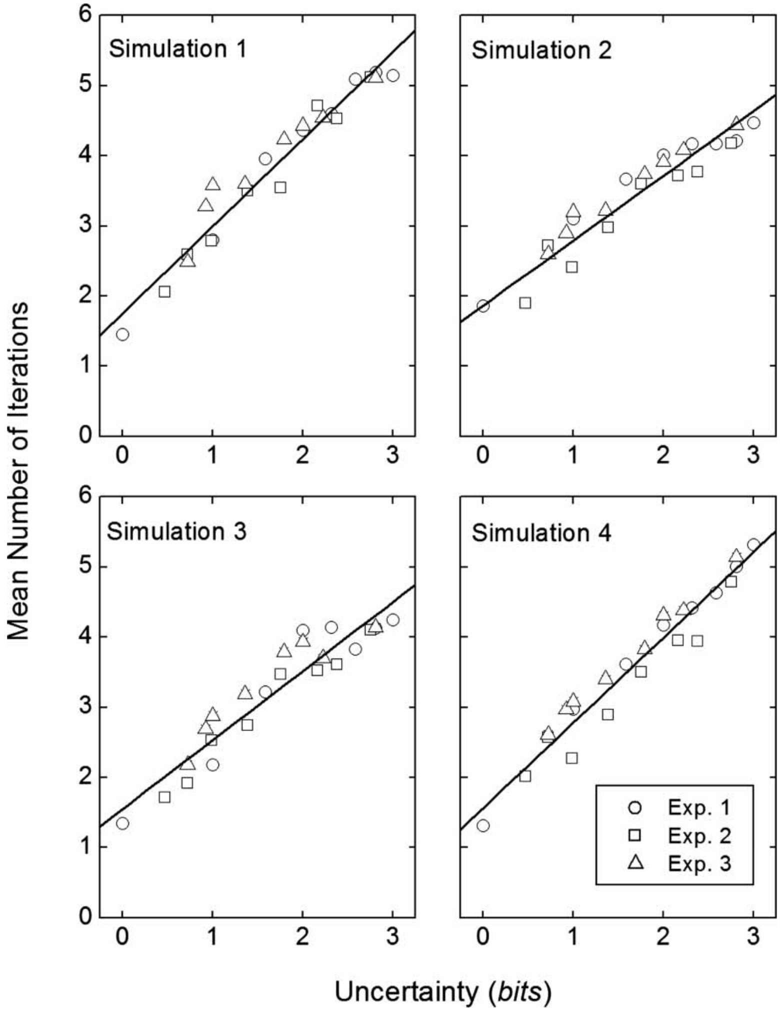

Figure 8. *Simulated response time as a function of stimulus uncertainty. Each panel represents an independent replication of the 24 uncertainty conditions.*

because it quantifies structure remarkably well (e.g., Jamieson & Mewhort, 2005, 2009; Pothos & Bailey, 1999; Stadler, 1992; Stadler & Neely, 1997; Tulving, 1962). That said, we concur with Luce that the theory has not yet lived up to its promise. We see two main reasons. First, the early authorities confused prediction for explanation based on a new metric—information measured in bits. The ability to quantify structure is a critical first step for understanding its relation to performance, but an explanation requires a mechanism to link structure to behaviour. Secondly, the early authorities identified the wrong locus at which to search for a mechanism. Namely, they framed the problem as a perceptual issue: Hake and Hyman's (1953) companion paper to Hyman's (1953) paper, for example, was entitled "Perception of the statistical structure of a random series of binary symbols". As the present simulation illustrates, the use of statistical structure reflects what we know and how we retrieve information from memory. The mechanism responsible for the use of redundancy in stimuli is memorial, not perceptual. Indeed, when an SRT task is run so that the response can be selected without having to look it up from memory, contingencies in the series no longer affect response time (Kveraga, Boucher, & Hughes, 2002).

GENERAL DISCUSSION

As participants practise identifying targets in a redundant series, their response time decreases. Because the participants cannot characterize the rules that make the sequence redundant or anticipate the next items in the series, even after hundreds of practice trials, the learning implied by speeded response time is taken widely to reflect implicit knowledge of the structure in the sequence. We have confirmed the standard results in a new experiment and have adapted an exemplar model of retrieval from memory to understand performance in it. The model provided a close fit to performance in our target experiment and experiments conducted in other labs.

According to the model, after each response, participants store a trace of the event that comprises the current stimulus, the response associated with it, and the context provided by an immediately preceding response. When the next stimulus is presented, the participants use it to retrieve the correct response. As they practise the task, the redundancy of the series provides extra information pointing to the correct response, and the extra information is responsible for the decrease in response time.

The model handled performance with materials constructed using simple sequential grammars and using a repeated series. The fit to our own experiment, and to the results of Nissen and Bullemer (1987), Stadler, (1997), and Stadler and Neely (1997), were all obtained without changing representation, encoding, or retrieval assumptions in the model. The same model also accommodated the long-established linear relation between response time and stimulus uncertainty, a function known as the Hick–Hyman law.

We conclude that learning in a SRT task does not point to a specialized implicit learning system. Instead, we suggest that participants use very local memory for events to speed responding. The larger implication is that performance in the SRT task can be explained by the same principles used in explicit-memory tasks. In an earlier paper (Jamieson & Mewhort, 2009), we showed that the principles also accommodate performance in a judgement-of-grammaticality task.

Stadler argued in 1992 that the judgement-of-grammaticality and the SRT task might be accommodated by the same representation assumptions and learning operations. The work presented here confirms Stadler's (1992) position: Performance in both tasks can be understood by storage of events from a single trial and parallel retrieval from memory. The present work takes his argument one step further by showing that the representation and retrieval assumptions are the same as those needed to accommodate a number of other heretofore seemingly unrelated learning and remembering behaviours to which Minerva 2 has been applied.

We have modelled performance as a process by which information retrieved from memory is clarified by an increasingly trace-specific search over time. By contrast, models of RT are often based on the idea that decision depends on information accumulated across time (e.g., Brown & Heathcote, 2005). Information accumulation models were developed in the context of perceptual decisions (e.g., Ratcliff, 2006; Vickers, 1970). The basic idea is that the perceptual system provides a signal that is too weak to support an immediate decision. As a result, the signal must be cumulated (by resampling) until sufficient evidence is obtained. Specific models based on the evidence-resampling idea successfully accommodate a variety of phenomena including the shape of the RT distribution and speed–accuracy trade-off. To date, however, none have described how the resampling idea could accommodate redundancy in a series of stimuli.

It might be attractive, therefore, to combine elements of the present model with elements of the resampling idea. The idea would be to create a hybrid that could inherit the best properties of both its parents.

A hybrid model is clearly an idea that deserves to be explored. Development of the idea, however, is well beyond the scope of the present work. The present simulations are sufficient to make our main point: Performance in the SRT task does not require an implicit system to abstract rules underlying redundancy. That said, a hybrid model might be too much of a marriage of convenience. First, the resampling idea seems better suited to a perceptual example than to a memorial one. It is easy to imagine that a perceptual transducer generates a signal that is too weak to support an immediate decision. As illustrated in Experiment 1, and in the other examples explored here, however, the initial evidence was strong enough to support error-free performance. Nevertheless, there was a clear performance gain as a function of practice and stimulus redundancy. Such a gain in performance does not map easily onto the idea of a weak signal that is gradually strengthened until it supports a decision. Rather, the gain suggests a strong initial signal that was made even stronger when predictability was introduced into the sequence (and sufficient practice allowed the participant to use it). In our account, stimulus redundancy helps performance because it modifies the rate at which the signal-to-noise level reaches criterion during retrieval: Redundancy increases the signal-to-noise ratio in the echo by reducing the contribution of traces that predict opposing responses to the probe. Our simulations are based on noise reduction during retrieval whereas the resampling models are based on signal amplification during signal accumulation. It is not clear whether the two ideas should be combined, and, to the extent that they are thought to conflict, it is certainly not clear what experimental evidence might distinguish the two positions.

Throughout, we used sequences with first-order sequential redundancy. The model can be extended to handle higher order redundancies. There is some evidence that the model already handles second-order sequential structure, but our examples of second-order sequential structure are confounded with lower order structure, a problem almost always found in nature. Nevertheless, we do not wish to make a too-strong claim regarding higher order contingencies. Remillard (2008) has recently unconfounded orders of structure and thereby extended the kind of constraint that can be used in SRT tasks. We are currently testing the model to see whether we can extend it to materials based on Remillard's methods.

In a thoughtful and comprehensive review, Cleeremans and Dienes (2008) focused attention on recently developed computational models of implicit learning. It is informative to highlight the main difference between our account for the SRT task and the accounts that they reviewed. Although the details differ, virtually all of the other models focus on how the subject can learn the series as a series. In effect, many models treat the SRT task as an example of an immediate serial recall task. In our model, participants do not learn anything about the series as a series; instead, they use local information about their own behaviour. Local information is sufficient to reduce response time across trials, but it is insufficient to allow a participant to recite the series when prompted.

The contrast in approaches highlights one of our main points. Our simulations show how performance in SRT tasks can be understood without invoking two separate learning systems: Participants know the mapping of the stimulus to the response, and they retrieve the response to the current probe. The speed of retrieval improves because the structure in the series restricts the information contributed by competing responses. The models reviewed by Cleeremans and Dienes (2008) also simulate performance, but, importantly, they do so by first accepting the idea of two learning systems and the idea that learning develops a representation of the rules that underlies the sequence: Cleeremans, Servan-Schreiber, and McClelland (1989) showed that a recurrent network trained on grammatical strings develops a pattern of activation on its hidden units that corresponds to the grammar (see also Perruchet & Pacton, 2007).

How can we distinguish our position from the position reviewed by Cleeremans and Dienes (2008)? One strategy would be to the fit competing accounts to the same data and to select the winner based on a goodness-of-fit criterion. However, given that both classes of model predict learning, a decision based on goodness of fit is unlikely to convince anyone. Rather, the question is an empirical one: We must identify conflicting predictions from the two accounts and then test the predictions to choose which view is correct.

Throughout this paper, we have accepted Nissen and Bullemer's (1987) view that the SRT task provides a fair test of the implicit-learning position (a view shared in all the papers modelled here). That view is, of course, open to question. Some theorists suggest that the SRT task can sometimes tap explicit learning as well as implicit learning and try to distinguish the possibilities in terms of a participant's awareness (e.g., Destrebecqz & Cleeremans, 2001). To assess awareness, we asked our participants to tell us as much as they could about the sequence presented to them. That technique is time honoured, but it may be insufficient. In our view, it is difficult, if not impossible, to achieve a psychophysics of awareness that is both convincing and theoretically neutral.

By contrast, Destrebecqz and Cleeremans (2001) used an indirect way to measure awareness and argued that the measure correlates with use of the implicit system. An alternative possibility is that the measure has little to do with the implicit/awareness dimension but rather correlates with the participant's preferred strategy for the SRT task: Under some circumstances, people try to learn something about the series, and under others they are content to use local information, as described in our model. We have no doubt that people can learn about a series if pressed to do so, only that they are unlikely to do so without provocation or purpose. In any case, our point remains that performance in the SRT tasks modelled here can be understood in terms of the use of local information during retrieval from memory. As a measure of private experience, awareness needs an explanation but, by definition, has no explanatory power.

The work presented here, and the work on which it builds (Jamieson & Mewhort, 2005, 2009), provides a common account of putatively distinct "implicit" and "explicit" tasks. Jamieson and Mewhort (2005) studied recall of sequences of symbols ordered by a grammar. They quantified the redundancy of the grammar and of the exemplars derived from it. The two measures are confounded, but when they were separated, local redundancy, rather than the redundancy of the grammar itself, predicted performance. Such results argue against the view that people abstract the grammar implicitly. Instead, they suggest that participants use local information from the exemplars themselves. Jamieson and Mewhort (2009) showed that Minerva 2's mechanism for retrieval from memory predicts performance in the judgement-of-grammaticality task; Pothos and Bailey (2000) have demonstrated very similar results using Nosofsky's (1986) classification model. Finally, the present work shows that the principles of retrieval from memory anticipate performance in the SRT task. We are encouraged that common principles allow us to understand performance across so many tasks and behaviours and, in particular, across tasks often distinguished as "implicit" and "explicit". Nature often uses a single mechanism for many purposes.

Original manuscript received 13 May 2008
Accepted revision received 17 September 2008
First published online 13 February 2009

REFERENCES

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory & Language*, *39*, 371–391.

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results*. New York: Holt, Rinehart, & Winston.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber

(1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, *120*, 316–323.

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*, 117–128.

Clark, S. E. (1997). A familiarity-based account of confidence–accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 232–238.

Cleeremans, A., & Dienes, Z. (2008). Computational models of implicit learning. In R. Sun (Ed.), *The Cambridge handbook of computational modeling* (pp. 396–421). New York: Cambridge University Press.

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

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. (1992). Connectionist and memory-array models of artificial grammar learning. *Cognitive Science*, *16*, 41–79.

Dienes, Z., Broadbent, D. E., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 875–887.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*, 180–209.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*, 91–101.

Fitts, P. M. (1966). Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, *71*, 849–857.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks-Cole.

Gaito, J. (1965). Unequal intervals and unequal *n* in trend analysis. *Psychological Bulletin*, *63*, 125–127.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.

Garner, W. R. (1974). *The processing of information and structure*. New York: Wiley.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Hake, H. W., & Hyman, R. (1953). Perception of the statistical structure of a random series of binary symbols. *Journal of Experimental Psychology*, *45*, 64–74.

Heathcote, A. (1988). Screen control and timing routines for the IBM microcomputer family using a high-level language. *Behavior Research Methods, Instruments, & Computers*, *20*, 408–412.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*, 11–26.

Hintzman, D. L. (1984). MINERVA-2: A simulation model of human memory. *Behavior Research Methods, Instruments & Computers*, *16*, 96–101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Hintzman, D. L. (1987). Recognition and recall in MINERVA-2: Analysis of the "recognition failure" paradigm. In P. E. Morris (Ed.), *Modelling cognition* (pp. 215–229). London, Wiley.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.

Hintzman, D. L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology*, *41*, 109–139.

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

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning & Verbal Behavior*, *22*, 485–508.

Jamieson, R. K., & Mewhort, D. J. K. (2005). The influence of grammatical, local, and organizational redundancy on implicit learning: An analysis using information theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 9–23.

Jamieson, R. K., & Mewhort, D. J. K. (2009). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology*, *62*, 550–575.

Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 524–531.

Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology*, *42*, 61–112.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia:

Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*, 172–179.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 79–91.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 169–181.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of non-local musical rules: Implicitly learning more than chunks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1417–1432.

Kveraga, K., Boucher, L., & Hughes, H. C. (2002). Saccades operate in violation of Hick's law. *Experimental Brain Research*, *146*, 307–314.

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, *12*, 703–710.

Kwantes, P. J., & Mewhort, D. J. K. (1999). Modeling lexical decision and word naming as a retrieval process. *Canadian Journal of Experimental Psychology*, *53*, 306–315.

Kwantes, P. J., & Neal, A. (2006). Why people underestimate y when extrapolating in linear functions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1019–1030.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Luce, R. D. (2003). Whatever happened to information theory in psychology? *Review of General Psychology*, *7*, 183–188.

Manza, L., & Reber, A. S. (1997). Representing artificial grammars: Transfer across stimulus forms and modalities. In Diane C. Berry (Ed.), *How implicit is implicit learning?* (pp. 73–106). Oxford, UK: Oxford University Press.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1083–1100.

McAndrews, M. P., & Moscovitch, M. (1985). Rule-based and exemplar classification in artificial grammar learning. *Memory and Cognition*, *13*, 469–475.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109–165). New York: Academic Press.

Merkel, J. (1885). Die zeitlichen Verhältnisse der Willensthätigkeit. [The temporal relations of activities of the will]. *Philosophische Studien*, *2*, 73–127.

Mewhort, D. J. K. (1990). Alice in Wonderland, or psychology among the information sciences. *Psychological Research*, *52*, 158–162.

Mewhort, D. J. K., & Johns, E. E. (2005). Sharpening the echo: An iterative resonance model for short-term recognition memory. *Memory*, *13*, 300–307.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

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

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Perruchet, P., & Pacton, S. (2005). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*, 233–238.

Pomerantz, J. R., & Lockhead, G. R. (1991). Perception of structure: An overview. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 1–20). Washington, DC: American Psychological Association.

Pothos, E. M., & Bailey, T. M. (1999). An entropy model of artificial grammar learning. In *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society*, (pp. 549–554). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pothos, E. M., & Bailey, T. M. (2000). The importance of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 847–862.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review*, *97*, 285–308.

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, *53*, 195–237.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, *6*, 855–863.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, *81*, 115–119.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.

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

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. (2008). Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential dependencies. *Quarterly Journal of Experimental Psychology*, *61*, 400–424.

Reynolds, L. (2004). Wellcome witnesses: The Medical Research Council Applied Psychology Unit. *History of Psychology*, *7*, 85–93.

Rossnagel, C. S. (2001). Revealing hidden covariation detection: Evidence for implicit abstraction at study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1276–1288.

Semon, R. (1923). *Mnemic psychology* (B. Duffy, Trans.). London: George Allen & Unwin. (Original work published 1909.)

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

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

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 and 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.

Stadler, M. A., & Neely, C. B. (1997). Effects of sequence length and structure on implicit serial learning. *Psychological Research/Psychologische Forschung*, *60*, 14–23.

St. John, M. F., & Shanks, D. R. (1997). Implicit learning from an information-processing standpoint. In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 162–194). Oxford, UK: Oxford University Press.

Swanson, J. M., & Briggs, G. E. (1969). Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, *81*, 223–229.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, *69*, 344–354.

Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed–accuracy tradeoff. *Journal of Mathematical Psychology*, *46*, 704–715.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*, 37–58.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 328–344.

Vokey, J. R., & Brooks, L. R. (1994). Fragmentary knowledge and the processing-specific control of structural sensitivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1504–1510.

Vokey, J. R., & Higham, P. A. (2005). Abstract analogies and positive transfer in artificial grammar learning. *Canadian Journal of Experimental Psychology*, *59*, 54–61.

Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, *122*, 227–248.

Whittlesea, B. W. A., & Wright, R. L. (1997). Implicit (and explicit) learning: Acting adaptively without knowing the consequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 181–200.

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory and Cognition*, *27*, 561–572.

Wright, R. L., & Whittlesea, B. W. A. (1998). Implicit learning of complex structures: Active adaptation and selective processing in acquisition and application. *Memory and Cognition*, *26*, 402–420.