

Comparing Direct and Indirect Measures of Sequence Learning

Luis Jiménez and Cástor Méndez
Universidad de Santiago

Axel Cleeremans
Université Libre de Bruxelles

Comparing the sensitivity of similar direct and indirect measures is proposed as the best way to provide evidence for unconscious learning. The authors apply this approach, first proposed by E. M. Reingold and P. M. Merikle (1988), to a choice reaction-time task in which the material is generated probabilistically on the basis of a finite-state grammar (A. Cleeremans, 1993). The data show that participants can learn about the structure of the stimulus material over training with the choice reaction-time task, but only to a limited extent—a result that is well predicted by the simple recurrent network model of A. Cleeremans and J. L. McClelland (1991). Participants can also use some of this knowledge to perform a subsequent generation task. However, detailed partial correlational analyses that control for knowledge as assessed by the generation task show that large effects of sequence learning are exclusively expressed through reaction time. This result suggests that at least some of this learning cannot be characterized as conscious.

Over the last 10 years, interest in human learning has steadily increased, thanks in part to the development of connectionism and to renewed attention to the *cognitive unconscious* (Reber, 1993). Indeed, there is now a large body of experimental evidence that suggests that people are able to develop sensitivity to complex stimulus covariations without intention to learn or even without awareness that learning is taking place (see Berry, 1994; Reber, 1989a, for extensive reviews). In system-control tasks, participants can learn to control a simulated system without being able to answer explicit questions about the behavior of the system (e.g., Berry & Broadbent, 1988). Artificial grammar-learning studies have shown that participants can classify strings of letters as grammatical or not grammatical after practice at memorizing similar strings and without being able to report on the rules that define grammaticality (e.g., Mathews et al., 1989). Finally, sequence-learning studies have demonstrated that participants can become sensitive to the regularities contained in sequences of stimuli presented in a choice-reaction setting despite remaining unable to report on the sequence or to perform well in other direct tests, such as generation, where they are asked to predict the next stimulus instead of reacting to the current one (e.g., Nissen & Bullemer, 1987).

Despite this wealth of research, little progress has been

accomplished on what is arguably the defining feature of implicit learning, that is, the independence of the resulting knowledge from conscious experience and awareness. As Shanks and St. John (1994) discussed extensively, the basic problem appears to be one of methodology and interpretation: How does one establish that learning is unconscious in the absence of any clear and accepted criteria of awareness? Which empirical measures would best reflect the operation of implicit or explicit learning processes?

The goal of this article is to reflect on the conditions required to demonstrate unconscious learning and to present new experimental data aimed at fulfilling these conditions. We also present simulation work with the simple recurrent network (SRN; Cleeremans & McClelland, 1991; Elman, 1990) to assess how well this model is able to account for various aspects of performance in this situation.

Many authors have started to reflect on the conditions required to compare performance on different tests of implicit and explicit knowledge. The typical problem can be described as follows: Assume that participants in an experiment are first presented with a learning task during which they are required to process (e.g., memorize) a set of stimuli. Over training, learning is assessed by some measure of performance. After a given amount of practice, they are then assessed for their explicit, reportable knowledge of some features of the training material. What kind of test of explicit knowledge should be used in this context? One could argue, along with Reber (1989b), that free verbal reports are the only good measures of the contents of awareness because other possible measures (e.g., recognition or discrimination) all tend to involve some degree of contextual cueing and hence are not free from the influence of potential unconscious determinants (cf. Reber, Allen, & Regan, 1985). On the other hand, one can argue just as well that free reports do not reflect all the contents of awareness on which performance is based and thus that more sensitive and directed tests (e.g., forced-choice discrimination) should be used when assessing explicit knowledge (see for instance Dulany, Carlson, & Dewey, 1984, 1985; Perruchet & Pacteau, 1990, 1991).

Shanks & St. John (1994) analyzed these issues by arguing

Luis Jiménez and Cástor Méndez, Facultad de Psicología, Universidad de Santiago, Santiago de Compostela, Spain; Axel Cleeremans, Université Libre de Bruxelles, Brussels, Belgium.

Portions of the data reported in this article were previously presented at the 16th Annual Conference of the Cognitive Science Society, Atlanta, Georgia, August 1994. This work was partially supported by Grant D.O.G. 1-10-1989 from the Consellería de Educación e Ordenación Universitaria da Xunta de Galicia (Spain). Axel Cleeremans is a Research Associate of the National Fund for Scientific Research, Belgium. We thank Alain Content, Zoltán Dienes, and Pierre Perruchet, for providing valuable comments on previous versions of this article.

Correspondence concerning this article should be addressed to Luis Jiménez, Facultad de Psicología, Universidad de Santiago, 15706 Santiago de Compostela, Spain. Electronic mail may be sent via Internet to jimenez@usc.es.

that valid demonstrations of unconscious learning should be based on dissociations between measures of implicit learning and awareness that satisfy two criteria: the information criterion and the sensitivity criterion. The information criterion requires that awareness tests should demonstrably tap on the same knowledge that was needed to support performance in the corresponding implicit test. The sensitivity criterion requires that awareness tests should demonstrably be sensitive to all of a participant's conscious knowledge.

Consider now how any given measure could simultaneously comply with these two criteria. Verbal reports, for instance, obviously do not fit the information criterion in that the experimenters cannot possibly guarantee that their questions are not inducing participants to respond on the basis of information that differs from what they used during learning. Verbal reports also do not fit the sensitivity criterion in that there is no way to guarantee that all of a participant's conscious knowledge will be reported. Unfortunately, however, this latter problem applies to every single measure of awareness as there is simply no way to ascertain whether participants use all of their relevant explicit knowledge when responding to a given verbal or discriminative test, even if the experimenter selects the most sensitive measure available. Hence, it may turn out to be impossible to find a single measure that is simultaneously (a) exhaustively sensitive to the relevant contents of awareness and (b) exclusively sensitive to this knowledge.

This ongoing debate about the validity of different methodological approaches to the assessment of awareness is similar to other long-standing debates in the perception literature (e.g., Holender, 1986) and in some other areas confronted with the general issue of unconscious processing (Schacter, Bowers, & Booker, 1989). In all these cases, one is faced with the necessity of formulating an explicit theory of awareness and of incorporating valid operational indices of awareness into the theory. Many authors (e.g., Allport, 1988; Marcel & Bisiach, 1988; Reingold & Merikle, 1988; Velmans, 1991) have suggested that this task is far from easy, and some of them (e.g., Reingold & Merikle, 1988) have explored the theoretical and methodological flaws underlying the widespread assumption that some given measure of performance may be taken as an *absolute index* of awareness (e.g., Jacoby, 1991; Reingold & Merikle, 1988).

Reingold and Merikle (1988) have argued that it may be impossible to consider some index of performance both as an exhaustive and as an exclusive measure of relevant conscious knowledge because researchers have no way of ascertaining that tasks are process pure and because researchers do not yet have a clear theoretical understanding of awareness. Hence, instead of requiring that absolute criteria of awareness be used, Reingold and Merikle suggest that a more productive strategy may be one that consists of comparing the sensitivity of various measures of the same relevant conscious information. They start by assuming that discrimination tasks in general may involve both relevant conscious information as well as some kind of unconscious sensitivity. Thus, no measure is likely to involve either kind of knowledge and processing in isolation. However, a given measure may be characterized as a direct or as an indirect test of the relevant knowledge depending on the relationship between the discrimination that it

requires and the definition of the task that participants are instructed to perform. For instance, recognition is a direct test of participants' ability to discriminate between old and new items when they are instructed to perform precisely this task. The old–new distinction, however, can also influence performance in other tasks: Merikle and Reingold (1991) have shown that judgments about the visual contrast of stimuli are affected by whether or not these stimuli had been presented before. In this case, the visual-contrast judgment task would be an indirect test of the old–new distinction.

Comparing similar direct and indirect measures of the same discrimination could thus be a way to determine whether performance is influenced by unconscious determinants. However, to do so, it is necessary to make assumptions about their relative sensitivity to conscious knowledge. Reingold and Merikle (1988) proposed that researchers make the following assumption: Direct tests of a given discrimination should not be less sensitive to conscious, task-relevant information than comparable indirect tests are. Thus, all other factors being equal, if participants are instructed to respond to information that is available to consciousness, then their use of this knowledge should not be worse than in cases where they are not directly required to use it. A straightforward implication of this assumption is that whenever an indirect measure shows greater absolute sensitivity to some relevant knowledge than a comparable direct measure does, one can conclude that this knowledge is not conscious, given that conscious knowledge alone could not explain the advantage observed in the indirect task.

Reingold and Merikle (1988; see also Merikle & Reingold, 1991) have applied this approach both to the demonstration of unconscious perception and to the study of problems of awareness in the field of implicit memory. The main goal of this article is to apply the same general logic to the study of unconscious learning effects in the context of serial-choice reaction-time (SRT) tasks. We start by discussing the reasons why sequence learning seems to be best suited to reveal such dissociations and report on an experiment designed specifically to address these issues.

Associations and Dissociations in Implicit Learning

The method advocated by Reingold and Merikle (1988) requires that two conditions be fulfilled. First, comparisons should be made between direct and indirect tests of the same discrimination only. Second, the two tests should be as comparable as possible with each other in terms of task context and demands. Does the existing body of evidence about implicit learning fulfill these conditions? A review of the literature about artificial grammar learning and system-control paradigms suggests that Reingold and Merikle's (1988) conditions are not fully respected in many cases. For instance, the critical comparisons used to assess awareness often involve two different but equally direct tests. In other cases, the two tests may not involve the same discrimination—a problem also expressed by Shanks and St. John's (1994) information criterion.

Consider, for instance, artificial grammar learning. Participants are typically instructed to first memorize a set of letter strings and then to make direct discriminations about the grammaticality of new strings that either conform or do not

conform to the rules used to generate the training material (Dienes, Broadbent, and Berry, 1991; Dulany et al., 1984; Mathews et al., 1989; Perruchet & Pacteau, 1990; Reber & Allen, 1978). For some authors, the fact that participants are found to perform above chance on the grammaticality test despite remaining unable to describe the rules of the grammar in free verbal reports constitutes good evidence that performance on the grammaticality test was implicit (e.g., Reber & Allen, 1978). For other authors, though, this kind of result is far from convincing because the free verbal reports might not be sensitive enough to reveal the extent of explicit knowledge possessed by participants. For this reason, many other direct tests have been proposed to explore the relationship between implicit and explicit knowledge in the context of artificial grammar learning. Thus, participants have been required to state the general features shared by the strings (e.g., Mathews et al., 1989), to mark the specific parts that make a string grammatical or nongrammatical (Dulany et al., 1984), or to assess the grammaticality of fragments of those strings (Dienes et al., 1991; Perruchet & Pacteau, 1990). In all these cases, a direct discrimination performance (i.e., grammaticality judgment) is thus compared with the results of other equally direct tests. Large associations have often been found, thus suggesting that discrimination performance is, in fact, based on knowledge that is available to conscious inspection. However, in the absence of clear theoretical assumptions about which kind of test is better suited to assess awareness, it is far from clear why one should interpret the association (or dissociation) results in precisely this way.

One crucial problem with this type of design is that if many of these tests of awareness differ from the measures of implicit performance both in terms of their task context and in terms of their relative sensitivity to conscious (and unconscious) information, they do not differ in terms of the direct versus indirect distinction. This is a problem because different task contexts could cue participants to retrieve different information regardless of whether this information is conscious or unconscious. Conversely, similar contexts should elicit responses based on the same information, again independently of whether this information is available to consciousness. Thus, it is not at all surprising, for instance, that participants' ability to make grammaticality judgments about complete letter strings is better related to their ability to make grammaticality judgments about fragments of letter strings than to their ability to verbalize the rules underlying the grammar (e.g., Dienes et al., 1991). Hence, it would seem that most of the associations or dissociations observed in this paradigm merely tend to reflect task similarities or differences rather than their relative sensitivity to conscious or unconscious processes.

Similar conclusions seem to hold for most of the *systems-control experiments* (Berry & Broadbent, 1984, 1987, 1988; Hayes & Broadbent, 1988; Marescaux & Karnas, 1991; Sanderson, 1989). In these experiments, participants are first required to learn to control a simulated system by setting the value of an input variable on each trial and by observing the output of the system. Their ability to reach and maintain a given target level on the output is typically considered as the main learning measure.

System-control performance has been described as implicit because participants are able to successfully control the system

despite remaining unable to answer explicit questions about the behavior of the system and because changes in one measure tend to be uncorrelated with changes in the other one. Again, however, the task contexts of the implicit and explicit measures differ considerably, and both measures are best characterized as direct. Thus, here also, there does not seem to be any good reason to infer that the differences exhibited by these two measures must be attributed to differences in their relative sensitivity to conscious or unconscious processes.

Consider, for instance, the following situation, which has often been used after the control task to assess participants' conscious knowledge (e.g., Berry & Broadbent, 1988; Marescaux & Karnas, 1991): A learning episode consisting of the previous state of the system and a hypothetical input is presented to participants, who are required to predict the next state adopted by the system as the outcome of this episode. Of course, complete explicit knowledge of the system's rules would provide all the information necessary to produce an accurate response to this question. Participants' prediction performance is typically far from perfect, however, and this is precisely why learning has been described as implicit in this situation. But there may also be alternative accounts that may explain the observed dissociation between control and prediction performance without appealing to differences in the conscious-unconscious dimension. For instance, one such alternative account is that some fragmentary knowledge of a reduced set of contingencies obtained during training with the control task is not transferred to the prediction task regardless of whether this knowledge is conscious or unconscious. In accordance with this claim, Marescaux and Karnas (1991) have observed that transfer between control and prediction performance is strongly dependent on whether the "questions" presented during both tasks refer specifically to the same contingencies (see also Dienes & Fahey, 1995). This pattern of results illustrates how the equivalence between the information required to perform any of the to-be-compared tasks acts as a preliminary condition that does not always hold in the context of system-control studies but that must be fulfilled before attributing a dissociation between any given measures to their differential sensitivity to conscious or unconscious effects.

Sequence Learning and the Generation Task

The third paradigm through which implicit learning has been studied is sequence learning. In these experiments, participants are typically placed in an SRT task in which the material has sequential structure: Some stimuli are more likely to appear than others in specific sequential contexts. Even though participants are kept unaware of this fact, their performance typically improves as training progresses. As has been shown repeatedly (see Cleeremans, 1993), this performance improvement reflects not only unspecific practice effects, but also an encoding of the sequential constraints present in the material. This developing sensitivity is clearly indirect in that the discriminations between predictable and unpredictable trials that it reflects are not directly required by the speeded-identification instructions given to participants. To assess awareness, SRT performance is typically compared with performance on a subsequent generation task, in which

participants have to perform the same discrimination between predictable and unpredictable sequence elements, but they perform these discriminations directly, that is, by explicitly predicting what they think the next stimulus will be. For instance, the *standard generation task* first proposed by Nissen and Bullemer (1987; Willingham, Nissen, & Bullemer, 1989) consisted of a modification of the SRT task in which participants were required to press the key corresponding to where they thought the next stimulus would appear instead of pressing the key corresponding to the current stimulus. Participants were told to pay more attention to accuracy of responding than to speed, but they were not explicitly told about the existence of a pattern. In addition, the stimulus remained present until a correct prediction was made so that several guesses could occur between any two trials of the generation task (e.g., Nissen & Bullemer, 1987).

It is far from clear, however, whether this standard generation task is the direct measure that is most comparable with the indirect measure provided by the reaction time (RT) task. This has led many authors to use alternative measures instead (e.g., Cleeremans & McClelland, 1991; Cohen, Ivry, & Keele, 1990; Perruchet & Amorim, 1992; Willingham, Greeley, & Bardone, 1993). In this section, we review the main problems that have been identified with the standard generation task and the alternative measures that have been proposed to address these problems.

Perruchet and Amorim (1992) describe three main problems with the standard generation task. First, the instructions do not mention the existence of a sequence and the fact that participants should make their predictions on the basis of the sequence knowledge obtained during the SRT task (i.e., it is not clear what they are actually trying to predict unless they specifically received such instructions). Second, the fact that participants are given feedback during generation allows them to learn about the material. This is a problem because it makes it hard to assess whether the generation task is measuring knowledge previously acquired during the RT task or knowledge acquired during generation itself. Third, the continuous guessing made mandatory by the task's design could interfere with memory of previous elements of the sequence and hence could make the generation task less sensitive to sequential knowledge than the SRT task is. To address these problems, Perruchet and Amorim have proposed two alternative measures: a version of the generation task that they called the *free generation task* and a recognition task. In free generation, participants are merely told to generate an entire sequence of trials that resembles the observed sequence, without receiving any feedback about how closely their generated sequence resembles the actual one. In the recognition test, participants are presented with a sequence of elements for a number of trials and are then asked to decide whether or not they had previously seen that particular sequence fragment on the basis of their experience with the SRT task material.

Although both of these measures may present some advantages compared with the standard generation task in terms of their global sensitivity to the acquired knowledge, it is far from clear whether they help in ensuring that they tap precisely the same information as used by participants during the SRT task. It is worth pointing out again that according to Reingold and Merikle (1988), the main issues when contrasting direct and

indirect measures of some knowledge are (a) that they involve precisely the same discrimination and (b) that they test this discrimination in contexts that are as comparable as possible. From this perspective, it is unclear whether free generation or recognition constitute more appropriate direct tests of the specific knowledge used during the SRT task than any other version of the generation task.

Indeed, consider for instance the discriminations that are required by recognition and by the SRT task. During the latter, it is assumed that knowledge of which sequence elements are likely and which are less likely to follow the contexts provided by previous elements is indirectly influencing the speed and accuracy of responses. Therefore, this discrimination between likely and unlikely sequence elements depends exclusively on the relative likelihood of the successors to these contexts. During recognition, however, participants are told to discriminate between old and new sequences as a whole, and their responses are thus influenced not only by the relative likelihood of the last element, but also by the perceived likelihood of all of the other transitions of the sequence. It seems obvious that this difference between the kind of information elicited by the two tasks tends to undermine arguments based on observed associations or dissociations between them. For this reason and despite its presumed advantages in terms of sensitivity, recognition may not fulfill the information criterion of Shanks and St. John (1994), just as it is not fulfilled by the other direct tests it was meant to improve on.

As for the free generation task, it also presents important problems that make it difficult to consider it as a good measure of the knowledge used by participants during the SRT task. First, as it was originally proposed by Perruchet and Amorim (1992), this task requires participants to freely generate a series of 100 consecutive keystrokes with the only constraint being that the generated sequence should resemble the sequence presented during the SRT task. Because of this, however, this task probably provides more information about what participants expect after their self-generated sequence fragments (which cannot be guaranteed to be part of the experimental set) than about what they expect after the experimental ones. Further, it is also possible that participants never come to generate specific sequence fragments, which therefore results in lost opportunities for comparing generation and SRT performance.

The main problem of the free generation task, however, is that the absence of any feedback—a feature that Perruchet and Amorim (1992) presented as a major advantage of this measure—may in fact be construed as a problem in that it greatly decreases this task's similarity with the SRT task in which indirect feedback is always present. In other words, whereas participants can always assess the quality of their anticipations during the SRT task simply by observing at which locations successive stimuli appear, they cannot do so in the free generation task. We surmise that this difference between the two tasks is likely to influence a number of factors involved in performance such as, for instance, participants' motivation to continue to try to anticipate the successive elements. Hence, to keep the direct task (i.e. generation) and the indirect task (i.e., SRT) as similar as possible to each other, it seems preferable to use what we could call a *continuous* version of the

generation task (see also Cleeremans & McClelland, 1991; Cohen et al., 1990) in which the next stimulus as prescribed by the sequential structure is presented regardless of participants' prediction responses, rather than using either the standard or the free generation tasks.

One concern with using a direct test that incorporates feedback information, however, is that the presence of feedback allows participants to learn, which makes it hard to separate the effects of previous knowledge from those of newly acquired knowledge. To address this problem, some authors (e.g., Cohen et al., 1990) have proposed to consider only the first few generation trials as the data with which to assess the knowledge acquired during the SRT task, but this technique presents reliability problems as the number of trials to be considered is necessarily very small. As a better way to cope with this difficult issue, we propose to adopt the sequence-learning paradigm developed by Cleeremans and McClelland (1991), in which the material to be learned was generated on the basis of a probabilistic finite-state grammar and in which random material was interspersed with structured material. This generation procedure results in stimulus material that is much more complex than typical stimulus material is and which Cleeremans and McClelland presented for considerably more trials. For two reasons, these features make it less likely for intrageneration learning to occur. First, because the material is probabilistic and follows complex generation rules, it is less likely that participants can learn about it in an explicit, goal-directed way within the few hundred trials typically used in other paradigms. Second, because participants in Cleeremans and McClelland's paradigm are exposed to several tens of thousands of RT trials, little is left to be learned about the material by the time the generation task is presented. Both of these features would tend to minimize intrageneration learning, thus enabling us to incorporate a greater number of generation trials in the relevant analyses. Consistently and in contrast with all other sequence-learning studies that we are aware of, Cleeremans and McClelland did not observe any learning over 465 trials of a generation task that followed 62,000 SRT trials.

In summary, our main goal in this article is to explore the relationship between SRT and continuous-generation performance in a probabilistic sequence-learning paradigm similar to the one developed by Cleeremans and McClelland (1991). We should point out that we assume neither that performance in the continuous generation task is exclusively determined by conscious contents nor that it should be considered as an exhaustive measure of knowledge available to consciousness. However, we believe that it can be taken as the direct measure of sequence learning that is most compatible with the typical indirect measure of this learning. Likewise, our approach does not require the assumption that SRT performance exclusively reflects unconscious learning but requires only that participants will not use explicit knowledge any more poorly when they are directly instructed to do so, as in the generation task, than when they are not, as in the SRT task. This assumption from Reingold and Merikle (1988) thus provides us with a conceptual framework that enables us to design experiments so as to assess whether participants show any evidence of expressing more knowledge when they are not directly in-

structed to use this knowledge than when they are. Any such effect would then have to be interpreted as evidence for unconscious learning.

There are additional reasons for us to think that this paradigm provides the optimal conditions in which to compare (direct) generation performance with (indirect) SRT performance. Indeed, the probabilistic nature of the stimulus material as well as the fact that it incorporates stimuli that either conform or do not conform to the sequential-generation rules allow us to consider large numbers of SRT and generation trials while minimizing the risk of having participants discover and memorize the sequence. It also provides for a great degree of continuous, intraparticipant control and allows learning to be assessed at different degrees of complexity and in great detail both for SRT and for generation performance.

In this article, we also examine whether inducing participants to use either an incidental or an intentional-learning orientation has any effects on performance in both tasks. Manipulation of the instruction set given to participants has obvious relevance to the issue and constitutes another angle of attack on the general problem of assessing the relationship between implicit and explicit performance. Indeed, theories that assume that learning is essentially implicit in this situation would predict that there should be no differences between incidental and intentional participants. Finally, we also assess how well the SRN model of sequence learning (Cleeremans & McClelland, 1991) can account for the data.

Method

Participants were exposed to a six-choice RT task. The experiment consisted of 20 sessions. Each session consisted of 20 blocks of 155 trials each, for a total of 62,000 trials over the entire experiment. On each trial, a stimulus could appear at one of six positions arranged horizontally on a computer screen. Participants were to press as quickly and as accurately as possible the key corresponding to the current location of the stimulus. As in Cleeremans and McClelland's (1991) experiments, the sequential structure of the material was manipulated by generating the sequence on the basis of a noisy finite-state grammar. This fact was revealed to participants only in the *intentional condition*. These participants were also told that trying to discover the sequence would help them perform better. By contrast, participants in the *incidental condition* were kept unaware of the presence of regularities in the material. All participants were exposed to 465 trials of a generation task after completion of the RT task. On each trial, participants were asked to predict the location at which the next stimulus would appear.

Participants

Twelve participants, all students of introductory courses in psychology at the University of Santiago de Compostela in Spain, participated in the experiment. Six participants were randomly assigned to each orientation condition (incidental vs. intentional). Participants were paid about \$50 for participating in the experiment and could earn an additional \$34 to \$62 depending on performance (see below).

Apparatus and Display

The experiment was run on an IBM PS/2 computer. The display consisted of six dots arranged in a horizontal line on the computer's

screen and separated by intervals of 3 cm. At a viewing distance of 57 cm, the distance between any two dots subtended a visual angle of 3.0°. Each screen position corresponded to a key on the computer's keyboard. The spatial configuration of the keys was entirely compatible with the screen positions (i.e., the key furthest to the left corresponded to the screen position furthest to the left etc.). The stimulus was a small white X 0.35 cm high that appeared on a black screen and was centered 1 cm above one of the six dots. The timer was started at the onset of the stimulus and was stopped by the participant's response. The response-stimulus interval was 120 ms.

Tasks

The experiment consisted of two tasks presented successively: an SRT task and a generation task. The SRT task was carried out during 20 training sessions, each composed of 20 blocks of 155 trials. After the last session, participants performed a generation task during which they were required on each trial to predict the location of the next stimulus by pressing the corresponding key. The generation task consisted of 465 trials over 3 blocks of 155 trials each. The generation task material followed exactly the same sequence as presented to each participant during 3 blocks of the penultimate SRT session. Therefore, both the specific sequence and the environment (i.e., the response-stimulus interval, the characteristics of the screen, and the keyboard layout) were kept very similar between the SRT and generation tasks. In contrast to the procedure used by Cleeremans and McClelland (1991), neither the SRT task nor the generation task provided explicit feedback. During generation, however, participants could come to know about their performance by comparing their predictions with the actual stimuli, just as they could become aware of the accuracy of their responding during the SRT task.

Procedure

Each participant performed the task at a rate of two sessions per day. The generation task was presented immediately after the last session of the SRT task. Participants were randomly assigned to one of the two experimental conditions. All participants were told that the goal of the experiment was to analyze the effects of extended practice on performance in a simple task. They were also told about the structure of the experiment and about how to place their fingers on the keyboard. Both accuracy and speed were emphasized as means of increasing earnings. Participants in the intentional condition received additional information about the stimulus material. They were told that a complex set of rules determined where each stimulus could appear, depending on the previous sequence. They were urged to try to discover these rules, not only as a way to improve their performance in the main SRT task, but also because they would later be asked to predict successive stimuli in a subsequent generation task. To increase their motivation to engage in explicit, code-breaking behavior, participants in the intentional condition were given specific instructions to search for the rules during each rest break. Further, they were told that their prediction accuracy during the generation task would be used to multiply the total value of their earnings during the SRT task.

After receiving the instructions appropriate to their condition, all participants were given three practice blocks of 15 random trials each, during which they were given a last chance to ask questions and to interact with the experimenter about the task setting. The first experimental session was then initiated. To eliminate early variability in performance, each block began with a set of 5 random and unrecorded trials. Participants were then exposed to 150 structured trials. Successive blocks were separated by a rest break, the length of which participants could control. After an entire session had been completed, participants were presented with feedback about their

performance. The computer displayed information about accuracy and mean RT obtained during the last session and also indicated how much bonus money had been earned so far in the experiment. The following formula was used to determine earnings for each session:

$$E = (ACC - 84) \times (880 - RT) \times .078,$$

where E represents earnings and ACC represents accuracy. The parameters were empirically selected both to set a lower limit on performance and to maintain the level of earnings (in pesetas) within the desired limits. After the final experimental session, all participants received instructions about the generation task. They were told that the sequence of stimuli had followed a pattern during the SRT task and that they would now have to try to predict where the next stimulus would appear instead of responding to the current stimulus. The instructions emphasized the fact that the sequential structure of the material would be the same in the generation task as it had been in the SRT task. Participants were also told that RTs were no longer recorded and that they could get feedback about their performance by comparing the location that they had indicated by their keypress with the actual location at which the next stimulus appeared. Finally, the importance of being as accurate as possible was again emphasized by pointing out that accuracy during generation would be used to compute a number that would be used to multiply earnings during the SRT task. This parameter was set to be 1.0 plus the proportion by which accuracy exceeded the chance level during generation.

Stimulus Generation

Stimuli were generated on the basis of a noisy finite-state grammar similar to the one used by Cleeremans and McClelland (1991), with a small proportion of random stimuli (15%) interspersed with structured ones. As shown in Figure 1, there are two main differences between the grammar used in this study and the one used by Cleeremans and McClelland. First, to further control for both the short-term priming effects and the verbalizable knowledge that may result from the presence of salient patterns in the stimulus material, the loops on Nodes 2 and 4 were eliminated. Second, the grammar was designed so as to ensure that some contexts remained ambiguous (i.e., may have been pointing to more than a single node) up to (and only up to) a given length. Stimulus generation proceeded in three phases. First, a sequence of 60,000 grammatical letters was generated on the basis of the grammar by selecting an arc coming out of the current node and recording the corresponding label on each trial. The current node was set to be Node 0 on the first trial of each block and was updated on each trial to be the node pointed to by the selected arc. Second, there was a 15% chance of substituting a randomly selected label for the recorded one (identity substitutions were not allowed). Third, the label was used to determine the screen position at which the stimulus would appear by following a 6 × 6 Latin square design, so that each label corresponded to each screen position for exactly 1 of the 6 participants in each condition.¹ Finally, a set of 5 completely random and unrecorded trials were added at the beginning of each block of 150 trials to control for initial response variability.

¹ The mapping between labels and screen positions was randomly determined for the first participant and was then modified for subsequent participants by shifting the sequence of screen positions one step to the right for each participant. Thus, for Participant 1 of each condition and with the consecutive screen positions labeled from left to right with the numbers 1 to 6, the consecutive labels A through F corresponded to the following series of locations: 1, 4, 5, 2, 6, and 3, respectively. For Participant 2, the same labels corresponded to the series 2, 5, 6, 3, 1, and 4 and so on for the other participants.

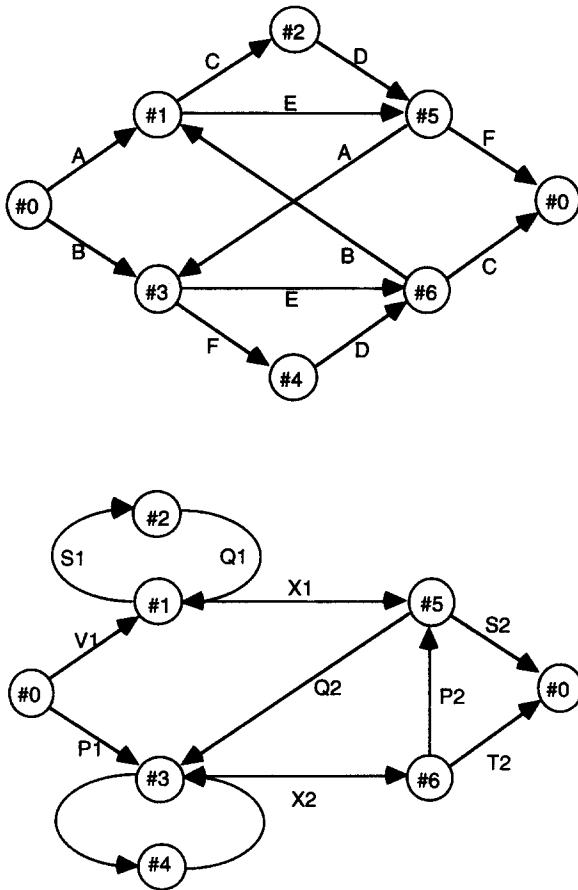


Figure 1. The finite-state grammar used to generate the stimulus material (on the top) compared with the grammar used by Cleeremans and McClelland (1991; on the bottom). Note that in both grammars the first and last nodes are the same.

Results

Participants were exposed to 20 sessions of an SRT task and were subsequently asked to try to predict each successive event in a follow-up generation task. We first present the SRT data.

Learning in the SRT Task

To assess whether participants learned about the sequential structure contained in the stimulus material, we compared their responses to a given stimulus when it had appeared after a context that it could or could not legally follow according to the grammar. Indeed, if participants become progressively more sensitive to the constraints expressed by the grammar over training, one would expect to observe a progressive facilitation for grammatical stimuli compared with nongrammatical stimuli.

Before we describe the analyses in detail, it is worth describing the relevant properties of the grammar. Consider the grammar illustrated in Figure 1 from the point of view of a system that is attempting to reduce the uncertainty of the next stimulus on the basis of the temporal context, that is, on the basis of the few previous trials.

A first point is that even if the system knew the grammar (i.e., had some internal representation similar to Figure 1), there would still be some uncertainty about which label would occur next because arcs are selected at random among the possible arcs emanating from a particular node during stimulus generation. Most nodes (i.e., all but Nodes 2 and 4) bear two outgoing arcs. Hence, the best any system can do to reduce the uncertainty associated with the next stimulus is to identify the current node. However, the only information available to the system is the sequence of previous labels. One can therefore ask how many previous labels are necessary to maximally reduce the uncertainty associated with the next element. In this grammar, as many as three previous labels are necessary to do so, but in most cases, two labels are sufficient. One label is never sufficient to determine the current node because each label may occur twice on different arcs. For instance, the label *A* appears on arcs pointing to both Nodes 1 and 3. When preceded by the label *D*, however, *A* can point only to Node 3. Some second-order sequences are still ambiguous, however. For instance, the context *AE-* can lead to Node 5 or to Node 6, but *CAE-* and *DAE-* resolve the uncertainty and point only to Node 5 or 6, respectively.

Consider now how the substitution procedure interacts with this structure. In 15% of the cases, a random label is substituted for the label prescribed by the grammar. Hence, the material incorporates cases where the current stimulus is nongrammatical in the context of the previous stimuli. To assess learning of the constraints embodied by the grammar, we can compare RTs for such nongrammatical stimuli with RTs for the same stimuli when their occurrence is consistent with the structure of the grammar. For instance, to determine whether participants are sensitive to the identity of Sequence Element *t-1* when responding to Element *t*, we can compare the response to Element *t* in cases where it is preceded by an element that it can or cannot legally follow (e.g., the response to *A* when *A* appears in the paths *D-A* vs. *B-A*). We would expect the RT to *A* to be faster in cases where its occurrence is consistent with the context than in cases where it is not. Further, we can conduct these comparisons for contexts of varying lengths. In the following section, we present analyses for contexts of Lengths 1 to 3 (henceforth referred to as L1, L2, and L3 contexts). For each length, three constraints were used to define which paths were entered into the analysis.

First, paths involving immediate and alternating repetitions (e.g., *A-A*, *AC-A*) were eliminated from all comparisons to avoid the short-term priming effects observed in Cleeremans and McClelland (1991).

Second, only grammatical contexts were considered, with successors selected in such a way that their grammaticality was exclusively dependent on the first element of the context.² Consider, for instance, the paths *BC-A* and *DC-A*. The

² In the remainder of the article, we use the word *path* to refer to any sequence of elements. Paths consist of a *context* consisting of all the elements of the path except for the last one. We refer to this last element as the *successor element*.

Table 1
List of Grammatical (G) and Nongrammatical (NG) Paths Considered for Each Context Length (L1, L2, and L3)

L1		L2		L3	
G	NG	G	NG	G	NG
		DC-A EC-A	BC-A		
C-A D-A E-A F-A	B-A	CD-A DF-A EF-A	FD-A BF-A	DBE-A	CBE-A FBE-A
		DC-B EC-B	AC-B		
C-B D-B E-B F-B	A-B	FD-B DF-B EF-B	CD-B AF-B	DAE-B	CAE-B FAE-B
A-C B-C D-C E-C	F-C	FA-C DB-C EB-C	DA-C EA-C FB-C	DAE-C CBE-C FBE-C	CAE-C FAE-C DBE-C
		AC-D BC-D	EC-D		
C-D F-D	A-D B-D E-D	AF-D BF-D	EF-D		
A-E B-E	C-E D-E F-E				
A-F B-F D-F E-F	C-F	DA-F EA-F CB-F	CA-F DB-F EB-F	CAE-F FAE-F DBE-F	DAE-F CBE-F FBE-F

contexts *BC-* and *DC-* are both legal sequences. Further, *A* is a legal successor of both the contexts *C-* and *DC-* but is specifically illegal after *BC-*.

Third, a given path was entered into the analyses only if there were matching grammatical and nongrammatical instances (i.e., differing only in their initial element) that conformed to the first two constraints. For instance, the grammatical path *CD-F* was not included in the analyses because the only nongrammatical instance with which it could be compared was *FD-F*, a path that is contaminated by an alternation pattern.

These successive eliminations of candidate paths resulted in sets of 30 (L1), 34 (L2), and 18 (L3) paths for further analysis. These paths are shown in Table 1. The effect of learning on performance on the SRT task was defined as a progressive increase in speed, accuracy, or both in response to predictable trials that could not be attributable to a trade-off between both indices. Hence, to consider both effects simultaneously, we adopted a multivariate approach using both RTs on correct responses and percentage of hits (i.e., accuracy) as the dependent variables (Pachella, 1974). Orientation was included as a grouping variable, and both grammaticality and practice were used as within variables. Wilks lambda was the multivariate

test selected for the multivariate analyses of variance (MANOVAs). The reported *F* represents the Rao approach to the *F* distribution of the Wilks lambda ratio (see Bray & Maxwell, 1985).

Because the frequency of each path is lower for longer paths, we aggregated several sessions of training within each level of practice for the higher complexity conditions. Hence, we considered 20 levels of practice for L1 paths, 10 for L2 paths, and only 5 for L3 paths. Figure 2 shows average RT and accuracy performance for each of the 20 experimental sessions plotted separately for grammatical and nongrammatical trials after L1 contexts and for incidental and intentional learning conditions. Figure 2 shows that both practice and grammaticality have strong effects on performance: Participants improved as training progressed, and the difference between grammatical and nongrammatical trials tended to increase with practice. These effects were present and similar in both the incidental

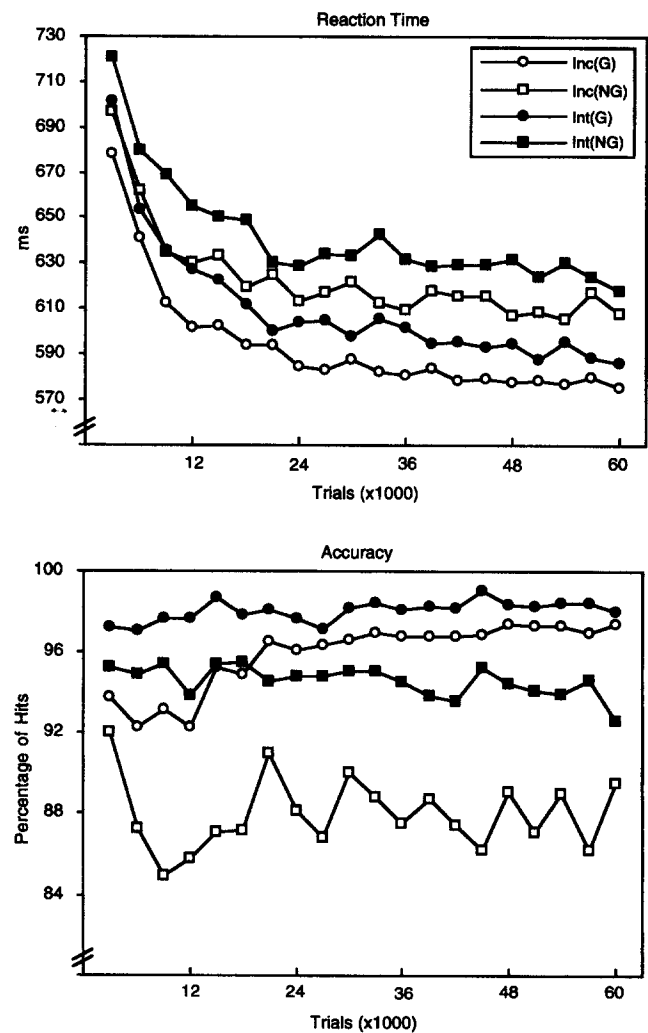


Figure 2. Reaction time and accuracy performance for each of the 20 experimental sessions plotted separately for grammatical (G) and nongrammatical (NG) successors after Length-1 contexts for incidental (Inc) and intentional (Int) learning conditions.

and the intentional conditions, but intentional participants seemed to respond in a slightly more cautious way, as expressed by their more accurate but slower responses.

A mixed MANOVA conducted with orientation (2 levels), practice (20 levels), and grammaticality (2 levels) variables on both RT and accuracy confirmed these observations. It revealed significant effects of practice, $F(38, 378) = 12.05, p < .0001$, and of grammaticality, $F(2, 9) = 108.08, p < .0001$, as well as a significant Practice \times Grammaticality interaction, $F(38, 378) = 2.30, p < .0001$. The orientation effect did not reach significance within this multivariate approach, $F(2, 9) = 3.25, p < .09$, although the univariate analyses conducted on each one of these variables showed that intentional participants responded more accurately, $F(1, 10) = 5.25, MSE = 419.84, p < .05$, but produced slower responses, $F(1, 10) = 5.19, MSE = 7017.70, p < .05$, than incidental participants. None of the interactions that included orientation reached significant levels in the MANOVA.

Finally, the univariate, follow-up analyses conducted on each of the dependent variables to explore their separate contributions to the general learning effect confirmed both effects, as inferred from the significance of the Practice \times Grammaticality interaction both for RT, $F(19, 190) = 2.00, MSE = 57.41, p < .01$, and for accuracy, $F(19, 190) = 3.23, MSE = 3.38, p < .01$. Similar results were obtained for L2 contexts as shown in Figure 3. Again, practice, $F(18, 178) = 15.26, p < .0001$; grammaticality, $F(2, 9) = 54.07, p < .0001$; and the Practice \times Grammaticality interaction, $F(18, 178) = 2.23, p < .01$, were significant in the $2 \times 10 \times 2$ MANOVA, confirming (a) that a progressive discrimination between grammatical and nongrammatical trials is taking place with practice at this level and (b) that this effect amounted to increased speed and accuracy for predictable trials. Orientation also reached significance for this level, $F(2, 9) = 4.42, p < .05$, thus indicating that incidental participants responded faster but less accurately than intentional participants. None of the remaining interactions including orientation reached significance. Again, there were no learning differences between the two conditions, as revealed by the absence of a triple Practice \times Grammaticality \times Orientation interaction ($F < 1, p > .50$).

The univariate analyses of the contribution of the two dependent variables to the observed learning effect confirmed the significance of the Practice \times Grammaticality interaction both for the RT, $F(9, 90) = 2.46, MSE = 54.20, p = .01$, and for the accuracy measures, $F(9, 90) = 2.06, MSE = 1.60, p < .05$. Figure 4 shows the results obtained for L3 contexts. As can be seen in this figure, there was again a progressive and generalized improvement in the SRT performance with practice, $F(8, 78) = 10.45, p < .0001$, as well as a nonsignificant trend toward a difference between intentional and incidental conditions regarding participants' response criteria, $F(2, 9) = 3.72, p < .07$. However, neither grammaticality, $F(2, 9) = 1.17, p > .30$, nor its interaction with practice ($F < 1, p > .40$) approached significance, thus suggesting that participants did not discriminate between successors of L3 contexts. The triple interaction involving grammaticality, practice, and orientation also did not reveal any significant effect ($F < 1, p > .90$), which provides a further indication that this lack of learning applies to both the intentional and the incidental conditions.

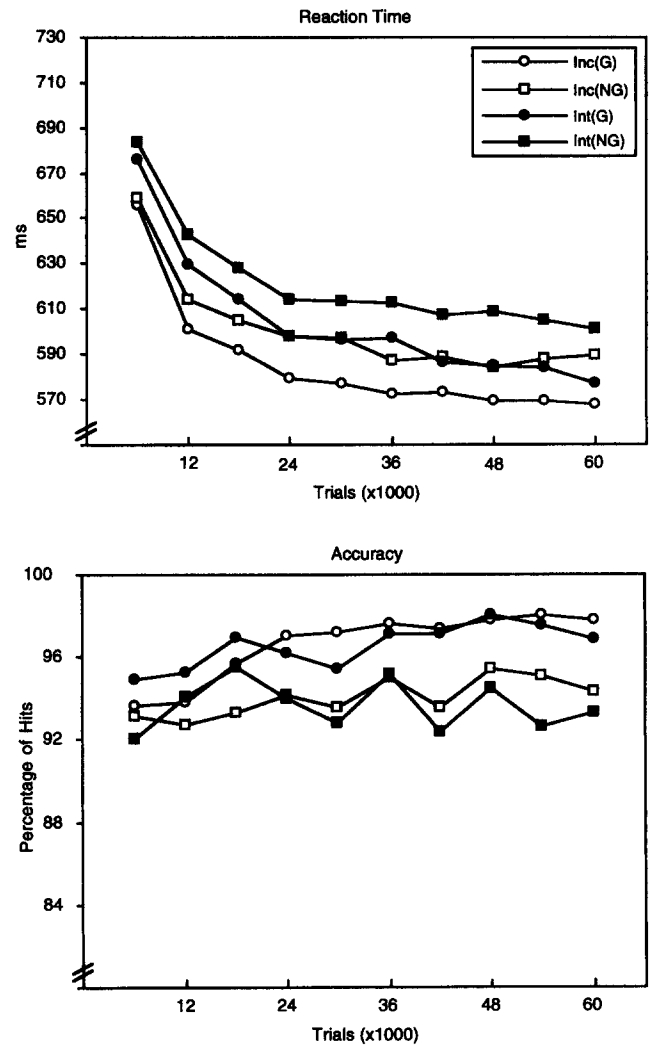


Figure 3. Reaction time and accuracy performance for each of the 10 blocks of two experimental sessions plotted separately for grammatical (G) and nongrammatical (NG) successors after Length-2 contexts for incidental (Inc) and intentional (Int) learning conditions.

In summary, our analyses suggest (a) that participants appear to be able to learn about the structure of the material but not about the constraints set by more than two previous trials at most, (b) that this sensitivity is expressed through increased speed and accuracy of responding to grammatical trials compared with nongrammatical trials, and (c) that orientation did not significantly influence learning even though intention to learn tends to affect the speed-accuracy trade-off.

Generation Task Performance

To assess whether there was any evidence for learning within the generation task, we first conducted a conventional analysis of hits percentage on each of the three successive blocks and for each orientation condition. As can be seen in Table 2, these results show that participants had knowledge that enabled them to perform better than could be expected on

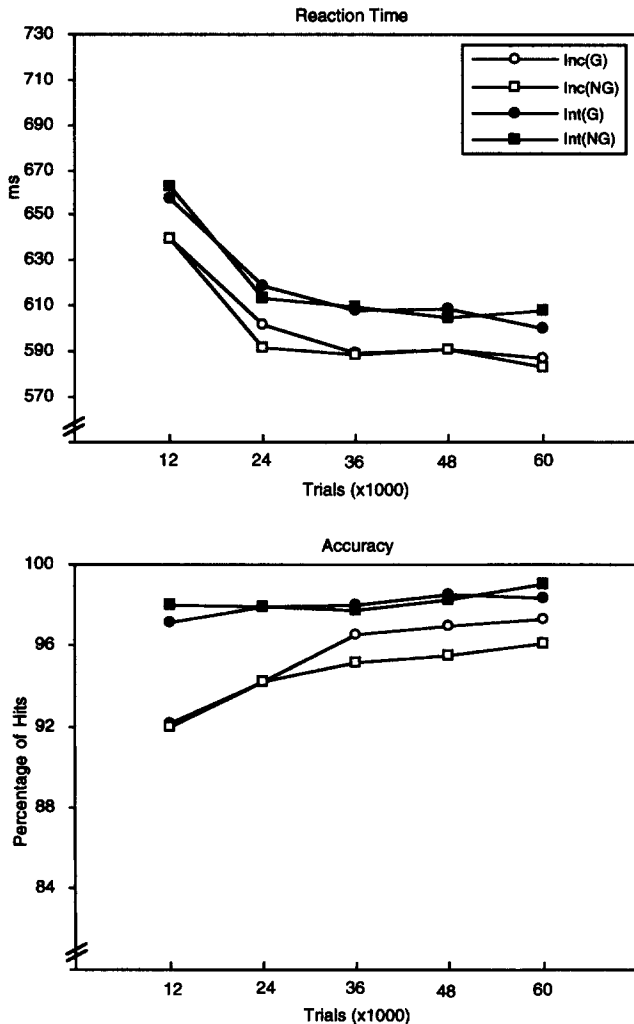


Figure 4. Reaction time and accuracy performance for each of the five blocks of four experimental sessions plotted separately for grammatical (G) and nongrammatical (NG) successors after Length-3 contexts for incidental (Inc) and intentional (Int) learning conditions.

the basis of a completely random prediction strategy (chance level around 17%) even during the first block, $t(11) = 10.60, p < .0001$, and even enabled them to perform somewhat better than could be expected for a participant who would never predict immediate repetitions (chance level: 20%), $t(11) = 6.90, p < .0001$. However, contrary to the usual trend observed in deterministic tasks (e.g., Cohen et al., 1990; Frensch, Buchner, & Lin, 1994; Frensch & Miner, 1994; Nissen & Bullemer, 1987), but consistent with the results reported in other experiments with probabilistic material (Cleeremans & McClelland, 1991, Experiment 2), an analysis of variance (ANOVA) with practice (3 levels) and orientation (2 levels) as independent variables did not reveal any evidence of learning within the generation task, $F(2, 20) = 1.79, MSE = 20.90, p > .15$. Orientation also did not reach significance in this analysis, $F(1, 20) = 0.25, MSE = 6.93, p > .50$, and the absence of practice effects was present both in the intentional and in the incidental groups, $F(2, 20) = 0.15, MSE = 1.70, p > .80$.

These results further reinforce the notion that the generation task as used here essentially taps knowledge acquired during the SRT task and does not provide new opportunities for learning. However, it may still be possible that all the sequence knowledge observed within the generation task could have been acquired by participants on the basis of their experience with the first generation block (i.e., all of the intrageneration learning would have taken place during the first block of the generation task). If this were the case, the generation task would actually overestimate the knowledge acquired during the SRT task, and it is therefore important to rule out this explanation. To control for this possibility, we asked 6 new participants to perform the generation task without any prior experience with the SRT task. The results of this control group are shown in Table 2. The data suggest that participants who have not been given prior training with the SRT task do not perform over chance during the first block of 155 generation trials, $t(5) = 0.21, p > .40$. Although it would seem that participants in the control group quickly learn about the stimulus material, the effect of practice did not reach significance, $F(2, 19) = 1.53, MSE = 33.15, p > .25$. Presumably then, participants in the control group could learn about simple features of the stimulus material, such as the fact that direct repetitions almost never occur. This single factor may account for control participants' increase in performance between the first and the second generation blocks, $t(5) = 2.01, p = .05$. Experimental participants, by contrast, had already acquired this knowledge during training with the SRT task and hence did not show any further learning during generation.

A more important analysis consists of assessing whether the average generation probability of any given sequence element differs depending on whether this element is or is not consistent with the context set by previous elements. To this effect, we analyzed the generation data in a way that is exactly analogous to the way we analyzed the SRT data; that is, we computed the conditional probability that participants would generate each successor to each context of the same paths used for the SRT analysis (see Table 1) and averaged these data over grammatical and nongrammatical cases for the different possible context lengths. Hence, this analysis provides us with a measure that reflects the extent to which grammatical and nongrammatical successors tended to be generated by participants after each of the selected contexts. Note that this measure differs from the more conventional measures of generation accuracy (i.e., hits percentage) in an important way. Indeed, accuracy measures tend to underestimate learning when applied to a probabilistic structure in that even perfect

Table 2
Percentage of Hits During Generation Averaged Over Blocks and Conditions

Condition	Block		
	1st	2nd	3rd
Incidental	25.85	23.60	24.67
Intentional	26.65	23.77	26.33
Control	16.90	20.53	21.30

knowledge of the rule system could still result in poor prediction performance, precisely because several grammatical successors are possible after each context and are selected at random by the generation procedure. Hence, participants who know which elements are grammatical in a given context may still not generate precisely the specific element that would appear next. Our measure, by contrast, considers any generated sequence element that is grammatical after a given context as a hit.

Table 3 shows the average probability of grammatical versus nongrammatical generation for each orientation group and for each context length, aggregated over the selected paths during all three blocks of the generation task. Independent ANOVAs conducted on the data corresponding to each context length confirmed the existence of significant effects of grammaticality both for L1 contexts, $F(1, 10) = 64.28$, $MSE = 23.50$, $p < .0001$, and for L2 contexts, $F(1, 10) = 11.74$, $MSE = 27.90$, $p < .01$, but not for L3 contexts, $F(1, 10) = 1.61$, $MSE = 24.80$, $p > .20$. None of the effects or interactions involving orientation approached significance.

In summary, the results obtained in both the SRT and generation tasks revealed that participants appeared to be able to anticipate or to predict the location of stimuli on the basis of information about the two previous sequence elements. This sensitivity to sequential constraints was limited however, in that participants' responses did not appear to be based on the constraints set by sequence elements that had appeared three trials before the current one. Both intentional and incidental participants produced nearly identical response distributions, even though the instructions given to each group appeared to have been efficient in promoting a different response attitude in each group. If this last result is consistent with the notion that learning is essentially implicit in this task, other results suggest otherwise. Indeed, we found that participants exhibited sensitivity to equally complex contingencies in both the SRT and the generation measures, despite the fact that the number of observations was disproportionately lower in the generation task as compared with the SRT one. To further evaluate this equivalence on a more comparable number of observations, we conducted another set of analyses of SRT performance with only the three blocks of trials that were also presented to participants during generation (and that had been arranged to include exactly the same trials in both tasks). As illustrated in Table 4 and confirmed by the MANOVAs, the grammaticality effect was again significant both for L1, $F(2, 9) = 10.89$, $p < .01$, and for L2 contexts, $F(2, 9) = 5.99$, $p <$

Table 3
Average Probability of Generation of Each Successor to Each Context

Successors	Condition					
	Incidental			Intentional		
	L1	L2	L3	L1	L2	L3
G	.246	.311	.160	.246	.306	.220
NG	.082	.250	.205	.092	.219	.227

Note. L1, L2, and L3 refer to Lengths 1, 2, and 3, respectively; G = grammatical; NG = nongrammatical.

Table 4
Means of Reaction Times and Percentages of Hits for Successors of Each of the Paths of L1, L2, and L3 Obtained During Three Blocks of Session 19

Successors	Condition					
	Incidental			Intentional		
	L1	L2	L3	L1	L2	L3
RT						
G	583.9	572.6	574.4	589.1	580.0	587.6
NG	614.2	609.4	612.1	613.5	602.6	594.9
ACC						
G	97.0	97.6	95.5	98.2	98.4	98.9
NG	85.2	94.7	100.0	96.3	95.5	100.0

Note. L1, L2, and L3 refer to Lengths 1, 2, and 3, respectively; RT = reaction time; G = grammatical; NG = nongrammatical; ACC = accuracy.

.05, but not for L3 contexts, $F(2, 9) = 3.44$, $p = .077$. Actually, both L1 and L2 responses tended to show a pattern of learning in which both speed and accuracy improved for grammatical but not for nongrammatical trials, whereas the slight RT advantage observed for grammatical L3 items was achieved only at the expense of an opposite difference in accuracy. Orientation again did not reach significance for all three of the analyses, and the same absence of effect was also obtained for the interaction between grammaticality and orientation. Thus, the general pattern of results described so far seems to reinforce the notion that SRT and generation performance reflect the same amount of learning about the sequential structure of the stimulus material.

These results could be taken, along with those reported by Perruchet and Amorim (1992), as providing evidence against the existence of a dissociation between direct and indirect measures in sequence-learning paradigms. However, we think that neither these data nor those reported by Perruchet and Amorim can be safely taken as strong evidence against the existence of an implicit learning process. In the following section, we describe more detailed analyses that support this claim.

Global Associations and Specific Dissociations

Consider the results obtained in both the SRT and generation tasks. Both measures show that participants are using information about the previous two elements to anticipate or predict the identity of the next one. However, the analyses we have presented so far have not enabled us to assess whether the knowledge used during the SRT task is similar to or different from the knowledge used during generation. Typically in this field, associations are taken to support the hypothesis that a single mechanism is sufficient to account for all the learning effects, whereas dissociations are interpreted as indicating the existence of some independent process of implicit learning. However, although this theorizing may seem to be straightforward, it is important to note that any less-than-perfect association between two comparable measures could amount to a dissociation if it is not attributable to random

variance (cf. Perruchet & Gallego, 1993). Thus, a fine-grained analysis of the knowledge revealed through each of the two measures is necessary to test whether there are significant differences between their respective sensitivities to the same contingencies or, alternatively, whether all the differences observed between the two measures are the result of random variation.

Perruchet and Amorim (1992) undertook such an analysis by obtaining the value of each of the to-be-compared measures for each part of their sequence and by computing correlations between the resulting distributions as an index of the degree to which they are related. Our analysis goes beyond theirs by explicitly testing whether all the learning effects observed through the SRT performance could be derived from the knowledge reflected in generation performance or, on the contrary, whether some sequence knowledge is exclusively reflected through the indirect measure, that is, the SRT task.

To test these possibilities, we proceeded in the following way. First, we computed the distribution of the generation probabilities for the last element of each of the selected paths within each context length as well as the corresponding distributions of RT and accuracy measures (aggregated over five groups of four consecutive sessions). These analyses were conducted separately for each participant. Second, we compared these learning distributions with each other by computing their correlation. Third, we also compared each distribution with a third one, namely the observed distribution of the conditional probabilities of the last element of the same selected paths. These conditional probabilities (CPs) were computed on the basis of the actual series of 60,000 structured trials presented to participants (all of whom were exposed to the same sequence, albeit with different mappings between labels and screen locations) simply by recording the probability of observing each successor after each of the contexts of all the paths listed in Table 1. The resulting distribution represents the optimal information about the sequential constraints embedded in the stimulus material that any system could obtain on the basis of training with the stimulus set. Indeed, any learning that is specific to the sequential nature of the material should be reflected in the data as a correspondence with the distribution of conditional probabilities at different context lengths. Hence we can operationalize learning as the extent to which either SRT or generation response distributions are associated with the sets of CPs (see Figure 5).

Further and of most importance, we can now assess whether direct and indirect measures of learning are similarly sensitive to the sequential structure of the stimulus material by computing the partial correlations between the distributions of the indirect measures and the CPs after their common variance with the direct measures has been removed. The square of this partial correlation would indicate the proportion of variance of the indirect measures that is exclusively explained by the variance in the CPs when the effect of the variance of the direct measures has been partialled out (i.e., the intersection between the corresponding diagrams that is not shared with the one representing the variance of direct performance). Any remaining association between indirect measures and CPs would thus reflect knowledge about the sequential nature of

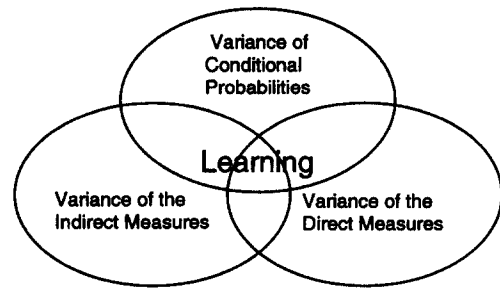


Figure 5. Venn diagram of the variances of the conditional probabilities and of the direct and indirect measures of performance. Sequence learning is operationalized as the extent to which the variances of either of the distributions of direct and indirect measures are associated with the variance of the conditional probabilities.

the material that cannot be accounted for by the knowledge expressed through the direct measures.

We proceeded in the following way. First, we transformed correlations and partial correlations for each participant into Z scores before computing averages or conducting MANOVAs. Second, as orientation did not influence learning, these Z scores were averaged among all 12 participants. Finally, these averages were then transformed back into correlation scores. These average correlations are summarized in the Appendixes. Figure 6 represents the average partial correlations between CPs and the results obtained on each of the two indirect measures (RT and accuracy) for each level of practice, controlling for the learning manifested through the generation task. The figure clearly shows that there are significant and significantly growing partial correlations between both indirect measures and the distribution of CPs (controlling for the effects observed in generation), at least for the contexts of L1 and L2. Three 2×5 (Orientation \times Practice) MANOVAs conducted separately for each context length with the Z -transformed scores of both RT and accuracy as dependent variables confirmed this impression by yielding significant practice effects for L1 contexts, $F(8, 78) = 9.78, p < .0001$, and L2 contexts, $F(8, 78) = 4.19, p < .0005$, but not for L3 contexts, $F(8, 78) = 1.22, p > .30$. None of the effects or interactions involving orientation approached significance.

To illustrate these data in a more intuitive way, we constructed Venn diagrams that represent the proportion of variance (i.e., squared correlation coefficients) that is shared by or is specific to each of the three data distributions corresponding to the CPs and to the direct and indirect measures (the latter data were averaged over the last four sessions of the SRT task). These data are represented separately for L1 and for L2 contexts in Figure 7. Visual inspection of the diagram corresponding to L1 contexts shows that if there seems to be some knowledge about the statistical structure of the material that is exclusively expressed through the generation task (accounting for around 5% of the variance in CPs), the percentage of variance explained exclusively through the indirect measures is clearly higher, amounting to over 20%.

For L2 contexts (see Figure 7), the variance of the CPs that is expressed exclusively through the indirect measures still amounts to over 10% of its total variance, whereas the

knowledge expressed exclusively through generation responses amounts to less than 0.2%. Overall, if these results suggest that generation and SRT performance are indeed related in that both measures reflect sensitivity to sequential constraints of the same length and in that both measures are even moderately correlated, they also make it clear that generation and SRT performance may be sensitive to partially different contents. Indeed, as the partial-correlations data indicate, some grammatical knowledge is expressed exclusively through the indirect measures, resulting in increased speed and accuracy that is proportional to the conditional probability of the stimuli.

More generally, our results clearly illustrate that tasks are not process pure and that the relations between tasks and knowledge are complex ones. However, even if some knowledge about the statistical structure of the material tends to be exclusively expressed through our direct measure, the main point we wish to emphasize is that some of the knowledge

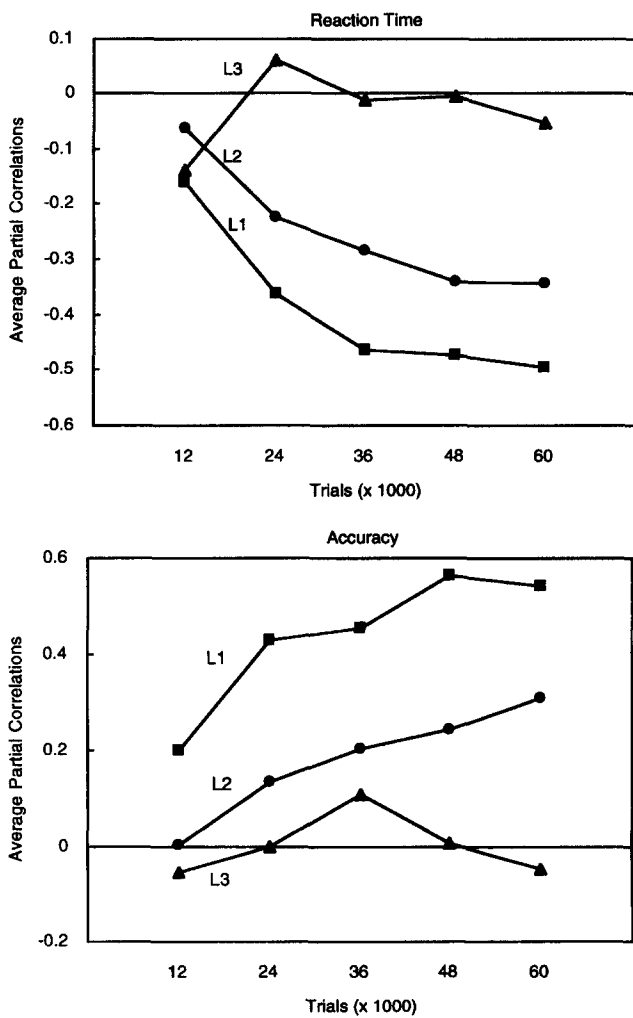


Figure 6. Average partial correlations between the distribution of the conditional probabilities and either reaction time or accuracy for each level of practice and each context length (L1, L2, and L3), controlling for the learning expressed through the generation task.

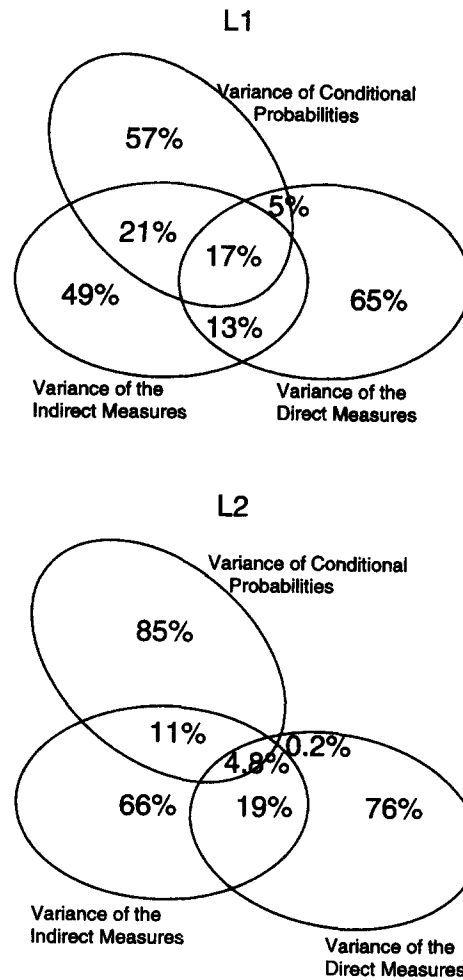


Figure 7. Venn diagrams of the proportions of the variance that are shared among the distributions of the conditional probabilities and of the direct and indirect measures of performance for contexts of Length 1 (L1) and Length 2 (L2).

acquired during the SRT task appears to be exclusively expressed in the context of this SRT task. Hence, from the theoretical perspective adopted in this article and if we accept Reingold and Merikle's (1988) assumption that conscious knowledge should not be expressed more poorly through a direct measure than through a comparable indirect measure, then we cannot help but conclude that the knowledge that is exclusively expressed in the context of the SRT task is best characterized as unconscious.

Finally, one may wonder what specific properties of the stimulus material make it more expressible indirectly. In other words, is it possible to determine which specific contingencies are responsible for the dissociation we obtained? To try to answer this question, we devised an exploratory analysis aimed at identifying which sequences had elicited different patterns of responding either in the SRT or in the generation task. For instance, if the knowledge used in both tasks was perfectly correlated and obeyed grammatical constraints, then we would expect that responses to grammatical successors would have

elicited the fastest RTs during the SRT task and that the same grammatical successors would also be those that were most often generated after the same contexts during the generation task. In the following analysis, we therefore focused on whether there were cases that are in violation of this example.

To make both direct and indirect measures more comparable with each other, we transformed all of the distributions into *Z* scores. Because our data incorporated two different indirect measures between which the net indirect effect could be shared (RT and accuracy), we first obtained a measure of "efficiency" by appropriately averaging the corresponding *Z* scores for these measures. Based on these data, we can now construct a dispersion diagram for each context length, representing the position of each path within the space defined by the generation and efficiency coordinates. These diagrams are represented in Figure 8.

As a first hint to the interpretation of these results, we could see that grammaticality was almost uniformly effective for the indirect measure; that is, the solid points in the figures are almost all to the right of the open points. In contrast, the effect of the direct measure is not so neat: Although some of the solid points are located further upward on the graph than almost any of the open points, there are both open and solid points at the lower end of the graphs (i.e., a number of grammatical paths that remain relatively undergenerated).

One possible account of these data consists of assuming that instead of matching the conditional probabilities that they were exposed to during training, participants tended to generate most often the most likely successor to each context. Such a *response-bias account* presents the advantage of not appealing to separate knowledge bases. Indeed, it merely assumes that a single knowledge base is acquired during training. Participants would subsequently use that knowledge in a biased way during generation, for instance by systematically producing the most likely successor to a given context instead of producing them according to their distribution in the training set. Is such a theory consistent with the data? To find out, we first explored the data and attempted to identify consistent cases. We did not find any clear examples. Next, we proceeded to analyze the data again by comparing the distribution of generation responses with a new theoretical distribution that reflects the influence of a potential bias. To construct this new theoretical distribution, we simply identified which successor was most likely in each context. That particular successor was assigned a probability of 1.0. All the other successors were assigned probabilities of 0.0. In cases where several successors were simultaneously most likely, each was assigned a probability equal to 1.0 divided by the number of tied successors. This new distribution therefore reflects the notion that during generation participants would tend to merely access their memory, retrieve the most likely successor(s) to the context that they were currently exposed to, and always output the retrieved item as their response.

If this account is correct, we would expect to observe that the actual distribution of generation responses correlates better with this new biased distribution than with the distribution of CPs used in the previous analyses. We conducted this analysis for each condition and context length and observed that generation performance was always better or equally corre-

lated with the original distribution of CPs than with the new biased distribution of theoretical probabilities. Therefore, these results lead us to rule out generation bias as a potential source of the dissociation between direct and indirect measures in this task.

Discussion

In this article, we have addressed three separate but related issues about learning in SRT tasks. First, we explored the effects of orientation on learning performance. We did not find any difference between groups of participants who were told about the existence of sequential contingencies and those who were kept unaware of this feature of the stimulus material.

Second and most important, we compared performance on comparable direct and indirect tests of sequence learning and examined the extent to which they are dissociable. Our data suggest that knowledge about the sequential structure of the stimulus material that participants expressed through an indirect RT task is to some extent dissociated from the knowledge that participants could express through a very similar but direct generation task. Specifically, we found that if the two knowledge bases were largely overlapping, some of the knowledge about the sequential structure of the material appeared to be expressible only through its effects on RT performance. Because this knowledge does not appear to be expressible when participants are explicitly asked to use it, as during the generation task, we believe it can be described only as nonconscious.

Third, the experiment was also designed to allow us to assess how much information participants can maintain about the temporal context. We found that participants were sensitive to information provided by sequential stimuli that had appeared one or two but not three time steps before the target. This result is at odds with other results obtained in very similar settings (e.g., Cleeremans & McClelland, 1991). It is therefore important to determine whether this discrepancy can be attributed to variability among participants or testing conditions or whether it reflects the influence of systematic differences such as, for instance, differences in the structural properties of the training material. A natural starting point to try to answer this question is provided by the SRN model of performance in SRT tasks, which was proposed by Cleeremans and McClelland (1991) and has been successfully applied to a wide range of sequential-choice RT situations.

We address these three issues in more detail in subsequent sections of this discussion. There is a fourth issue that we would like to address first, however: What kinds of mechanisms may account for our data? In particular, do they entail separate knowledge bases and learning mechanisms or can we assume that this pattern of results could be produced by mechanisms that operate on a single knowledge base? In the following section, we discuss this issue and explore whether the SRN model is relevant to it.

Simulation of the Learning Results

A major question about the implications of the observed dissociation between direct and indirect measures is whether it

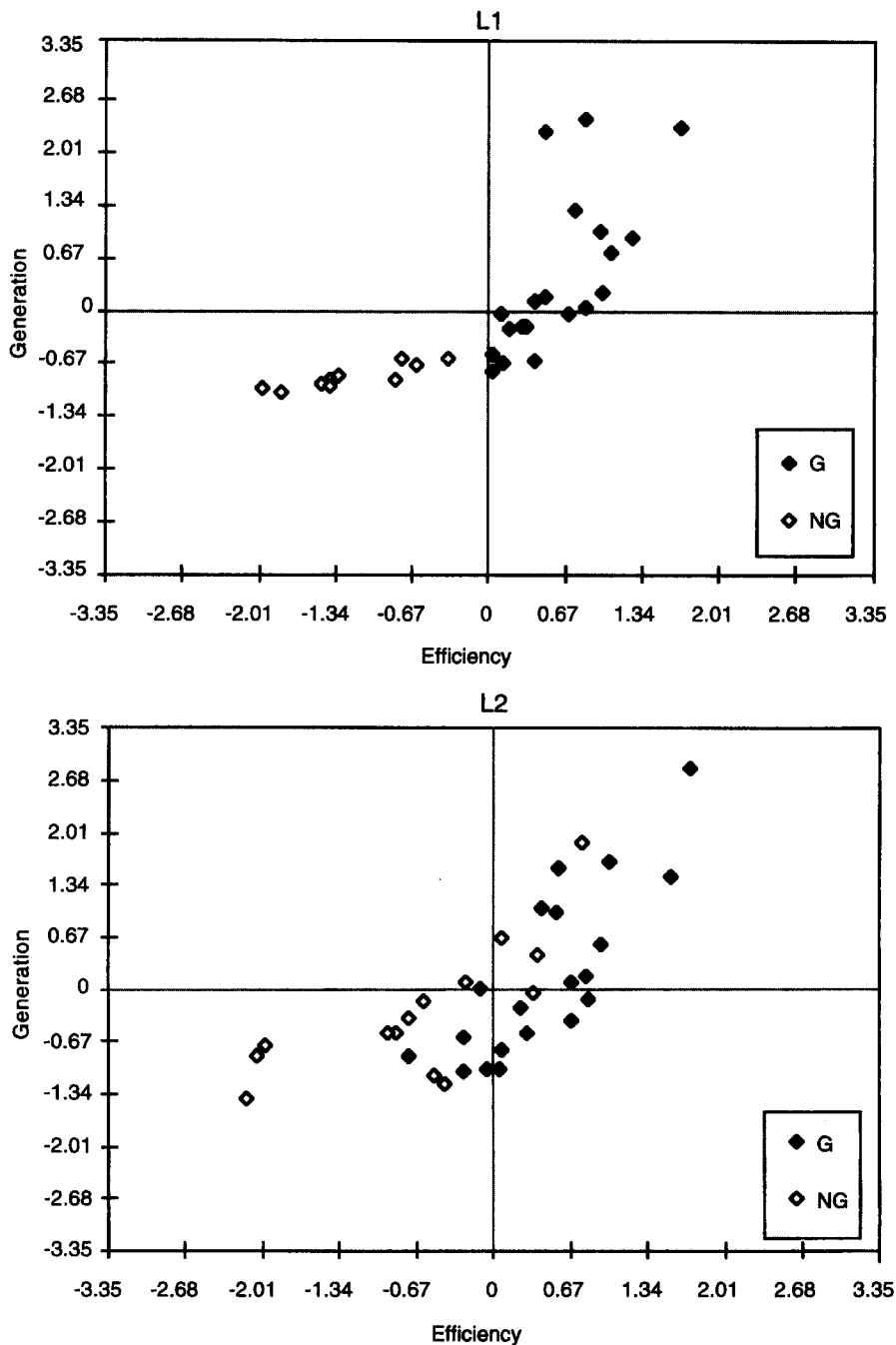


Figure 8. Dispersion diagram representing the position of grammatical (G) and nongrammatical (NG) paths of Length 1 (L1) and Length 2 (L2) within the space defined by the generation and efficiency coordinates.

could have been produced by a single set of learning mechanisms or, conversely, whether it should be taken as reflecting the existence of separate knowledge bases that would independently affect each measure of performance. How to interpret dissociations has recently been addressed by Whittlesea and Dorken (1993), who concluded that dissociations may often express specific differences in the processing operations re-

quired to satisfy the demands of each task rather than expressing deep architectural differences between hypothetical underlying learning mechanisms. In this section, we do not wish to rule out the logical possibility that such an explanation may be applied to the dissociation between SRT and generation performance that we obtained, but we would like to address this issue by assessing whether an SRN with a single

processing and representational pathway could produce a dissociation between direct and indirect measures of its performance.

In the following paragraphs, we first present the SRN and determine through simulation whether it can account for our SRT data. Next, we explore whether the model could be adapted to enable it to produce both identification and generation responses and whether such a model would be able to produce dissociations between these two performance measures.

The SRN, first proposed by Elman (1990) and adapted by Cleeremans and McClelland (1991) to simulate sequential effects in RT tasks is shown in Figure 9. The network uses back-propagation to learn to predict the next element of a sequence on the basis of only the current element and a representation of the temporal context that the network has elaborated itself. Over training, the relative activation of the output units representing each possible successor come to approximate the optimal conditional probabilities associated with their appearance in the current context and can thus be interpreted as representing implicit preparation for the next event. Previous work (see Cleeremans, 1993; Cleeremans & McClelland, 1991) has shown that the SRN is able to account for about 80% of the variance in SRT data.

The main source of discrepancies between our results and those obtained in previous studies that were successfully simulated by the SRN concerns the amount of sequence information that participants are sensitive to while responding to the current element. It appears that participants can learn to respond on the basis of three elements of the temporal context in some cases (Cleeremans, 1993; Cleeremans and McClelland, 1991) but not in others such as the experiment described in this article. A trivial hypothesis about why context sensitivity varies with experiments is that there is variability among participants or in the specific testing conditions. However, it may also be the case that context sensitivity varies as a function of the structure of the sequence used during training. This is supported by the fact that different simulation models of sequence learning exhibit different properties with respect to the type of sequence they can learn. Hence understanding why context sensitivity may vary as a function of the specific

sequence used is important because it may help constrain the range of possible context representations instantiated by different models such as the SRN, the Jordan network (Jordan, 1986; see also Jennings & Keele, 1990), or buffer networks (see Cleeremans, 1993). Conducting a detailed comparison among these models is beyond the scope of this article, but we think it is important to determine whether our data are consistent with the predictions of the SRN model.

One key aspect of learning in the SRN is that the material needs to be "prediction relevant" at every step for its representation to be maintained in the context layer, whereas many other models of sequence learning have been built on the basis of a simpler "moving-window" paradigm (e.g., Cleeremans, 1993; Frensch & Miner, 1994; Jordan, 1986). Of course, if the length of the context window is kept constant either in real time (Frensch & Miner, 1994) or in the number of successive items allowed to be considered (Jordan, 1986), there is no way of explaining why context sensitivity may depend on the sequential structure of the stimulus set presented during training. By contrast, learning in models such as the SRN can take place only if each element of the sequence is useful in predicting the next one.

Does the material used in our experiment fulfill the prediction-relevance condition in all cases? Consider the problem of predicting which elements are legal after an *E*. The label *E* can point to either Node 5 with legal successors *A* and *F* or Node 6 with legal successors *B* or *C*. Hence, to reduce the uncertainty associated with *E*, the network needs to distinguish between its two occurrences in the grammar; that is, it needs to become sensitive to the context in which each instance of *E* may occur. However, in the grammar, both instances of *E* occur only in the context of either *A* or *B*. Hence, knowing that *A* or *B* occurred before *E* is not useful in distinguishing between the two instances of *E*. Therefore, the only way for the network to distinguish between two instances of *E* is to encode which element was presented prior to the *AE*- or *BE*- contexts, that is, *C*, *D*, or *F* according to the grammar. Thus, the contexts *CAE*- and *FAE*-, for instance, unambiguously point to Node 5, whereas the contexts *CBE*- and *FBE*- point to Node 6. Normally, the SRN can become sensitive to constraints set by sequence elements that occurred three time steps before the target, but in the case of this particular grammar, it is almost impossible for it to do so because the material is not fully prediction relevant. Indeed, consider the predictions that the network can make on the basis of *A* or *B*. Both *A* and *B* predict the exact same set of successors, that is, *C*, *E*, and *F*, because in the grammar they both point to exactly the same nodes (Nodes 1 and 3). Because *A* and *B* share the same set of successors, they are not prediction relevant and tend to elicit very similar internal representations during training regardless of which element occurred before *A* or *B*. Distinguishing between *A* and *B*, however, is crucial because contexts such as *CAE*- and *CBE*- point to different nodes, and it is therefore necessary for the network to encode this information if it is to be able to distinguish between the different instances of *E*.

In summary, the SRN therefore actually predicts that encoding third-order regularities will be very hard with this particular grammar, in contrast to what was obtained with other grammars such as the one used by Cleeremans and

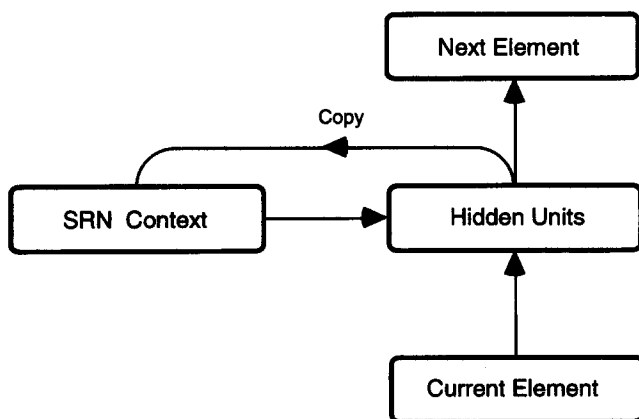


Figure 9. The simple recurrent network (SRN).

McClelland (1991). This prediction depends crucially on a limiting property of the SRN that it does not share with many other architectures for sequence processing, that is, that the material has to be prediction relevant at every step for the network to learn about the material. In the following, we report on simulations that verify this prediction in the context of our experiment.

To assess how well the SRN was able to account for SRT performance in this experiment, we conducted six simulations in which the model was trained on the same material and for the same number of trials as human participants were with the parameters and architecture used by Cleeremans and McClelland (1991). We used an SRN with 15 hidden units and local representations on both the input and output pools (i.e., each unit corresponded to one of the six stimuli). To account for short-term priming effects, the network used dual-connection weights and running-average activations on the output units as described in Cleeremans and McClelland. The network was trained to predict each element of a continuous sequence of stimuli generated in exactly the same conditions as those used for human participants. On each step, a label was generated from the grammar and presented to the network by setting the activation of the corresponding input unit to 1.0. Activation was then allowed to spread to the other units of the network, and the error between its response and the actual successor of the current stimulus was then used to modify the weights. During training, the running average activation of each output unit was recorded on every trial and transformed into Luce ratios (Luce, 1963) to normalize the responses. For the purpose of comparing simulated and observed responses, we assumed (a) that the normalized running average activations of the output units represent response tendencies and (b) that there is a linear reduction in RT proportional to the relative strength of the unit corresponding to the correct response. The network's responses were subtracted from 1.0 to make increases in response strength compatible with reduction in RT.

The data from six simulations were collected and transformed as described above and were then averaged for further analysis. We analyzed these average responses in exactly the same way as for responses from human participants. Figure 10 shows simulated response differences between grammatical and nongrammatical trials after the contexts of L1, L2, and L3 over the 60,000 trials of practice together with a representation of the corresponding RT performance. One can see that the model, like human participants, did not learn L3 contingencies and that it learned at about the same rate as human participants did. Overall, the correlation between the distributions of human RTs and simulated responses was 0.87. However, because human learning effects are shared between RT and accuracy, the degree of overall adjustment between the simulated responses and either of these empirical measures could presumably never be perfect. Of importance, both the model and the human participants did not become sensitive to constraints set by L3 contexts.

In summary, the comparison between simulated and human responses in this experiment thus provides further support for the SRN model as a model of SRT performance. Indeed, the fact that both human participants and the SRN appear incapable of becoming sensitive to such conjunctions lends

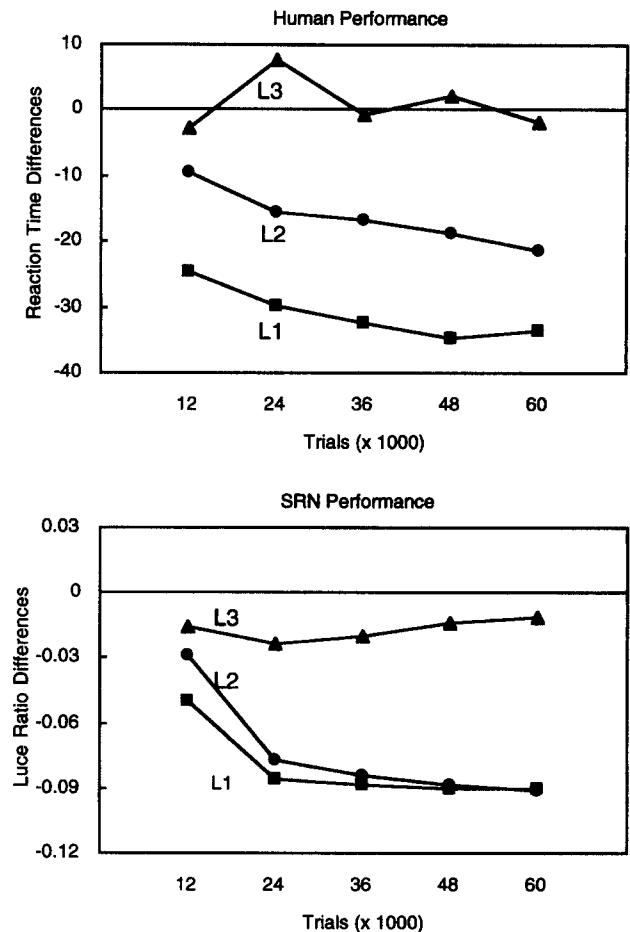


Figure 10. Reaction times for human participants and simulated response differences between grammatical and nongrammatical trials after the contexts of Lengths 1, 2, and 3 (L1, L2, and L3) over the 60,000 trials of practice. SRN = simple recurrent network.

support to the notion that local prediction relevance is an important dimension of sequence learning and runs against the idea that a simple moving-window model would be sufficient to account for sequence-learning data in general.

We now turn to the question of assessing whether the model is relevant to the main issue that we explored in this article, that is, whether the model would be able to produce dissociations between identification and generation responses that would be similar to the empirical dissociation we observed. A first comment is that the architecture as it stands does not distinguish in any way between identification and generation responses because the network is in fact trained to predict the next element to simulate performance in the SRT task. As discussed above, this has proved successful in previous work because when one interprets the network's prediction responses as representing implicit preparation for the next event, the model is capable of successfully accounting for a wide range of human data. However, if one merely interprets the network's responses as reflecting preparation in the context of SRT tasks and prediction in the context of generation, then the model would predict that SRT and generation performance

are always perfectly correlated—a prediction that is demonstrably wrong.

A simple alternative may consist of taking into account the different demands required respectively by the SRT and generation tasks. For instance, during the SRT task, both the model and human participants are indirectly making use of sequence information to prepare for any of the possible successors. It is therefore reasonable and has been demonstrated as accurate to assume that response preparation is proportional to the relative likelihood of these successors. During generation, however, participants, as described in the Results section, are asked to predict the most likely successor and hence to select a single response out of several activated memory representations of the possible successors. Exactly the same response rule can be applied to the model's responses. When we apply this rule to the model's responses, we obtain a distribution of generation responses that is linearly correlated neither with the CPs nor with the network's distribution of responses before the transformation. In other words, the SRN appears to be able to produce a dissociation between implicit preparation and prediction responses. However, and crucially, we have already concluded (see the Results section) that this dissociation is actually inconsistent with the data, because human generation responses always correlate better with the original distribution of CPs than with a new distribution designed to express such a selection rule.

To conclude this section, it would appear that models such as the SRN, which base their performance on a single processing pathway are incapable of accounting satisfactorily for our dissociation data. If this reinforces our conclusion that SRT and generation performance are based on distinct processes, it is difficult at this point to assess whether this SRN's inability to simulate the dissociation results stems from essential or ancillary properties of the architecture. We are currently in the process of assessing whether models that use different structures to support identification and prediction responses are more successful in producing dissociations consistent with our data.

Independence From Orientation

Participants in the intentional condition did not perform better than incidental participants in either the SRT task or the generation task. This is consistent with results from several other studies with the artificial grammar-learning paradigm (Dienes et al., 1991; Dulany et al., 1984; Mathews et al., 1989; Perruchet & Pacteau, 1990) and suggests that implicit learning involves the acquisition of contingencies in the absence of conscious efforts to do so (Berry, 1994; Hayes & Broadbent, 1988; Reber, 1989a, 1993). As Mathews and his coworkers have proposed (Mathews et al., 1989; Stanley, Mathews, Buss, & Kotler-Cope, 1989), such absence of orientation effects should be expected when participants are unlikely to develop and use mental models subsuming the structure of the material, when the salience of the stimulus material is low, and when learning is best described as based on some raw form of memory of each specific contingency. We think that these features provide a good characterization of the probabilistic sequential structure used in these experiments, essentially

because there were no salient features that could guide participants' search for contingencies and because potential rules that could have been induced may always be invalidated by the occurrence of the interspersed random trials.

There may also be other ways to account for the absence of orientation effects in this situation. For instance, one may think that intentional participants gave up on the search for rules early during training and started behaving as incidental participants thereafter. Two points argue against such an interpretation. First, intentional participants differed from incidental participants in that they tended to produce more accurate but slower responses. Second, intentional participants were constantly reminded to search for the rules and that doing so could help them increase their total earnings through better generation performance. Nevertheless, intentional participants did not differ from incidental participants in the distribution of their RTs or in how much they learned about the sequence. Hence, we feel confident that these results suggest that learning of the sequential structure in this paradigm is indeed independent of conscious attempts to do so.

This conclusion stands in sharp contrast with other recent results. In particular, both Frensch and Miner (1994) and Curran and Keele (1993) reported large advantages for participants who knew that the material was sequentially structured and who were asked to look for rules (Frensch & Miner, 1994) or who knew exactly what the sequence was (Curran & Keele, 1993). The most likely hypothesis that may account for the discrepancy appears to be related to the salience of the stimulus material. Both Curran and Keele and Frensch and Miner used extremely simple repeating deterministic sequences, whereas the stimulus material used in this experiment was probabilistic and generated from a finite-state grammar. It seems obvious that conscious attempts to identify regularities will tend to be more successful when the material is both more easily memorized and structured in a deterministic way. The fact that intentional participants from Frensch and Miner's experiment performed better than incidental participants even on the very first trials of the generation task further reinforces the hypothesis that their advantage could be based on the availability of better explicit knowledge about the sequence rather than on the dependence of implicit learning on participants' orientation. Taken together, both results suggest that using complex contingencies such as those instantiated by finite-state grammars is a better strategy to study implicit learning than using simpler, deterministic sequences. In the former case, indeed, the probabilistic nature of contingencies acts as a control for the orientation effects observed in deterministic tasks.

Sequence Learning and Awareness

The results we obtained suggest that learning in this paradigm is not only independent from participants' orientation, but also that some knowledge obtained during training appears to be better used when participants are not directly told to use it (i.e., during the SRT task) than when they are (i.e., during generation). This conclusion is strongly supported by the results of the partial correlational analyses that showed that there is a significant proportion of the variance of the

distribution of performance in the indirect measures that is exclusively explained by the structure of the sequence and that is independent from the variance of the distribution of responses to the generation task.

The comparison between SRT and generation performance on which these results are based fulfills most of the essential requirements of the approach advocated by Reingold and Merikle (1988): We compared otherwise similar direct and indirect measures of the same discrimination that were obtained under identical contextual constraints. Hence, the dissociation observed in this experiment does not appear to be easily discarded as reflecting contextual or dimensional differences between both measures, nor can it be attributed to higher sensitivity of the indirect task to some specific conscious processes. In addition, the use of many more generation trials than are generally used in deterministic tasks (Cohen et al., 1990; Frensch, Buchner, & Lin, 1994; Frensch & Miner, 1994; Nissen & Bullemer, 1987; Willingham et al., 1989) and the use of a more sensitive learning measure than the one used in other probabilistic tasks (Cleeremans & McClelland, 1991) are additional features that make our generation measure more similar to the SRT measure, hence lending support to the validity of this comparative approach.

One potential concern about the use of a continuous generation task such as the one used in this experiment is that it may not be as sensitive as alternative measures such as free generation or recognition, both of which were proposed by Perruchet and Amorim (1992). Detailed inspection of their data and methods makes it somewhat difficult to assess this concern. Indeed, there are several important procedural differences between the two situations. First, Perruchet and Amorim used deterministic material during training, whereas we have used more complex, probabilistic sequences. Under these circumstances, it may be difficult to assess whether lower scores on our direct measure should be interpreted either as evidence for its lack of sensitivity or as an indication that participants do not acquire as much explicit knowledge in probabilistic situations as in deterministic situations. Furthermore, Perruchet and Amorim obtained correlations between direct and indirect measures, but they did not analyze to what extent the observed correlations could be attributed either to sequence learning or to some other bias that would affect both direct and indirect measures. Finally, Perruchet and Amorim computed correlations based on aggregate data averaged over all participants, whereas we computed correlations individually for each participant and subsequently averaged them using a *Z* transformation. We presumed that learning may be idiosyncratic and hence that knowledge expressed by a given participant would correlate better with his or her own performance on another measure than with the performance of the entire group.

Despite these differences, it may still be interesting to compare the two sets of data. For instance, Perruchet and Amorim (1992) have shown that RTs to one sequence position were related to recognition of the four-trial sequence ending in this position and to the number of times that a three-trial or a four-trial sequence ending in this position had been freely generated. However, when other lengths of sequences were considered (either higher or lower), they reported only low

and unreliable correlations. By contrast, we have obtained significant correlations between direct and indirect responses to successors of contexts of L1 and L2 but not to successors of contexts of higher length.

In short, the source of these discrepancies is an interesting problem for additional research, but the methodological differences between the two experiments make it very hard to explore them further in this article. However, the main point we wish to emphasize is that we think that there is no evidence whatsoever, either in the experiment of Perruchet and Amorim (1992) or in its comparison with the results reported in this article, that recognition or free-generation performance constitute more sensitive measures of the knowledge expressed through the SRT task than does the continuous generation task.

In the remainder of this section, we discuss three additional issues that may be used as arguments against interpreting our results as evidence for the existence of some unconscious learning influencing the indirect measure of performance. First, one may ask whether the dissociation we observed could also have arisen because of the lack of reliability of one of the measures. A second issue is whether lack of comparability between the two measures could have produced the dissociation. Finally, a third issue is whether similar dissociations could be obtained between other pairs of measures such as, for instance, two indirect measures. We discuss each issue in turn.

The first point is a valid concern. Indeed, it could be argued that the pattern of partial correlations we obtained is artifactual because it could have resulted from the lower reliability of generation performance as compared with RTs. This is particularly plausible in this case because there are far fewer trials involved during generation than during the SRT task. Although RT and accuracy are naturally noisier than generation (which can be compensated for by averaging over many more trials, as we did), it remains possible that the low number of generation trials we used has resulted in some specific loss of reliability for the direct measure as compared with the indirect measures. If that were the case, then it would come as no surprise that RTs correlate with CPs even when controlling for shared variance with generation scores. To assess whether reliability was a concern in our experiment, we used the split-halves method on each measure for each participant and each context length and also for each practice level in the case of the RT and accuracy measures. Surprisingly, this analysis revealed consistently equivalent or higher reliabilities for generation performance than for the indirect measures. Averaging over participants, generation reliability was 0.989 for L1 contexts, 0.954 for L2 contexts, and 0.954 for L3 contexts. Thus, it appears unlikely that reliability problems are sufficient to account for the dissociation we observed.

The second point is about the possibility that differences other than the direct versus indirect nature of the measures could have resulted in the observed dissociation. It is both obvious and unavoidable that measures such as RT, accuracy, and generation performance all have different metrics and place different temporal demands on participants. However, in this study, we have tried to minimize these differences. One important contextual difference that remains between the two tasks is that they place very different time demands on

participants. Indeed, during the SRT task, participants are instructed to respond as quickly as possible, whereas during generation, there is no time pressure whatsoever. This difference between the two tasks, however, could in fact be construed as enabling participants to maximize their use of conscious knowledge during generation, because they are free to ponder about their response for as long as they want. Therefore, it seems unlikely that this difference could have made generation any less sensitive to conscious information than the comparable indirect measures were.

A third issue about the interpretation of our dissociation results is that similar dissociations may also be observed between other measures such as, for instance, two indirect measures. It may be argued that observing such dissociations between two indirect measures would undermine the argument that dissociations between direct and indirect measures result from system-level dissociations rather than from surface differences between the tasks (see Perruchet & Baveaux, 1989, for a discussion of this point in the field of implicit memory research). However, we believe that such an argument is largely irrelevant. Indeed, if we agree that dissociations could be obtained between many different pairs of measures, we do not think that they necessarily have any implications about the question of awareness. Consider for instance measures such as RT and accuracy during an SRT task. Both RT and accuracy are indirect measures that involve the same discrimination and that are obtained in the same context. Depending on whether or not participants are trading speed for accuracy, these two measures could be positively or negatively correlated. Speed and accuracy may even be completely uncorrelated in some cases. In all these cases, however, the pattern of results would say nothing informative about the conscious or unconscious nature of the underlying knowledge merely because these two measures have not been designed to differ in precisely the way that is required to reflect the influence of conscious or unconscious knowledge. By the same token, a dissociation between given direct and indirect measures could also fail to be informative about the question of awareness if there is any other difference between them that potentially accounts for the observed dissociation.

Hence, the fact that dissociations may also be obtained between other pairs of measures has little bearing on the conclusions that can be brought by following the method advocated by Reingold and Merikle (1988), from a dissociation between otherwise similar direct and indirect measures.

One may wonder, finally, why we consider, along with Reingold and Merikle (1988), that the direct versus indirect dimension is particularly informative with respect to the question of awareness as opposed to any other dimension along which measures of learning could be classified. Briefly, we found that this approach enabled us to avoid several problematic assumptions that have often plagued other approaches, such as the assumption that absolute measures of awareness exist or the assumption that tasks are process pure. In this respect, Reingold and Merikle's framework shares many features with the process-dissociation procedure suggested by Jacoby (1991, 1994; Jacoby, Toth, & Yonelinas, 1993). Indeed, both approaches rely on the notion that consciously established goals (i.e., goals established through

explicit instructions) have a privileged relationship with conscious information. Jacoby's approach is based on the reasoning that knowledge expressed even in cases where participants are explicitly told not to use it must be considered as automatic. Reingold and Merikle's approach is similarly based on the assumption that knowledge that is exclusively expressed when participants are not told to rely on it must also be considered as nonconscious. We believe that both frameworks constitute major methodological advances in the field of implicit learning because they enable us to free ourselves from otherwise untenable assumptions about the nature of awareness.

References

- Allport, D. A. (1988). What concept of consciousness? In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 159–182). Oxford, England: Clarendon Press.
- Berry, D. C. (1994). Implicit learning: Twenty-five years on. In M. Moscovitch & C. Umiltà (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 755–782). Cambridge, MA: Bradford.
- Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 36(A), 209–231.
- Berry, D. C., & Broadbent, D. E. (1987). The combination of explicit and implicit learning processes in task control. *Psychological Research*, 49, 7–15.
- Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit–explicit distinction. *British Journal of Psychology*, 79, 251–272.
- Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance*. London: Sage.
- Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120, 235–253.
- Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 17–30.
- Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 189–202.
- Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 875–887.
- Dienes, Z., & Fahey, R. (1995). The role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 848–862.
- Dulany, D. E., Carlson, R. C., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General*, 113, 541–555.
- Dulany, D. E., Carlson, R. C., & Dewey, G. I. (1985). On consciousness in syntactical learning and judgment: A reply to Reber, Allen, and Regan. *Journal of Experimental Psychology: General*, 114, 25–32.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Frensch, P. A., Buchner, A., & Lin, J. (1994). Implicit learning of unique and ambiguous serial transitions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 567–584.

- Frensch, P. A., & Miner, C. S. (1994). Individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*, *22*, 95–110.
- Hayes, N. A., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition*, *28*, 249–276.
- Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences*, *9*, 1–66.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.
- Jacoby, L. L. (1994). Measuring recollection: Strategic vs. automatic influences of associative context. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious processing* (pp. 661–679). Cambridge, MA: Bradford.
- Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*, 139–154.
- Jennings, P., & Keele, S. W. (1990). A computational model of attentional requirements in sequence learning. *Proceedings of the 12th Annual Conference of the Cognitive Science Society* (pp. 876–883). Hillsdale, NJ: Erlbaum.
- Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the 8th Annual Conference of the Cognitive Science Society* (pp. 10–17). Hillsdale, NJ: Erlbaum.
- Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.
- Marcel, A. J., & Bisiach, E. (1988). *Consciousness in contemporary science*. Oxford: Clarendon Press.
- Marescaux, P. J., & Karnas, G. (1991). *The implicit versus explicit knowledge distinction revisited: When finding associations between verbalizable knowledge and some performance criteria*. (Tech. Rep. No. 4PR3GK). Faculté des Sciences Psychologiques et Pédagogiques, Université Libre de Bruxelles.
- 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.
- Merikle, P. M., & Reingold, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study nonconscious memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 224–233.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.
- Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing* (pp. 41–82). Potomac, MD: Erlbaum.
- Perruchet, P., & Amorim, P. A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 785–800.
- Perruchet, P., & Baveaux, P. (1989). Correlational analyses of explicit and implicit memory performance. *Memory & Cognition*, *17*, 77–86.
- Perruchet, P., & Gallego, J. (1993). Association between conscious knowledge and performance in normal subjects: Reply to Cohen and Curran (1993) and Willingham, Greeley, and Bardone (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1438–1444.
- Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge. *Journal of Experimental Psychology: General*, *119*, 264–275.
- Perruchet, P., & Pacteau, C. (1991). The implicit acquisition of abstract knowledge about artificial grammar: Some methodological and conceptual issues. *Journal of Experimental Psychology: General*, *120*, 112–116.
- Reber, A. S. (1989a). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.
- Reber, A. S. (1989b). More thoughts on the unconscious: Reply to Brody and to Lewicki and Hill. *Journal of Experimental Psychology: General*, *118*, 242–244.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge*. Oxford, England: Oxford University Press.
- Reber, A. S., & Allen, R. (1978). Analogy and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition*, *6*, 189–221.
- Reber, A. S., Allen, R., & Regan, S. (1985). Syntactical learning and judgment, still unconscious and still abstract: Comment on Dulany, Carlson, and Dewey. *Journal of Experimental Psychology: General*, *114*, 17–24.
- Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*, *44*, 563–575.
- Sanderson, P. M. (1989). Verbalizable knowledge and skilled task performance: Association, dissociation, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 729–747.
- Schacter, D. L., Bowers, J., & Booker, J. (1989). Intention, awareness and implicit memory: The retrieval intentionality criterion. In S. Lewandowsky, J. C. Dunn, & F. Kirsner (Eds.), *Implicit memory: Theoretical issues*. Hillsdale, NJ: Erlbaum.
- Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable learning systems. *Behavioral and Brain Sciences*, *17*, 367–395.
- Stanley, W. B., Mathews, R. C., Buss, R. R., & Kotler-Cope, S. (1989). Insight without awareness: On the interaction of verbalization, instruction and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *41(A)*, 553–577.
- Velmans, M. (1991). Is human information processing conscious? *Behavioral and Brain Sciences*, *14*, 651–726.
- 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.
- Willingham, D. B., Greely, T., & Bardone, A. M. (1993). Dissociation in a serial response task using a recognition measure: Comment on Perruchet and Amorim (1992). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1424–1430.
- Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1047–1060.

Appendix A

Correlation (Averaged Over Participants) Between the Distribution of Conditional Probabilities (CPs) and Generation (GEN) on the Selected Paths

Length	r (CPs, GEN)
L1	0.47
L2	0.22
L3	-0.01

Note. L = length.

Appendix B

Correlation (Averaged Over Participants) Between the Distribution of Conditional Probabilities (CPs) and Either Reaction Time (RT) or Accuracy (ACC) Averaged Over Sets of Four Sessions on the Selected Paths

Session	L1		L2		L3	
	r (CPs, RT)	r (CPs, ACC)	r (CPs, RT)	r (CPs, ACC)	r (CPs, RT)	r (CPs, ACC)
1-4	-0.34	0.33	-0.15	0.07	-0.08	-0.08
5-8	-0.50	0.53	-0.27	0.23	0.08	-0.04
9-12	-0.58	0.55	-0.34	0.24	0.01	0.10
13-16	-0.59	0.64	-0.38	0.27	0.03	0.00
17-20	-0.60	0.62	-0.41	0.33	-0.03	0.01

Note. L = length.

Appendix C

Correlation (Averaged Over Participants) Between the Distribution of Generation (GEN) and Either Reaction Time (RT) or Accuracy (ACC) Obtained at the End of Training (Sessions 17-20) on the Selected Paths

	L1		L2		L3	
	r (GEN, RT)	r (GEN, ACC)	r (GEN, RT)	r (GEN, ACC)	r (GEN, RT)	r (GEN, ACC)
	-0.48	0.35	-0.43	0.22	-0.18	0.15

Note. L = length.

Received January 11, 1995
 Revision received August 7, 1995
 Accepted August 22, 1995 ■