Journal of Experimental Psychology:
Learning, Memory, and Cognition
2008, Vol. 34, No. 1, 80–96

Copyright 2008 by the American Psychological Association
0278-7393/08/$12.00 DOI: 10.1037/0278-7393.34.1.80

An Attention-Based Associative Account of Adjacent and Nonadjacent Dependency Learning

Sébastien Pacton
Université Paris Descartes and
Centre National de la Recherche Scientifique

Pierre Perruchet
Université de Bourgogne and
Centre National de la Recherche Scientifique

In 5 experiments, results showed that when participants are faced with materials embedding relations between both adjacent and nonadjacent elements, they learn exclusively the type of relations they had to actively process in order to meet the task demands, irrespective of the spatial contiguity of the paired elements. These results are consonant with current theories positing that attention is a *necessary* condition for learning. More important, the results provide support for a more radical conception, in which the joint attentional processing of 2 events is also a *sufficient* condition for learning the relation between them. The well-documented effect of contiguity could be a by-product of the fact that attention generally focuses on contiguous events. This reappraisal considerably extends the scope of approaches based on associative or statistical processes.

Keywords: attention, statistical computation, associative learning, language

The past 2 decades or so have been the theater of an attempt to account for a growing number of higher cognitive activities in a framework rooted in the notions of associative (or statistical) learning. This line of research has been especially lively in language, where the so-called statistical/distributional approaches are sometimes construed as viable alternatives to the more conventional rule-based perspective (e.g., Redington & Chater, 1998; Seidenberg & MacDonald, 1999; for brief overviews, see Perruchet, 2005; Perruchet & Pacton, 2006).

By and large, such approaches have focused on the human abilities to detect and exploit the relations between adjacent elements. This focus is consistent with the traditional literature on associative learning. In considering the domain of animal conditioning, or studies on paired-associate learning in humans, we observe that, in all cases, the items to be associated are displayed in close temporal or spatial proximity. This is also true for the more recent studies on implicit learning (e.g., Stadler & Frensch, 1998). For instance, the finite-state grammars that are commonly used in artificial grammar studies govern relations between contiguous elements. As a consequence, the models of implicit learning relying on the formation of chunks of adjacent events have proved to be especially successful in those situations (e.g., Perruchet, Vinter, Pacteau, & Gallego, 2002; Servan-Schreiber & Anderson, 1990).

Several studies have shown that adjacent relations are far more relevant for language acquisition than has been claimed in the past. In a series of oft-cited experiments, Saffran, Newport, and Aslin (1996) showed that the formation of the lexicon partly relied on statistical relations between adjacent syllables. Moreover, Redington, Chater, and Finch (1998) showed that highly local context provided a considerable amount of information about the syntactic categories of words. However, linguistic structures also embed remote dependencies. That is to say, a relation exists between two events, A and C, irrespective of the intervening events (hereinafter, this structure is referred to as A*X*C, where *X* stands for a variable event, which is statistically independent of both A and C). Such relations are found at different levels, from the subsyllabic level (e.g., the short vs. long pronunciations of vowels according to the presence of a "silent e" ending, irrespective of the intermediary consonant, as in CAP–CAPE, CAR–CARE; Stanback, 1992) to morphosyntactic relationships (e.g., between auxiliaries and inflectional morphemes, as in "*is* writ*ing*," irrespective of the verb stem) and hierarchical structures (e.g., in center-embedded sentences, such as "the rat the cat ate stole the cheese"; Perruchet & Rey, 2005). Nonadjacent dependencies are also present in other domains of high-level knowledge, such as music. Likewise, capturing the relationships between distant objects seems essential. As claimed by Turk-Browne, Jungé, and Scholl (2005), "people are constantly bombarded with noise in space and time that needs to be segregated in order to extract a coherent representation of the world, and people rarely encounter a sequence of relevant stimuli without any interruptions" (p. 562).

In the past few years, this state of affairs has given rise to a set of studies investigating the possibility of learning nonadjacent dependencies in artificial languages (Bonatti, Peña, Nespor, &

Sébastien Pacton, Institut de Psychologie, Laboratoire de Psychologie et Neurosciences Cognitives, Université Paris Descartes, Paris, France, and Centre National de la Recherche Scientifique, Paris, France; Pierre Perruchet, Laboratoire d'Etude de l'Apprentissage et du Développement / Centre National de la Recherche Scientifique, Université de Bourgogne, Dijon, France.

This work was supported by the Centre National de la Recherche Scientifique (CNRS; Formation de Recherche en Evolution [FRE] 2987 and Unité Mixte de Recherches 5022), the Université Paris Descartes, and the Université de Bourgogne.

Correspondence concerning this article should be addressed to Sébastien Pacton, Institut de Psychologie, Laboratoire de Psychologie et Neurosciences Cognitives, Université Paris Descartes (CNRS, FRE-2987), 71 Avenue Edouard Vaillant, 92774, Boulogne-Billancourt, Cedex, France. E-mail: sebastien.pacton@univ-paris5.fr

Mehler, 2005; Gómez, 2002; Newport & Aslin, 2004; Onnis, Monaghan, Richmond, & Chater, 2005; Peña, Bonatti, Nespor, & Mehler, 2002; Perruchet, Tyler, Galland, & Peereman, 2004), music (Creel, Newport, & Aslin, 2004; Kuhn & Dienes, 2005), and visual perception (Turk-Browne et al., 2005).[1] The results show a consensus on the conclusion that learning nonadjacent dependencies is possible, but under far more restrictive conditions than those required for learning the relations between contiguous events. The question to be debated is, What is the nature of these constraints? More specifically, can these constraints be encompassed within a general theory of associative learning, or are they specific enough to suggest that learning adjacent and nonadjacent dependencies constitutes radically different forms of learning?

An Overview of Earlier Studies

In an initial attempt to address the questions outlined above, Gómez (2002) showed that, in a situation where the successive A*X*C units were perceptually distinct, the degree to which the A_C relationships were learned depended on the variability of the middle element (*X*). More precisely, participants were presented with 2, 6, 12, or 24 different *X* elements, and learning was observed to increase markedly under conditions of greatest variability in both adults and infants (see also Gómez, 2006; Onnis, Christiansen, Chater, & Gómez, 2003). According to Gómez, when *X* is variable, the adjacent relations (i.e., A*X* and *X*C) are not stable enough to be learned, and thus people shift their focus onto the only stable relation available in the language, namely, that between A and C.

For Creel, Newport, and Aslin (2004), the crucial factor is the similarity between A and C, more precisely, the higher level of similarity existing between A and C than between A (or C), on the one hand, and *X,* on the other hand. Similarity can be assessed on an acoustic dimension. Using musical tone sequences, Creel et al. (2004) showed that nonadjacent dependencies were not acquired when all of the elements differed equally one from one another, whereas learning was successful when A and C were similar in pitch or timbre, and different from *X.* The role of acoustic similarity was also emphasized by Onnis et al. (2005), who concluded from their studies of artificial language that no learning was obtained without some degree of phonological similarity between A and C syllables. To account for their results, Newport and Aslin (2004) invoked the perceptual grouping principles identified by Gestalt psychologists and, notably, the law of similarity. In other studies from the same laboratory, similarity was assessed in terms of categorical classification. Newport and Aslin (2004) failed to observe learning with nonadjacent syllables (i.e., A, *X,* and C were syllables), whereas learning occurred when A and C were consonants and *X* was a vowel and, conversely, when A and C were vowels and *X* was a consonant.

According to Bonatti et al. (2005), consonants and vowels do not serve the same function in language, with consonants giving cues about the lexicon and vowels being tied more closely to syntax. Starting from the postulate that statistical analyses mainly serve to build a lexicon, Bonatti et al. predicted that nonadjacent relations between consonants would be learned, whereas nonadjacent relations between vowels would remain undetected. The authors provided some experimental support for their prediction. This proposal suggests that the conditions allowing the learning of nonadjacent dependencies are linked to language-specific features and, thus, are different from the general purpose mechanisms serving to discover the relations between contiguous events. Indeed, Peña et al. (2002) argued that a statistical or distributional approach is unable to account for the discovery of nonadjacent dependencies. As experimental support, Peña et al. showed that this form of learning was only possible when short pauses were introduced between the A*X*C words during the familiarization phase. The role of these pauses, Peña et al. asserted, was to make the speech stream more similar to natural language. By doing so, the pauses trigger computational processes oriented toward the discovery of the speech structure, which the authors argued are entirely different from those involved in the discovery of statistical patterns.

This brief overview leaves a somewhat fuzzy picture. Considering the different studies jointly suggests that none of the learning conditions that have been a focus up until now may be taken as decisive. For instance, the emphasis of Gómez (2002) and Onnis et al. (2003) on the need for a highly variable intervening event is challenged by the positive results obtained by other contributors, who generally used only two or three different *X*s (e.g., Newport & Aslin, 2004; Perruchet et al., 2004). Likewise, the emphasis on the learning of relationships between consonants (Bonatti et al., 2005) or the need for pauses between A and C units cannot be taken as universal prerequisites given the studies showing that relations between vowels are learned as well and that pauses are unnecessary (Newport & Aslin, 2004). As a consequence, the issue of whether learning nonadjacent dependencies actually requires conditions that are qualitatively different from those required for learning adjacent dependencies remains an open question.

A Unitary View of Adjacent and Nonadjacent Dependency Learning

The view proposed here is aimed at accounting for the learning of both adjacent and nonadjacent dependencies within a unified theoretical framework. Our framework relies heavily on the role of attentional processes. Emphasizing the role of attention in associative learning is far from new. Although the initial theories of conditioning mostly focused on the contiguity of the stimuli (Guthrie, 1930), a number of studies in the 1960s and 1970s showed that contiguity was not sufficient to trigger the learning of associations. Those studies led to the emergence of models centered on the need for learners' attentional involvement (e.g., Mackintosh, 1975). Since then, empirical data suggesting that learning does not occur without a minimal level of attention have accumulated, especially in the area of implicit, nonintentional forms of learning, which has

[1] Turk-Browne et al. (2005) reported successful learning of relationships between visual shapes of a given color when the related items are interrupted by irrelevant items from another color stream. Turk-Browne et al.'s study was not directly aimed at studying nonadjacent dependencies, and as discussed by the authors, the results are only suggestive of the possibility of what they called "interrupted statistical learning." Indeed, the occurrence of the intervening events was randomly determined, hence occasionally allowing the presentation of relevant events in immediate succession. However, a subsequent study by the same authors, currently published in abstract form (Jungé, Turk-Browne, & Scholl, 2005), suggests that learning persists when the "interrupting" elements are perfectly interleaved.

been a primary focus of recent research (e.g., Frensch, Buchner, & Lin, 1994; Hoffman & Sebald, 2005; Hsiao & Reber, 1998; Jimenez & Mendez, 1999; Whittlesea & Dorken, 1993). We consider this research in more detail in the General Discussion. It is sufficient here to point out that all of these studies converge toward the conclusion that attention is a *necessary* condition for learning. Indeed, what is shown is that even when the objective conditions that are, in principle, conducive to learning are fulfilled (especially those concerning the contiguity of events), learning may fail to occur whenever the learner's attention is not focused on the relevant events.

As such, a learning model positing the necessity of attention remains unable to account for the learning of nonadjacent relations. Indeed, the need for attentional involvement does not preclude other conditions from being effective, especially those concerning the temporal or spatial adjacency of events. To account for nonadjacent dependency learning within an attentional model, one must postulate, in addition, that attention is a *sufficient* condition for learning to occur. The claim that attention is a sufficient condition means that no other condition is required, neither in participants' dispositions (such as their intention to learn) nor in the external display (such as the spatial or temporal relationships between events). Thus the critical postulate of our model is that the joint attention given to a pair of events is a *necessary and sufficient* condition for the emergence of associative learning and memory. In line with this postulate, several authors have proposed that associative learning is an automatic process that links together all of the components that are present in the attentional focus at a given point (Frensch & Miner, 1994; Logan & Etherton, 1994; Perruchet & Vinter, 2002; Stadler, 1995; Treisman & Gelade, 1980). However, to the best of our knowledge, the potential of this proposal to account for nonadjacent dependency learning has yet to be exploited.

If the spatial or temporal contiguity of the relevant events is not a mandatory condition for associative learning and memory, one may wonder why this condition has received such considerable (and once exclusive; see Guthrie, 1930) attention throughout the past research on learning. It may be that, in an overwhelming proportion of earlier associative learning studies, the objective contiguity of the to-be-associated elements and the joint attentional processing of those elements were confounded. For instance, in standard classical conditioning studies, the conditioned and unconditioned stimuli were generally displayed within a temporal window of 500 ms or so, with a long blank interval between successive pairings. In those conditions, there is no doubt that the two relevant events are attentionally processed together, and this leaves open the possibility that the causal factor was the joint attentional processing, not the contiguity per se. More generally, the mental content composing the attentional focus at a given moment has a high chance of representing events that are close on spatial and/or temporal dimensions in the environment. Certainly, the attentional content may also encompass events that are not adjacent in the environment, but some specific reasons are needed to pay joint attention to those events. Thus, our attention-based view accounts for both the easy formation of associations between contiguous events (those events are generally represented as mental chunks) and the far more limited ability to build associations between nonadjacent events (the joint attentional processing of those events requires some special conditions). In this view, the conditions promoting learning nonadjacent dependencies are the same as those that allow the joint attentional processing of the relevant events.

This theoretical proposal appears to be compatible with the available findings on nonadjacent dependency learning. For instance, the variability of the intervening event (X) may draw participants' attention toward the repeated elements (A and C) as a result of their relative frequencies, hence favoring the discovery of their relations, as observed by Gómez (2002)[2] and Onnis et al. (2003). The factors revealed in the studies by Creel et al. (2004) and Newport and Aslin (2004; see also Onnis et al., 2005) originally accounted for by reference to the Gestalt law of similarity, may also be encompassed in a view centered on attentional processing. Indeed, as has been occasionally suggested (e.g., Driver & Baylis, 1998; Scholl, 2001), an equivalent claim could be made that an array of elements forms a perceptual group, as defined by the traditional Gestalt principles, with the assertion that attention automatically spreads across those elements, hence integrating them within a single chunk. More indirectly, an attention-based interpretation is also consistent with Peña et al.'s (2002) data. Indeed, Peña et al. failed to observe learning when the A*X*C words were played as an unbroken stream, whereas successful learning of nonadjacent dependencies was obtained when the speech stream was perceived as a succession of A*X*C words (as shown in Perruchet et al., 2004). Without pauses between the A*X*C units, there is no reason to jointly focus on A and C rather than on any other subset of syllables. By contrast, when participants perceive the auditory stream as a succession of A*X*C words, they pay more attention to the first and last syllables than to the intervening ones, because the start and the end of a sequence capture more attention than do the intermediary events (e.g., Cowan, 1991).

The Present Study

The fact that, by and large, an attention-based model of associative learning is able to provide an a posteriori account for earlier results provides only partial support for the model. This is especially true given that the nature of the features likely to capture attention is open to debate, and there is a risk of circularly defining these features as those that promote associative learning. Stronger evidence would be provided if one were able to draw new predictions from our model. It is worth noting in this context that attention can be captured by the intrinsic features of the environment but can also be internally driven. That is to say, attention can be oriented toward adjacent or nonadjacent aspects of the very same physical display as a function of task demands. This property allows original predictions, as is illustrated after a brief outline of our experimental approach.

[2] Gómez (2002) also conferred a role to attention, but in a different way. Indeed, Gómez assumed that attentional processes are captured by a stable relation. This means that detecting the stability of the relation is a prerequisite for this relation to fall into the focus of attention. Note, however, that Gómez's interpretation leaves open the question of the mechanisms allowing the detection of stable relations, which may be thought of as the hallmark of learning. By contrast, we assume that certain events capture attention, for instance, because of their intrinsic properties, with the detection of their relations being a by-product of the joint attentional processing of those events.

In the experiments described below, participants were faced with a set of problems, each consisting of a sequence of digits embedding both adjacent and nonadjacent regularities. The participants' task involved (a) the sequential processing of each individual digit of the sequence and (b) the joint processing of two selected digits. These two digits were adjacent for a first group of participants and nonadjacent for a second group. A subsequent recognition test explored how well participants from the two groups learned both adjacent and nonadjacent dependencies. The predictions of different theories are represented in Table 1. We reasoned as follows: If learning depends exclusively on the contiguity between elements, then participants should learn adjacent relations, irrespective of the task they have to perform. If learning depends on both attention and contiguity, then participants should learn only adjacent relations in the condition in which the task demand leads to a focus on those relations. This second possibility would be consistent with the prevalent associative learning theories, which posit that attention is necessary without concurrently relaxing the condition of contiguity. Finally, if learning depends exclusively on attention, irrespective of the objective relations between elements, then participants would learn adjacent relations when they focus on adjacent relations and nonadjacent relations when they focus on nonadjacent relations. In other words, when contiguity and attention are manipulated orthogonally, participants should learn the relations between the elements on which they focus to meet the task demand, irrespective of their objective contiguity. Our hypothesis was in keeping with this third possibility.

Experiment 1

In Experiment 1, participants were presented with a set of 20 problems. Each problem consisted of a digit on the left, which indicated the target for this problem, followed by a sequence of 20 digits. Each sequence consisted of two occurrences of an adjacent dependency (e.g., the digit 4 systematically followed the digit 6) and two occurrences of a nonadjacent dependency (e.g., the digit 7 systematically followed the digit 3, with a variable digit intervening between the digits 3 and 7). An example of a sequence is: 5 8 5 6 4 3 9 7 1 0 8 6 4 9 3 0 7 2 2 1 (see Figure 1). Participants had to browse the 20-digit sequences, from left to right, in order to detect the targets. When participants detected a target, they had to calculate the difference between two digits, either the two digits that immediately followed the target or the two digits that immediately surrounded the target. All of the participants had to circle the target and the two digits they had to subtract from each other when the absolute value of their difference equaled three. It is worth noting that participants never had to circle the elements forming the adjacent and nonadjacent dependencies, as neither the difference between 6 and 4 nor the difference between 3 and 7 equals 3 (or −3). After this study phase, the same recognition test was used in the two groups to assess the learning of both adjacent and nonadjacent dependencies.

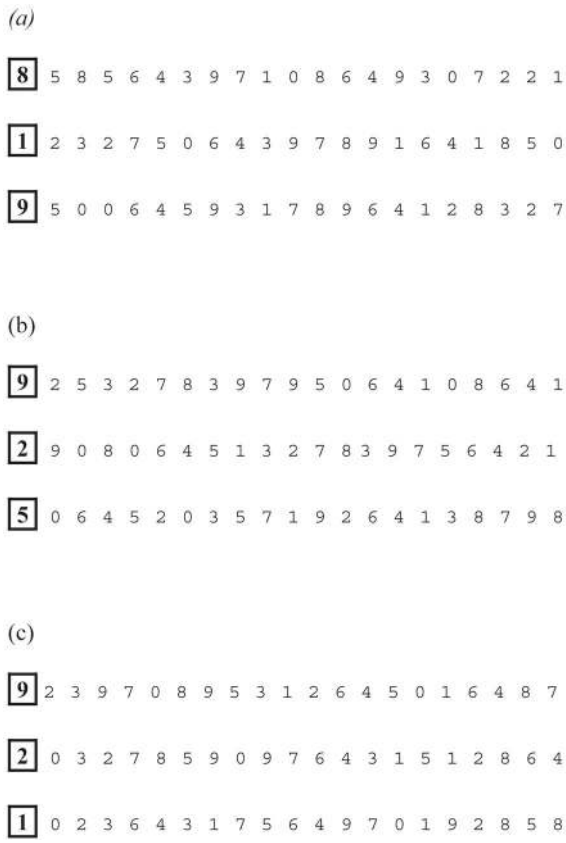

Figure 1. Examples of problems used in the three experiments. Each problem was formed with one target digit on the left, within a rectangle, and one 20-digit sequence on the right: (a) three problems used in Experiment 1, adjacent task; (b) three problems used in Experiments 1, 3, 4a, and 4b, nonadjacent task; and (c) three problems used in Experiment 2 nonadjacent task. In all of these examples, the adjacent regularity was *64*, and the nonadjacent regularity was *3X7*.

Table 1
Predictions of Three Theoretical Models of Learning

Relations assessed during the recognition test	Contiguity only	Contiguity plus attention	Attention only
Adjacent task-induced attention			
Adjacent	+	+	+
Nonadjacent	−	−	−
Nonadjacent task-induced attention			
Adjacent	+	−	−
Nonadjacent	−	−	+

Note. The table shows predictions of three theoretical models of learning as a function of whether the task focuses participants' attention on adjacent or nonadjacent dependencies. A plus sign indicates that learning should occur, and a minus sign indicates that learning should not occur.

An attention-based model predicts an interaction between the type of dependency and the task demand. Indeed, adjacent dependencies, but not nonadjacent dependencies, should be learned when participants have to process adjacent items, whereas the reverse is expected when participants have to process nonadjacent items. Note that this pattern of results cannot be attributed to intrinsic properties of the material, as the material included the very same dependencies for all of the participants.

Method

Participants. A total of 192 undergraduate students from the Université Paris Descartes, Paris, France, participated in the experiment in partial fulfillment of a course requirement. All participants were native French speakers. Participants were randomly assigned to one of the four Dependency Type × Digit Combination experimental groups ($n = 48$ for each group).

Materials. The stimuli used in the study phase consisted of 20 problems composed of a target digit inside a rectangle on the left (which indicated the value of the target for this problem) and a sequence of 20 digits on the right (see Figure 1). In every sequence, (a) each of the 10 digits (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) occurred twice, (b) a pair of digits forming an adjacent dependency (e.g., *64*) occurred twice, and (c) a pair of digits forming a nonadjacent dependency (e.g., *3X7*) occurred twice. The adjacent and nonadjacent pairs were identical for all of the problems presented to a given participant. For each problem, the target digit was selected randomly among the digits 0, 1, 2, 5, 8, and 9, that is, those digits that were not elements of the pairs forming the adjacent and nonadjacent dependencies. When the task involved a subtraction between adjacent digits, the target digit occurred once just before one of the two pairs of digits forming an adjacent dependency and once in another position. When the task involved a subtraction between nonadjacent digits, the target digit occurred once between the two pairs of digits forming a nonadjacent dependency and once in another position. Figure 1a shows three problems with the adjacent pair *64*, and with the nonadjacent pair *3X7*, when the task consisted in deciding whether the difference between the two digits following the target digit equals three (this is the case for the second problem in which the digits 8 and 5 follow the target 1). Figure 1b shows three problems with the adjacent pair *64*, and with the nonadjacent pair *3X7*, when the task consists in deciding whether the difference between the two digits surrounding the target digit equals three (this is the case for the second problem in which the digits 4 and 1 surround the target 2).

For each participant, a set of 20 problems was generated. For each problem, the value of the target digit, the position of the two pairs forming the adjacent dependency, the position of the two pairs forming the nonadjacent dependency, and the position of the residual digits were determined randomly. Different randomizations were used for each participant. This led to the construction of 192 different sets of 20 problems. For half of the problems, the adjacent dependencies were *64* and the nonadjacent dependencies were *3X7*, and for the other half, the adjacent dependencies were *37* and the nonadjacent dependencies were *6X4*. Each of the 192 sets of 20 problems was printed in Verdana font, size 12, on a single A4 sheet of paper (21.0 × 29.7 cm). The distances were 1.5 cm between the target digit inside the rectangle and the first digit of the sequence and 0.7 cm between two successive digits of the sequence. There was a distance of 1.0 cm between two successive problems.

A two-choice recognition test was used to assess participants' learning of adjacent and nonadjacent dependencies. The stimuli used in this test consisted of 16 pairs of three-digit strings. Each pair contained a legal string and an illegal string. Eight pairs were used to assess the learning of adjacent dependencies. For the participants who were exposed to *64* as the adjacent dependency and *3X7* as the nonadjacent dependency during the study phase, the legal string was *64X* for four pairs and *X64* for four other pairs. The illegal string resulted from a spacing violation in four pairs *6X4* and from an order violation in four other pairs *46X* or *X46*. Eight other pairs were used to assess the learning of nonadjacent dependencies. The legal string was *3X7* for the eight pairs. The illegal string resulted from a spacing violation in four pairs *37X* or *X37* and from an order violation in four other pairs *7X3*. Pairs were elaborated on the basis of the same principle for the participants who were exposed to the opposite digit combination, that is, *37* as the adjacent dependency and *6X4* as the nonadjacent dependency during the study phase. The digit *X* was always selected randomly among the six digits used as target digits in the study phase (0, 1, 2, 5, 8, 9). The digit *X* was the same in the legal string and in the illegal string of a pair (e.g., 640 – 604; 387 – 378). For each participant, the 16 pairs of three-digit strings were printed in Verdana font, size 12, on a single A4 sheet of paper. The distances were 0.7 cm between two successive digits of a sequence and 2.0 cm between the third digit of the left string and the first digit of the right string of a pair. There was a distance of 1.5 cm between two pairs.

Procedure. The experiment was carried out with groups of 15 to 20 participants. In the study phase, the participants were provided with the sheet of paper containing the 20 problems. Participants were asked to look for the target digit within each sequence of 20 digits while going through the sequence from left to right. They were informed that the target digit could occur more than once within a sequence. At this point, half of the participants were asked to calculate whether the absolute value of the difference between the two digits following the target digit equaled 3 (i.e., was +3 or −3), a task that entails the processing of two adjacent digits. The other half of the participants were asked to calculate whether the absolute difference between the two digits surrounding the target digit equaled 3, a task that entails the processing of two nonadjacent digits. The two groups had to circle the target and the two digits when this condition was fulfilled (this condition was fulfilled in at least two problems for each participant). The participants were not informed that they would have to perform a recognition test after the study phase. Completing the 20 problems of the study phase usually took less than 10 min.

After this task, the sheets of paper used in the study phase were withdrawn and replaced by those for the recognition test. All of the participants, regardless of the group to which they were assigned, were asked to select which of two strings of three digits was part of the digit sequences viewed in the study phase of the experiment. Participants were told that they had to make a choice for each of the 16 pairs, even if they sometimes felt that they were guessing.

Results

Figure 2 shows the mean percentage of correct responses in the recognition test for items evaluating the learning of adjacent and nonadjacent dependencies as a function of the task performed during the study phase. An analysis of variance (ANOVA) was performed on the number of correct responses, with type of subtraction (subtraction between adjacent digits vs. subtraction between nonadjacent digits) and digit combination (the adjacent pair *64* and the nonadjacent pair *3X7* vs. the adjacent pair *37* and the nonadjacent pair *6X4*) as between-subjects variables and dependency type (adjacent vs. nonadjacent) as a within-subject variable. In this and subsequent experiments, any effect with $p < .10$ is reported.

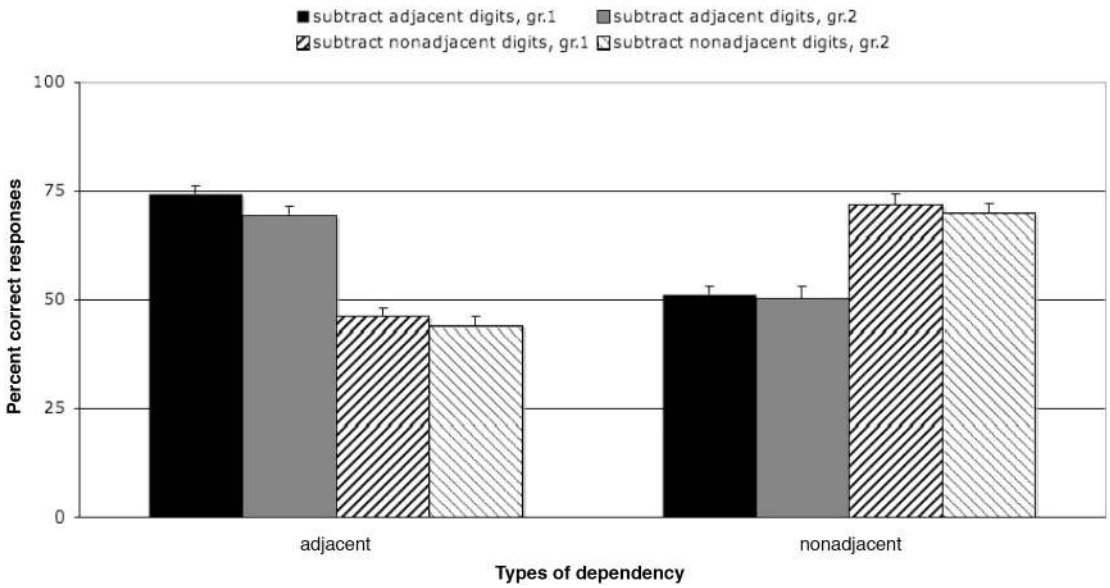

Figure 2. Mean percentage of correct responses (and standard error of the mean) for items evaluating the learning of adjacent and nonadjacent dependencies as a function of the task performed during the study phase in Experiment 1. Group 1 (gr.1) was presented with the adjacent pair *64* and the nonadjacent pair *3X7*, and Group 2 (gr.2) was presented with the adjacent pair *37* and the nonadjacent pair *6X4*.

Only the Type of Subtraction × Dependency Type interaction was significant, $F(1, 188) = 141.75$, $p < .01$, $MSE = 2.37$, $\eta_p^2 = .43$ (where η_p^2 represents partial eta squared). This interaction resulted from the fact that the type of dependencies learned varied as a function of the study task. Planned analyses showed that participants who subtracted adjacent digits during the study phase performed better on items evaluating adjacent dependency learning than on items evaluating nonadjacent dependency learning, $F(1, 94) = 50.70$, $p < .01$, $MSE = 2.66$, $\eta_p^2 = .35$. Analyses conducted with t tests indicated that the performance of such participants was above chance level for items evaluating the learning of adjacent dependencies ($M = 71.74$, $SE = 1.61$), $t(95) = 10.85$, $p < .01$, but did not differ from chance for items evaluating the learning of nonadjacent dependencies ($M = 50.78$, $SE = 1.69$), $t(95) = 0.36$, $p = .72$. By contrast, participants who subtracted nonadjacent digits performed better on items evaluating nonadjacent dependency learning than on items evaluating adjacent dependency learning, $F(1, 94) = 98.52$, $p < .01$, $MSE = 2.07$, $\eta_p^2 = .51$. These participants performed above chance level for items evaluating the learning of nonadjacent dependencies ($M = 70.83$, $SE = 1.74$), $t(95) = 10.28$, $p < .01$. Somewhat surprisingly, the performance on items evaluating the learning of adjacent dependencies ($M = 45.05$, $SE = 1.48$) was slightly, but significantly, below chance level, $t(95) = -2.57$, $p = .012$.

A final analysis compared the level of performance for adjacent and nonadjacent dependencies when participants' processing was triggered by the task instructions. The scores were 71.74 and 70.83, respectively, which did not differ significantly, $t(190) = 0.31$, $p = .75$. This means that when participants were asked to process two nonadjacent items, they learned their relations as well as they learned the relations between two adjacent items when they were asked to process adjacent items.

Discussion

Experiment 1 compared the learning of adjacent and nonadjacent dependencies by participants who performed a task requiring either the processing of adjacent elements or the processing of nonadjacent elements. Participants who focused on adjacent elements learned adjacent dependencies but did not learn nonadjacent dependencies. As shown in Table 1, this result was expected irrespective of whether contiguity and/or attention is construed as the causal factor. The crucial results concerned the performance of the participants who focused on nonadjacent dependencies. First, it appears that these participants learned nonadjacent dependencies. It is noteworthy that the recognition score of these participants was not significantly lower than the recognition score on adjacent dependencies reached by the participants who focused on adjacent dependencies. As illustrated in Table 1, only a model positing that attention is a necessary and sufficient condition for learning made this prediction. Second, there was no evidence for adjacent dependencies learning when the participants in the nonadjacent dependencies group were not in the attentional focus. This learning failure occurred despite the fact that (a) adjacent dependencies (e.g., *64* for half of the participants) were displayed twice within each problem (and hence 40 times on the whole study phase) and (b) the target searching task ensured that each digit of the sequences was individually processed. This result runs counter to any models focusing on event contiguity, whether contiguity is considered in isolation or in conjunction with attention.

Thus, overall, these results support the idea that the type of dependencies that are learned crucially hinges on the elements on which participants have focused during the study phase to meet the task demand. There was no evidence that the objective adjacency

of the events in the display played any role in and of itself when attentional factors were controlled.

Experiment 2

Experiment 2 explored whether the prior results obtained by the participants asked to process nonadjacent events could be replicated in a still more stringent condition. In summary, each problem used in Experiment 1 included two occurrences of the nonadjacent pairs (e.g., *3X7*). However, the task requirement ensured the attentional processing of only one of the two occurrences. For instance, in our example, the target was displayed only once between 3 and 7 within a problem. One may wonder what role, if any, was served by the *3X7* sequence in which *X* was not the target.

In Experiment 2, the problems were the same as those of Experiment 1, except that the nonadjacent dependency occurred only once per problem instead of twice. In this case, the target always served as the intervening event. To prevent any frequency bias in the sequence, the individual digits forming the nonadjacent dependency were nevertheless displayed twice (as any other digits in the 0–9 range), but on their second occurrence, they were displayed in random locations in the sequence.

The consequences of this change on the statistical properties of the materials are straightforward. Indeed, the change decreased (in fact halved) the frequency of the nonadjacent relations. Moreover, the decoupled presentation of the elements composing the dependency decreased their contingency, and contingency has been shown to be crucial for associative learning (e.g., Perruchet & Peereman, 2004). For instance, each of the digits composing the relevant association may become associated with other digits when they are displayed in isolation, hence generating competition and interference.

However, an attentional theory of associative learning does not predict a difference between Experiments 1 and 2. Indeed, the exposure to the A*X*C sequence when *X* is not the target (in Experiment 1) should have no positive effect on the learning of the A_C relationships, and likewise, the exposure to the decoupled A and C components (in Experiment 2) should have no detrimental effect, given that, in each case, the joint processing of the A and C elements was not required by the task instructions.

Method

Participants. A total of 96 undergraduate students from the Université Paris Descartes participated in the experiment in partial fulfillment of a course requirement. All participants were native French speakers. Participants were randomly assigned to one of the two digit combination experimental groups ($n = 48$).

Materials and procedure. The stimuli and the procedure of Experiment 2, including digit combination, were identical to those of Experiment 1, with the following exceptions: First, all of the participants received the instructions to process nonadjacent digits. Second, in each problem, the A*X*C sequence occurred only once, with *X* matching the target in each case. The A and C digits were also displayed once each as decoupled events. Their location in the sequence was randomly determined, with a different randomization for each problem and each participant.

Figure 1c shows three problems, with the adjacent pair *64,* which occurred twice; the nonadjacent pair *3X7,* which occurred once; and the digits forming the nonadjacent pair (3 and 7) displayed in random locations in the sequence, to prevent any frequency bias.

Results

Figure 3 shows the mean percentage of correct responses on the recognition test for items involving adjacent and nonadjacent dependencies. An ANOVA was performed on these data, with digit combination (*64* and *3X7* vs. *37* and *6X4*) as a between-subjects variable and type of dependency (adjacent vs. nonadja-

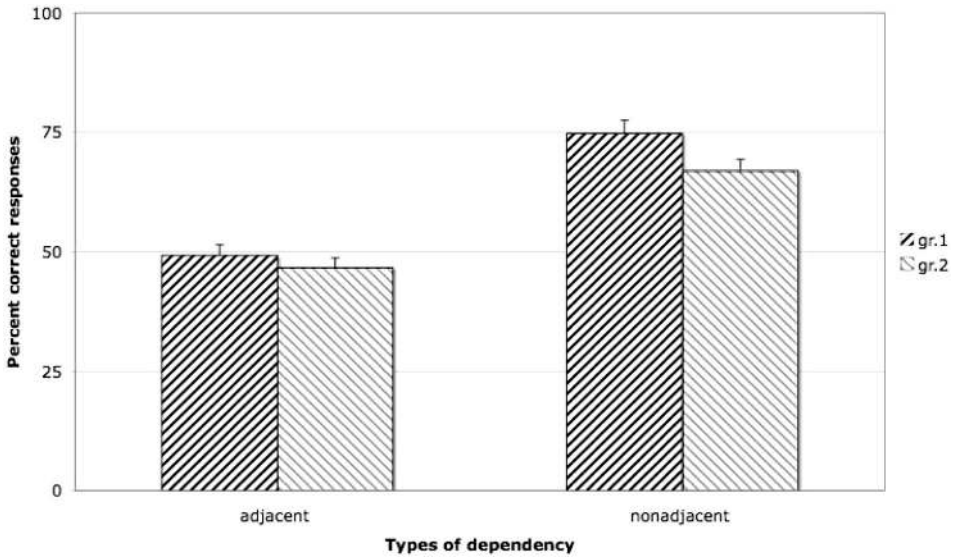

Figure 3. Mean percentage of correct responses (and standard error of the mean) on the recognition test for items evaluating the learning of adjacent and nonadjacent dependencies for the two groups of participants in Experiment 2. Group 1 (gr.1) was presented with the adjacent pair *64* and the nonadjacent pair *3X7*, and Group 2 (gr.2) was presented with the adjacent pair *37* and the nonadjacent pair *6X4*.

cent) as a within-subject variable. The type of dependency had a significant effect, $F(1, 94) = 54.44$, $p < .01$, $MSE = 2.96$, $\eta_p^2 = .37$, with correct responses occurring more frequently on the items involving nonadjacent dependencies ($M = 70.83$, $SE = 1.90$) than on the items involving adjacent dependencies ($M = 47.92$, $SE = 1.51$). The digit combination variable was marginally significant, $F(1, 94) = 3.57$, $p = .062$, $MSE = 2.34$, $\eta_p^2 = .04$, with correct responses occurring more frequently for the digit combination *64* and *3X7* ($M = 61.98$, $SE = 3.56$) than for the digit combination *37* and *6X4* ($M = 56.77$, $SE = 3.16$). There was no Digit Combination × Type of Dependency interaction ($F < 0.70$). Performance was above chance for nonadjacent dependencies, $t(95) = 9.36$, $p < .01$, whereas performance did not differ significantly from chance level for the adjacent dependencies, $t(95) = -1.07$, $p = .29$.

Discussion

Experiment 2 successfully confirmed the main conclusion of Experiment 1. First, participants who performed a task leading them to focus on nonadjacent digits learned nonadjacent dependencies. This result was obtained despite the fact that the nonadjacent dependency occurred only once in each problem and despite the decoupled presentation of the digits composing this dependency elsewhere in the sequences. It is worth noting that performances were very close to the corresponding results of Experiment 1,[3] in which the nonadjacent dependency occurred twice. This indicates that whether the A and C digits are displayed in their A*X*C format is inconsequential whenever *X* is not a target. Second, participants failed to learn adjacent dependencies. Taken together, these results are all the more striking, as nonadjacent dependencies were presented half as frequently as adjacent dependencies. The results strongly strengthen the view that the crucial factor in associative learning is the attentional processing of the stimuli during the study phase and not the objective contiguity of the events.

Experiment 3

Do the results reported in Experiments 1 and 2 persist over a transfer procedure? In the prior experiments, the intervening digits in the nonadjacent dependencies used during the test phase were the same as those seen during the study phase. This procedure has been used in prior studies (e.g., Gómez, 2002) and is commonly considered as providing a valid measure of nonadjacent dependency learning. However, as noted above, we defined nonadjacent dependency learning as the learning of the relation between two events, A and C, irrespective of the intervening events (*X*). The test used up until now does not fit exactly with this definition, because it is possible that performance improvement is conditioned on the use of the specific intervening events that were seen during the study phase. For instance, the possibility exists that participants simply memorized the A*X*C strings of digits as a whole, without extracting the A_C relations. To rule out this hypothesis, a transfer situation is needed, in which the *X* events seen in the test phase differ from the *X* events seen in the study phase. Prior studies have shown that such a change may have dramatic effects on performance. For instance, in Peña et al. (2002), the conditions promoting learning when the same intervening events were used in the study and test phases turned out to be no longer efficient when a transfer test was involved (but see Perruchet et al., 2004).

In Experiment 3, we explored the extent to which participants whose task leads them to focus on nonadjacent elements truly learned the association between disjoint elements, irrespective of the intervening events. The study task was the same as in Experiment 2, except that the target of the problems given to a participant could take only three values instead of six; hence, the *X* events seen in the test phase could differ from the *X* events seen in the study phase. In the recognition test, the digit intervening between two nonadjacent digits of a legal three-digit string was one of the three digits used as the target in the study phase for half of the test items (familiar test items). In the other half, one of the other three digits that were not used as target in the study phase (transfer test items) served this role. If participants have learned the association between nonadjacent digits irrespective of the digit embedded between these two nonadjacent digits, their performance should remain above chance level for transfer items. By contrast, if participants have simply memorized A*X*C strings of digits, their performance should be significantly above chance level for the familiar items (as observed in Experiments 1 and 2) but should not differ from chance for transfer items.

Method

Participants. A total of 160 undergraduate students from the Université Paris Descartes participated in the experiment in partial fulfillment of a course requirement. All participants were native French speakers. Participants were randomly assigned to one of the four Target × Digit Combination experimental groups ($n = 40$).

Materials. The problems used in the study phase of Experiment 3 were identical to the problems used in Experiment 1, with the exception that the target value of the problems given to a participant could take only three values. The target was randomly selected among the digits 0, 2, and 9 for half of the participants and among the digits 1, 5, and 8 for the other half. The target occurred once between the two pairs of digits forming the nonadjacent dependency and once in another, randomly determined, position. Different randomizations were used for each participant, leading to the construction of 160 different sets of 20 problems. Each of the 160 sets was printed on one A4 sheet of paper.

The recognition test used in Experiment 3 was the same as that used in Experiments 1 and 2, with one exception that enabled the creation of two categories of items. The digit *X* was selected randomly from among the three digits used as target digits in the study phase for one half of the test items involving adjacent dependencies and for one half of the test items involving nonadjacent dependencies (familiar items). The digit *X* was selected randomly among the three digits that were not used as target digits in the study for the other half of the test items evaluating the learning of adjacent dependencies and for the other half of the test items evaluating the learning of nonadjacent dependencies (transfer items). As in Experiments 1 and 2, the same digit *X* was used

[3] An ANOVA was performed on the correct responses, with experiment (Experiment 1—in which participants asked for the processing of nonadjacent digits only—vs. Experiment 2) and digit combination (*64* and *3X7* vs. *37* and *6X4*) as between-subjects variables and type of dependency (adjacent vs. nonadjacent) as a within-subject variable. The results indicated no main effect of experiment and no significant interactions for any of the four interactions involving this variable (all Fs < 0.59).

in the legal string and in the illegal string of a test pair (e.g., 640 – 604; 387 – 378). For each participant, the 16 pairs of three-digit strings were printed on one A4 sheet of paper.

Procedure. The procedure used in Experiment 3 was the same as that used in Experiments 1 and 2.

Results

Figure 4 shows the mean percentage of correctly selected test items evaluating the learning of adjacent dependencies, the learning of nonadjacent dependencies evaluated with the familiar test items, and the learning of nonadjacent dependencies evaluated with the transfer test items.

An initial analysis contrasted the learning of adjacent and nonadjacent dependencies, as in the prior experiments. The ANOVA included digit combination (*64* and *3X7* vs. *37* and *6X4*) and target (sample 0, 2, 9 vs. sample 1, 5, 8) as between-subjects variables and type of dependency (adjacent vs. nonadjacent) as a within-subject variable. The number of correct responses differed only as a function of the type of dependency, $F(1, 156) = 102.47$, $p < .01$, $MSE = 2.27$, $\eta_p^2 = .40$, with better performance on items involving nonadjacent dependencies ($M = 70.70$, $SE = 1.30$) than on items involving adjacent dependencies ($M = 49.38$, $SE = 0.93$). Performance was above chance level for the items involving nonadjacent dependencies, $t(159) = 12.98$, $p < .01$, whereas performance did not differ from chance level for the items involving adjacent dependencies, $t(159) = 0.52$, $p = .60$.

A second analysis examined the effect of transfer on nonadjacent dependencies. This ANOVA included digit combination (*3X7* vs. *6X4*) and target (sample 0, 2, 9 vs. sample 1, 5, 8) as between-subjects variables and type of intervening digit (familiar vs. transfer) as a within-subject variable. Only the type of intervening digit was significant, $F(1, 156) = 18.01$, $p < .01$, $MSE = 0.83$, $\eta_p^2 = .10$, with correct responses occurring more frequently on familiar items than on transfer items. Notably, however, performance was above chance level for both familiar items ($M = 76.09$, $SE = 1.46$), $t(159) = 13.75$, $p < .01$, and transfer items ($M = 65.31$, $SE = 1.87$), $t(159) = 7.09$, $p < .01$.

Discussion

First, Experiment 3 again replicated the results observed in Experiments 1 and 2, showing that participants who were asked to perform a task that leads them to focus on nonadjacent digits learned nonadjacent dependencies but did not learn adjacent dependencies. Second, this experiment showed that performance in the test of nonadjacent dependencies remained above chance level when the digit intervening between two nonadjacent digits of a legal three-digit string was not a digit used as a target in the study phase. Even though performance was poorer with these unfamiliar items than with familiar items, this finding provides especially compelling evidence that genuine learning of the A_C relationships occurred in this situation.

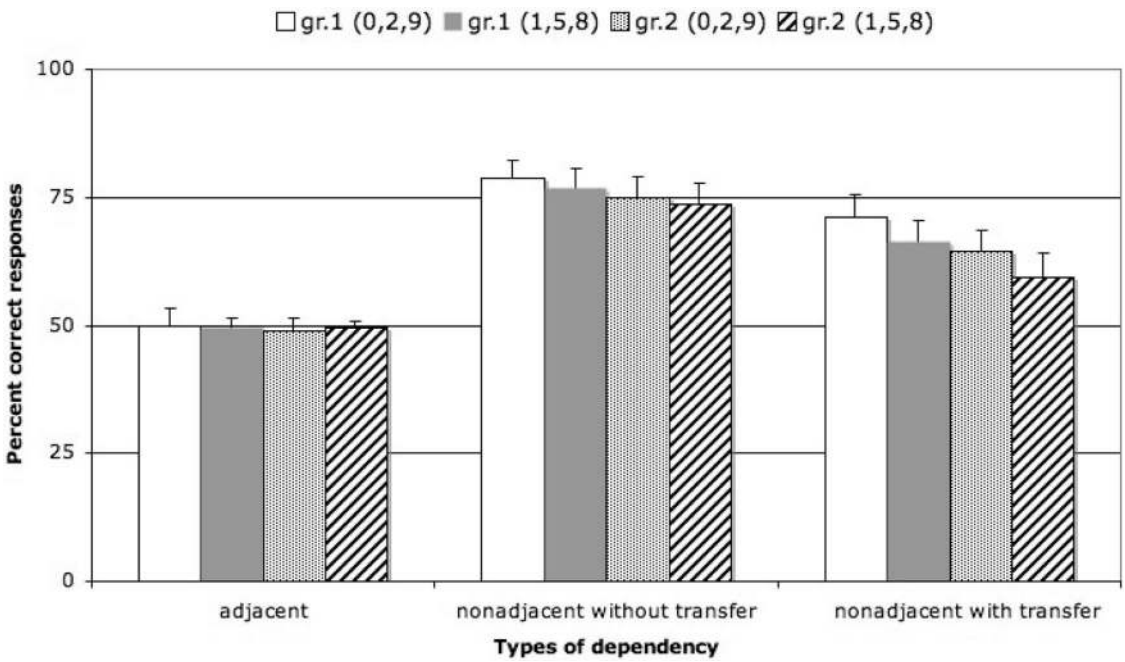

Figure 4. Mean percentage of correct responses (and standard error of the mean) as a function of the type of items used to assess the learning of adjacent and nonadjacent dependencies for the four groups of participants in Experiment 3. Group 1 (0, 2, 9) was presented with the adjacent pair *64*, the nonadjacent pair *3X7*, and the target digit 0, 2, or 9; Group 1 (1, 5, 8) was presented with the adjacent pair *64*, the nonadjacent pair *3X7*, and the target digit 1, 5, or 8; Group 2 (0, 2, 9) was presented with the adjacent pair *37*, the nonadjacent pair *6X4*, and the target digit 0, 2, or 9; and Group 2 (1, 5, 8) was presented with the adjacent pair *37*, the nonadjacent pair *6X4*, and the target digit 1, 5, or 8. gr.1 = Group 1; gr.2 = Group 2.

Experiments 4a and 4b

Taken together, Experiments 1 through 3 show that when participants are faced with stimuli embedding relations between both adjacent and nonadjacent elements, they learn exclusively the type of relation they had to actively process in order to meet the task demands, irrespective of the spatial contiguity of the paired elements. The learning score for nonadjacent dependencies when the task required the processing of adjacent digits (Experiment 1) and, more crucially, the learning score for the adjacent dependencies when the task required the processing of nonadjacent digits (Experiments 1, 2, and 3) were never above chance level.

A postulate underpinning these experiments was that all individual items are processed. If some items were not actually processed, the observed failure of participants to learn their relations would be trivial. Up until now, we have taken for granted that the processing of all individual items occurred, because participants had to scan the digit strings in order to detect the target. However, this postulate can be challenged.[4]

First, it is possible that participants quickly discover that each problem includes two, and only two, targets. Once participants have processed the second target, they might stop scanning the digit string to move on to the next one. Note that, although this possibility can hardly be ruled out, its consequences on our line of argumentation would be limited. If participants had decided to stop scanning the sequences after the second target, they would have been exposed to only a part of the materials. However, this remains substantial (namely, two thirds of each digit string, on average, for the digit strings occurring after the point in the session where participants noticed that there were two targets per problem), and it provides an unbiased sample of the whole strings (notably with regard to the type of displayed dependency).

Another, a priori more devastating possibility, is that the target pops out from the digit strings, without any of the other digits actually being scanned. This possibility is suggested by experiments in visual-search tasks: Typically, participants are asked to search for a line segment of which the orientation differs from the common orientation of the surrounding distractors. In this situation, the detection time is independent of the number of distractors, a result suggesting that the display was processed at a preattentive level (Treisman & Gelade, 1980).

At least two reasons militate against this alternative explanation. First, it has been suggested that the early conclusion according to which the distractors in a standard visual-search task were processed at a preattentive level may have been an overstatement, with a more appropriate conclusion being that they received relatively little attention (e.g., Joseph, Chun, & Nakayama, 1997). Second, our procedure differs strikingly from the standard situations used to investigate the pop-out phenomenon. In particular, in our task, the targets differed from one sequence to the next, and the same digits were used alternatively as target and distractor across trials. This makes the possibility that the nontarget digits were processed at a preattentive level somewhat unlikely (e.g. Ahissar & Hochstein, 1996). However, it remains possible that the digits that were not involved in the subtraction task were processed very cursorily and, crucially, faster than the digits that had to be subtracted from one another. If the amount of time allocated to two task-irrelevant digits was shorter than that allocated to the joint processing of two task-relevant digits, then the relevant factor could be the overall duration of processing rather than the fact of processing jointly, instead of separately, the two events at hand.

In two additional experiments, we investigated whether our results could be replicated using a task that ensures a more extensive processing of the items not involved in the subtraction task. Experiments 4a and 4b presented the same problems as in Experiment 3, but the sequence of digits appeared progressively, from left to right, instead of being displayed all at once. In Experiment 4a, the occurrence of each digit was self-paced by participants. In Experiment 4b, the digits appeared sequentially on the computer screen, with a fixed 500-ms interval between each digit. Other changes were minor and were a consequence of the necessary shift from group pencil-and-paper experiments to individual computer-controlled experiments. Notably, participants were now asked to report orally to the experimenter, on each target occurrence, whether the difference between the two digits surrounding the target equaled +3 or −3.

Experiment 4a

Method

Participants. A total of 24 undergraduate students from the Université Paris Descartes participated in the experiment in partial fulfillment of a course requirement. All participants were native French speakers. Participants were randomly assigned to one of the four Target × Digit Combination experimental groups ($n = 6$).

Materials. The problems and the test items were the same as in Experiment 3. Because the number of participants per group was now 6, a total of 6 problems and 6 tests were randomly drawn among the 40 that were generated for each group in Experiment 3.

The experiment was carried out with PsyScope for Mac OS X (J. D. Cohen, MacWhinney, Flatt, & Provost, 1993) on a PowerBook G4 (Apple Inc., Cupertino, CA) 15.4-in. (39.1-cm) widescreen display. The target digits, the sequences of 20 digits, and the test strings were printed in Chicago font, size 26. The distance between the left border of the screen and the first digit of the sequence of 20 digits, as well as the distance between two successive digits of the sequence, was 1.3 cm.

Procedure. The participants were tested individually in a sound-attenuated room. The experiment consisted of a study phase and a test phase. Before the study phase, participants were informed that at the beginning of each problem, a digit displayed at the top left of the screen indicated the value of the target for the problem; they were instructed to press the space bar to reveal each digit of a sequence of digits that would appear progressively from left to right, in the middle of the screen. The participants were told that when the digit displayed after a keystroke was revealed to be a target, they would have to calculate whether the difference between the two digits surrounding the target equaled +3 or −3. Participants were asked to provide their answer orally (yes or no; note that before responding, participants had to press the key once more to reveal the digit situated on the right of the target). Finally, as in prior experiments, participants were informed that the target digit could occur more than once within a sequence. At the end of each problem, a message invited participants to press the space bar to proceed to the next problem.

[4] We thank Brian Scholl and two anonymous reviewers of a draft of this article for drawing our attention to this problem.

Once the 20 problems of the study phase were completed, participants performed the recognition test. They were instructed to select which of two strings of three digits was among the digit sequences viewed in the study phase of the experiment. As in prior experiments, participants were told that they had to make a choice for each of the 16 pairs, even if they sometimes felt that they were guessing. They were to press the Q key of the (French) AZERTY keyboard if they felt that the correct answer was the string on the left and the M key if they felt that the correct answer was the string on the right. The screen was cleared immediately after the participant's keystroke, and there was a 2-s delay before the presentation of the following pair. The two strings of digits, separated with a slash mark, were displayed in the center of the screen.

Results

Figure 5 shows the mean percentage of correct responses on the recognition test for items with adjacent dependencies, for items with nonadjacent dependencies evaluated with the familiar test items, and for items with nonadjacent dependencies evaluated with the transfer test items. An initial ANOVA was performed, with digit combination (*64* and *3X7* vs. *37* and *6X4*) and target (sample 0, 2, 9 vs. sample 1, 5, 8) as between-subjects variables and type of dependency (adjacent vs. nonadjacent) as a within-subject variable. The number of correct responses differed only as a function of the type of dependency, $F(1, 20) = 31.50$, $p < .01$, $MSE = 1.06$, $\eta_p^2 = .61$, with better performance on items involving nonadjacent dependencies ($M = 68.23$, $SE = 3.53$) than on items involving adjacent dependencies ($M = 47.40$, $SE = 3.01$). Performance was above chance level for the items involving nonadjacent dependencies, $t(23) = 5.17$, $p < .01$, whereas performance did not differ from chance level for the items involving adjacent dependencies, $t(23) = 0.86$, $p = .40$.

A second analysis, with digit combination (*3X7* vs. *6X4*) and target (sample 0, 2, 9 vs. sample 1, 5, 8) as between-subjects variables and type of intervening digit (familiar vs. transfer) as a within-subject variable, examined the effect of transfer on nonadjacent dependencies. As in Experiment 3, performance was better for familiar items than for transfer items, although the effect of the type of intervening digit failed to reach significance, $F(1, 20) = 3.54$, $p = .07$, $MSE = 0.71$, $\eta_p^2 = .15$. Performance was again above chance level for both familiar items ($M = 73.96$, $SE =$

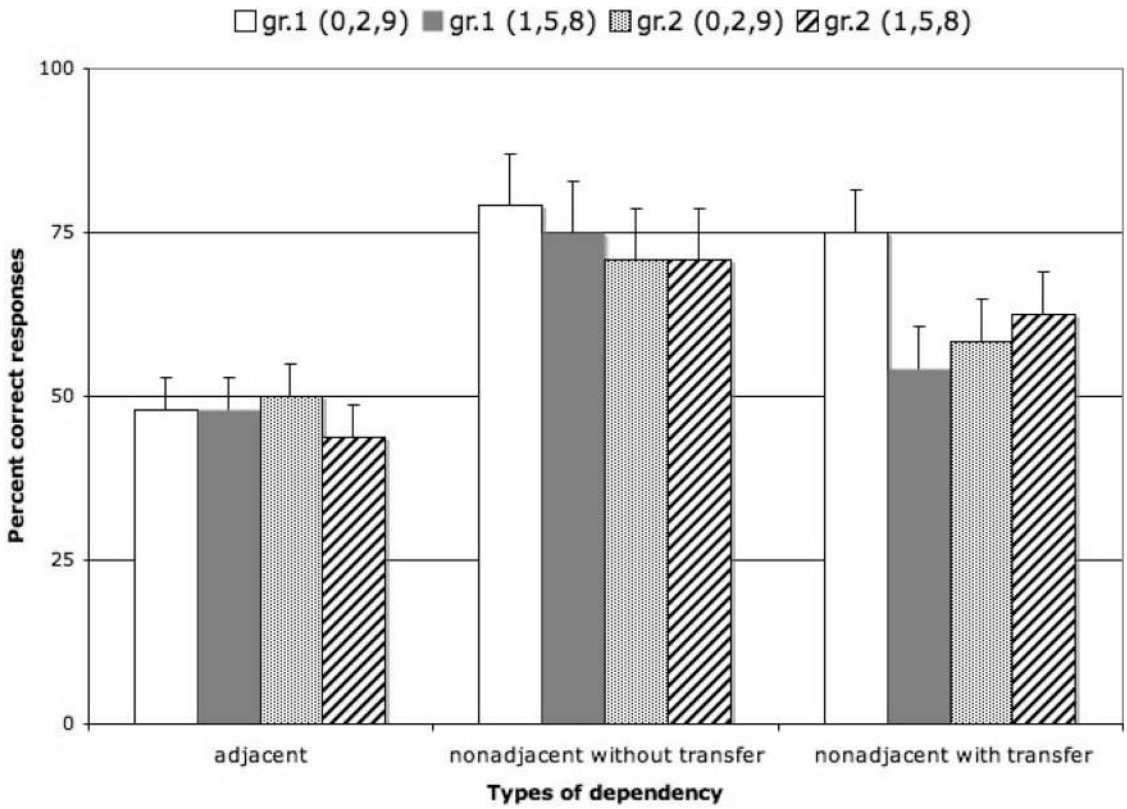

Figure 5. Mean percentage of correct responses (and standard error of the mean) as a function of the type of items used to assess the learning of adjacent and nonadjacent dependencies for the four groups of participants in Experiment 4a. Group 1 (0, 2, 9) was presented with the adjacent pair *64*, the nonadjacent pair *3X7*, and the target digit 0, 2, or 9; Group 1 (1, 5, 8) was presented with the adjacent pair *64*, the nonadjacent pair *3X7*, and the target digit 1, 5, or 8; Group 2 (0, 2, 9) was presented with the adjacent pair *37*, the nonadjacent pair *6X4*, and the target digit 0, 2, or 9; and Group 2 (1, 5, 8) was presented with the adjacent pair *37*, the nonadjacent pair *6X4*, and the target digit 1, 5, or 8. gr.1 = Group 1; gr.2 = Group 2.

2.06), $t(23) = 5.82$, $p < .01$, and transfer items ($M = 62.50$, $SE = 2.50$), $t(23) = 2.50$, $p = .02$.

Experiment 4b

Method

Participants. A total of 24 undergraduate students from the Université Paris Descartes participated in the experiment in partial fulfillment of a course requirement. All participants were native French speakers. Participants were randomly assigned to one of the four Target × Digit Combination experimental groups ($n = 6$).

Materials. The materials were the same as in Experiment 4a.

Procedure. The procedure was the same as in Experiment 4a, with one exception concerning the manner in which the sequences of digits were displayed on the computer screen during the study phase. Instead of being self-paced, the timing of the digit presentation was predetermined by the program. The target digit remained on the screen for 2,000 ms before the occurrence of the first digit of the 20-digit sequence, and a 500-ms interval separated the display of two successive digits of the sequence.

Results

Figure 6 shows the mean percentage of correct responses on the recognition test for items with adjacent dependencies, for items with nonadjacent dependencies evaluated with the familiar test items, and for items with nonadjacent dependencies evaluated with the transfer test items. An initial ANOVA was performed, with digit combination (*64* and *3X7* vs. *37* and *6X4*) and target (sample 0, 2, 9 vs. sample 1, 5, 8) as between-subjects variables and type of dependency (adjacent vs. nonadjacent) as a within-subject variable. The number of correct responses differed only as a function of the type of dependency, $F(1, 20) = 20.79$, $p < .01$, $MSE = 2.31$, $\eta_p^2 = .51$, with better performance on items with nonadjacent dependencies ($M = 74.48$, $SE = 2.86$) than on items with adjacent dependencies ($M = 49.48$, $SE = 4.02$). Performance was above chance level for the items with nonadjacent dependencies, $t(23) = 8.55$, $p < .01$, whereas performance did not differ from chance level for the items with adjacent dependencies, $t(23) = 0.13$, $p = .90$.

A second analysis included digit combination (*3X7* vs. *6X4*) and target (sample 0, 2, 9 vs. sample 1, 5, 8) as between-subjects variables and type of intervening digit (familiar vs. transfer) as a within-subject variable. As in Experiment 4a, the effect of the type of intervening digit failed to reach significance, $F(1, 20) = 3.71$, $p = .07$, $MSE = 0.68$, $\eta_p^2 = .16$, but in contrast to prior experiments, performance was descriptively better for transfer items than for familiar items. Notably, performance was above chance level for both familiar items ($M = 71.88$, $SE = 4.59$), $t(23) = 4.76$, $p <$

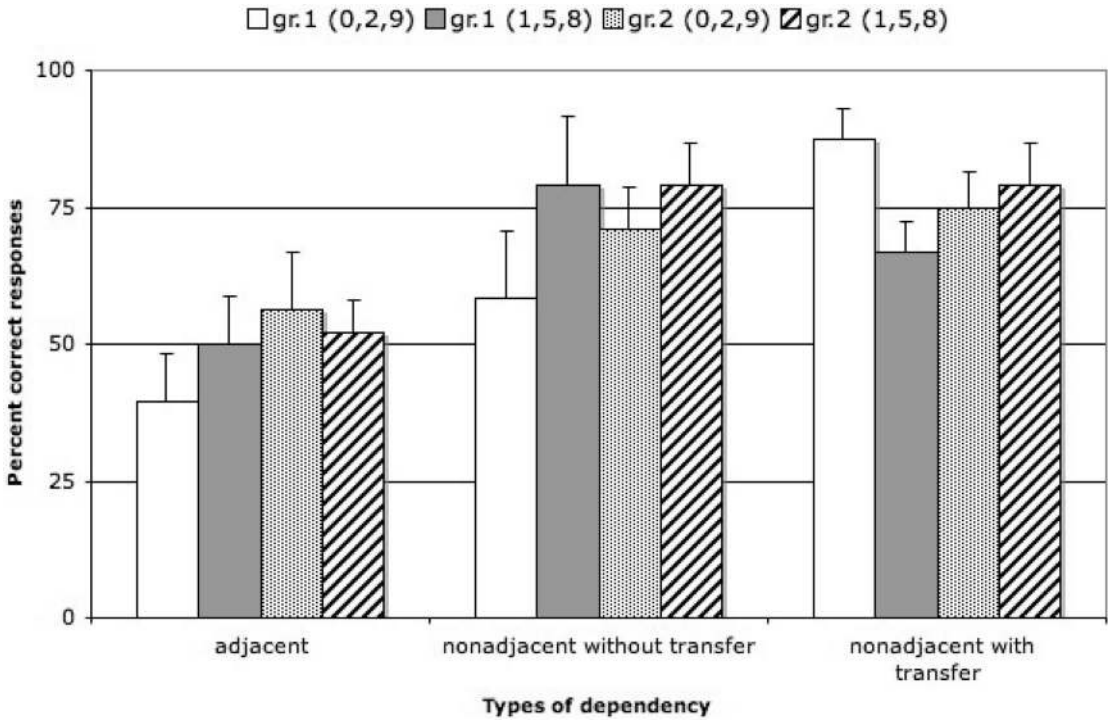

Figure 6. Mean percentage of correct responses (and standard error of the mean) as a function of the type of items used to assess the learning of adjacent and nonadjacent dependencies for the four groups of participants in Experiment 4b. Group 1 (0, 2, 9) was presented with the adjacent pair *64*, the nonadjacent pair *3X7*, and the target digit 0, 2, or 9; Group 1 (1, 5, 8) was presented with the adjacent pair *64*, the nonadjacent pair *3X7*, and the target digit 1, 5, or 8; Group 2 (0, 2, 9) was presented with the adjacent pair *37*, the nonadjacent pair *6X4*, and the target digit 0, 2, or 9; and Group 2 (1, 5, 8) was presented with the adjacent pair *37*, the nonadjacent pair *6X4*, and the target digit 1, 5, or 8. gr.1 = Group 1; gr.2 = Group 2.

.01, and transfer items ($M = 77.08$, $SE = 3.96$), $t(23) = 6.84$, $p < .01$, as in prior experiments.

Discussion

Experiments 4a and 4b confirm the conclusion of the first three experiments regarding the possibility of learning nonadjacent dependencies when the task demand requires joint attention to the relevant elements. As in Experiment 3, the recognition scores for the test items embedding the nonadjacent dependency were significantly above chance, even when the intermediary digits were not seen in the training phase. However, the difference in performance according to whether the intermediary digits of the test items were seen or not seen during training was somewhat inconsistent across experiments. Although this difference never reached significance, Experiment 4a tended to replicate the transfer decrement observed in Experiment 3, whereas the trend was reversed in Experiment 4b. Given that the only difference between Experiments 4a and 4b concerned the fact that the successive appearance of the digits was self-paced and computer-paced, respectively, this discrepancy is hardly interpretable. It is likely that the apparent reversal of the transfer decrement in Experiment 4b—namely, descriptively better performance in the transfer than in the familiar situation—should be attributed to noise in the data. Note that other studies in the implicit learning domain have also reported discordant results regarding the transfer decrement effect.

Considering the learning of an invariant as an example, we note that the use of transfer in place of familiar test stimuli in prior research usually has yielded no significant effect (e.g., McGeorge & Burton, 1990). However, both poorer (Stadler, Warren, & Lesch, 2000) and better (Newell & Bright, 2002, p. 1110) performance in transfer, as compared with familiar, situations occasionally has been reported.

The procedural changes introduced in Experiments 4a and 4b were devised to increase the amount of attentional processing devoted to the individual nontarget digits, which included those forming the adjacent dependency. These changes were intended to test our hypothesis that even adjacent relations are not learned when the task does not require the joint processing of the relevant elements, even though the duration of exposure to individual elements is controlled. The results were clear-cut: There was no evidence of learning in either experiment, even at a descriptive level. It is worth stressing, however, that we controlled the objective duration of the visual displays and not the amount of time actually allocated by the participants to the active processing of those displays. It remains possible that processing duration varies despite the equalized duration of objective exposure. A posteriori control over the actual processing engaged by the participants should be carried out in future studies, perhaps with eye-tracking measures. If eye-tracking measures reveal a sizable disproportion in the duration of processing allocated to the different items, then ruling out the hypothesis that the amount of processing time is a causal factor would certainly require a more radical change in the experimental procedure.

General Discussion

A Summary of Results

Participants who were asked to process adjacent elements learned adjacent dependencies but did not learn nonadjacent dependencies. As shown in Table 1, this pattern of results does not discriminate among the predictions of different theories. Indeed, this was expected irrespective of whether contiguity or attention is construed as the causal factor. The performance of the participants who processed nonadjacent digits to fulfill the task requirements was much more informative. Two main results emerged.

First, there was no evidence for adjacent dependency learning in the participants processing nonadjacent digits. This learning failure occurred despite the fact that (a) in each of the five reported experiments, adjacent dependencies were presented 40 times over the course of the entire study phase and (b) the target searching task ensured the individual processing of each digit of the sequences. Although processing may have been somewhat limited in the first three experiments, in which the digit strings were displayed as a whole, the digit-by-digit presentation of the problems used for Experiments 4a and 4b prompted a more extensive processing of each item.

Second, participants processing nonadjacent dependencies for meeting the task demand learned them in five independent experiments. It is noteworthy that the participants' recognition score was not significantly lower than the recognition score for adjacent dependencies reached by the participants who focused on adjacent dependencies in Experiment 1. This remained true when nonadjacent dependencies were displayed twice less frequently than adjacent dependencies in Experiment 2. Experiments 3, 4a, and 4b, showed that performance was also above chance level in a transfer test, in which the intervening events were changed between the study phase and the test phase. Performance was lower when the intervening events were changed between the study phase and the test phase than when the same intervening events were used in both phases in Experiment 3, although Experiments 4a and 4b showed no significant difference (and contradictory trends) according to whether the intervening items in the test were old or new.

As shown in Table 1, this pattern of results is only consistent with the attention-based theory of associative learning outlined in the introduction, which posits that selective attention is a necessary and sufficient condition for learning to occur. In this discussion, we first elaborate further on this theory and how it may be articulated with the extant literature on learning. Next, we focus on the sense in which our study and the theoretical conclusions it supports concern implicit, rather than explicit, forms of learning. Finally, we outline some of the problems that arise when attempting to apply the theory to specific domains, such as language, music, and visual perception.

An Attentional Account of Adjacent and Nonadjacent Dependency Learning

The failure to observe learning of adjacent dependencies when the instructions drew participants' attention toward nonadjacent dependencies provides new support for the models of learning positing that attention is a necessary condition for learning to occur. The earlier accounts of conditioning (e.g., Guthrie, 1930) focused exclusively on the contiguity of the conditioned and unconditioned stimuli (CS and US, respectively). However, further studies made it obvious that the repeated exposure to a pair of contiguous events is not sufficient to trigger associative learning. The well-documented effect of blocking (Kamin, 1968) illustrates

this point. In the first phase of a blocking experiment, a CS, namely A, is followed by a US. In the second phase, A is always followed by the same US, but it is systematically presented with another cue, B. The phenomenon of blocking designates the fact that the relation between B and the US is not learned in those conditions, or at least is not learned to the same extent as if B had been associated with the US in standard conditions. In other words, the contiguity of B and the outcome is not sufficient for learning to occur. The phenomenon is important in and of itself but also because it provides a simple explanation for other, apparently more complex, phenomena. For instance, Rescorla (1968) showed that adding an unpaired US during training decreased conditioning, although the level of contiguity (i.e., the number of CS–US pairings) was kept constant. This result was initially conceived of as a demonstration that animals are able to assess the genuine contingency between events. However, a few years later, Rescorla and Wagner (1972) suggested that the phenomenon of blocking could account for the results of Rescorla (1968), because the unpaired occurrences of the US elicit conditioning to some background stimuli, which in turn, blocks the conditioning to the experimental CS.

Of special interest for our concern, the phenomenon of blocking has been accounted for by a framework based on attention (Mackintosh, 1975). The relation between B and the outcome is not processed attentionally, it was argued, because the predictive value of A captures all of the learner's available attention. Associative blocking, and a few other related phenomena, such as the CS preexposure effect (Lubow, 1989) and, more recently, highlighting (Kruschke, 2001), have led contemporary researchers to emphasize the crucial role of attention in associative learning and to elaborate various attention-based models (see Kruschke, 2003, 2005).

The need for learners' attentional involvement has been confirmed by a number of studies in other experimental arrangements as well. In recent research on implicit learning, for instance (see Perruchet, in press; Perruchet & Pacton, 2006; Shanks, 2005, for reviews), the principle of these studies consists of adding a concurrent secondary task during the training session, then observing whether performance improvement is equivalent to that observed in a single-task, standard procedure. With regard to artificial grammars, Dienes, Broadbent, and Berry (1991) demonstrated that the accuracy of grammaticality judgments was lowered when participants had to perform a concurrent random number generation task during the familiarization phase. The studies carried out on dynamic system control tasks provide a more contrasting picture. The theory surrounding the early studies on these tasks posited a distinction between two forms of learning: selective (i.e., with attention) and unselective (i.e., without attention; e.g., Berry & Broadbent, 1988), and some initial studies seemingly supported this view. However, subsequent studies (e.g., Green & Shanks, 1993) failed to replicate earlier results despite extensive attempts to do so and found that, as a rule, the secondary task impaired performance irrespective of the complexity of the task. A similar story happened with serial reaction time tasks, with a few initial studies (e.g., A. Cohen, Ivry, & Keele, 1990) claiming that nonattentional learning is efficient for the simplest forms of sequential dependencies. The secondary task typically used in this context is a tone-counting task, in which a high- or low-pitched tone sound is emitted after each trial, and participants are required to keep a running count of one of them (e.g., high-pitched tones). Performance in these conditions typically remains above chance. However, observing performance improvement under dual-task conditions does not imply the existence of a nonattentional form of learning, because the secondary task may not deplete the attentional resources completely. As claimed by Stadler (1995), "even when implicit serial learning is observed in conjunction with the tone-counting task, . . . it cannot be said that learning occurred without attentional capacity—the participants certainly devoted attention to the serial RT task" (p. 683). Closing their survey of the role of attention in implicit sequence learning, Hsiao and Reber (1998) concluded, "We view sequence learning as occurring in the background of the residual attention after the cost of the tone-counting task and the key-pressing task. If there is still sufficient attention available to the encoding of the sequence, learning will be successful; otherwise, failure will result" (p. 487; for other approaches that emphasize the role of attention, see Frensch et al., 1994; Jimenez & Mendez, 1999).

The role played by selective attention in acquisition processes also has been identified in, for instance, the literature on automatisms (e.g., Fisk & Schneider, 1984), covariation learning (Hoffman & Sebald, 2005), word segmentation (Toro, Sinnett, & Soto-Faraco, 2005), visual perception (e.g., Baker, Olson, & Behrmann, 2004; Turk-Browne et al., 2005), contextual cueing (Jiang & Chun, 2001; Jiang & Leung, 2005), and memory (e.g., Roediger, 1990). This contention holds even for the so-called implicit memory phenomena, in which performance does not involve the recollection of the initial episodes. There is now overwhelming evidence that attention to the material at the time of encoding is a necessary condition for the observation of improved performance in subsequent implicit memory tests, such as word completion and perceptual identification tasks (e.g., Crabb & Dark, 1999), reading tasks (MacDonald & MacLeod, 1998), and object decision tasks (Ganor-Stern, Seamon, & Carrasco, 1998).

As noted in the introduction, however, acknowledging the necessity of attention in learning does not provide, as such, an account for nonadjacent dependency learning. In all of the studies evoked above, the conditions of contiguity are met, and hence their results are compatible with the claim that associative learning requires both the contiguity of events and the learner's attention. To account for nonadjacent dependency learning, learners' attentional involvement needs to be construed as a sufficient condition. This proposal is consonant with the position taken by several authors, who have construed associative learning as an automatic process that associates all the components that are simultaneously present in the attentional focus (Frensch & Miner, 1994; Logan & Etherton, 1994; Perruchet & Vinter, 2002; Stadler, 1995; Treisman & Gelade, 1980). Our result that participants learned nonadjacent dependencies when asked to process them provides a striking confirmation of this view. The objective adjacency of the events in the display would play no role in itself when attentional factors are controlled. Indeed, when contiguity and attentional processing were pitted against each other, there was no residual role for contiguity. The relation between two nonadjacent events can be learned whenever there is some reason for processing the two events jointly.

Prior studies also showed the possibility of learning nonadjacent dependencies, but only when the to-be-learned material possessed specific properties, such as a high variability of the *X* events (e.g.,

Gómez, 2002), a high level of similarity between A and C events (e.g., Onnis et al., 2005), or the presence of pauses between the A*X*C sequences (Peña et al., 2002). On the basis of those results, one could argue that this form of learning requires quite specific conditions, which are qualitatively different from the conditions required to learn adjacent relations. We suggest, instead, as noted in the introduction, that such variables could be effective because they facilitate the attentional processing of the relevant events. Our demonstration that participants' performance may change drastically as a function of task instructions, while the objective properties of the material are kept constant, provides especially strong evidence for this interpretation of the earlier literature.

Attention and Implicitness

The fact that our procedure involved a manipulation of participants' attentional allocation through the instructions they received could suggest that we actually contrasted incidental and intentional learning. In this account, learning either adjacent or nonadjacent dependencies would be possible only when participants intentionally search for those regularities. It is worth stressing that this is not our intended conclusion. Although we have no concerns with the claim that such learning can result from an analytic, explicit search, we were interested in the possibility of learning in incidental conditions, in keeping with the current focus of most learning researchers. Was our procedure actually well-suited to meeting this objective?

Our task was nominally incidental because, although the task involved the joint attentional processing of two digits, the instructions required neither a search for the repeated occurrences of some pairs of digits nor the explicit memorization of those digits. Beyond the instructions, however, it is important to examine whether participants could have shifted themselves to a strategy of memorization. A reasonable assumption is that participants could have done so if they had some reason to believe that learning the digits would help them to perform the task demands, which primarily consisted of detecting the target among an array of distractors. In fact, this is not the case because the repeated patterns provided a quite unreliable cue for the presence of a target. One of the two repeated patterns was never associated with the target (e.g., the adjacent dependencies for participants dealing with nonadjacent digits). Moreover, even if one restricts the analysis to the repeated pattern with which participants dealt while performing the task, half of the occurrences of this pattern happened without the target and half of the targets occurred outside this pattern. As a consequence, a strategy consisting of searching for this pattern in order to guide target detection would lead to missing half of the targets and to having attention drawn toward an irrelevant part of the strings in half of the cases. All of these characteristics make it quite unlikely that participants were engaged in an explicit search of the statistical regularities of the digit sequences.

The discussion above makes it clear that learning was certainly implicit in the sense that it was incidental. It is worth adding that we do not claim that the resulting knowledge of the dependencies was implicit in the sense that it was unavailable to consciousness. Although the possibility that some form of knowledge may affect performance without being available to consciousness cannot be ruled out, several authors have emphasized the lack of empirical support for such a claim (for reviews, see, e.g., Dulany, 1997; Shanks & St. John, 1994). Irrespective of this possibility, restricting a priori the manifestation of implicit learning to some uninformed guessing, similar in some ways to the vague feeling elicited in subliminal perception, appears to be unproductively restrictive and to leave behind the most important expressions of implicit learning in real-world settings. We subscribe here to the view that implicit learning processes primarily shape conscious percepts and representations (e.g., Perruchet & Vinter, 2002; Perruchet, Vinter, & Gallego, 1997). In keeping with this position, we measured learning through a forced-choice recognition test, which is traditionally construed as an explicit test. Note that this measure can hardly be conceived of as a departure from the current practice in implicit learning research. For instance, all of the studies on word segmentation following the seminal articles of Saffran and collaborators (Saffran et al., 1996; e.g., Newport & Aslin, 2004; Onnis et al., 2005) have also used a forced-choice recognition test to assess learning. To our best knowledge, this procedure has never been construed as preventing the consensual embodiment of this literature within the context of implicit learning research (for a review, see Perruchet & Pacton, 2006).

Questions for Further Research

A theory of associative learning positing that joint attention on the relevant events is both a necessary and a sufficient condition for learning associations provides a forceful argument for extending the relevance of associative learning principles to complex domains, such as natural language, music, and object perception. Indeed, the theory allows us to account for the learning of remote dependencies, which occur routinely in those domains, while avoiding the apparently intractable problem of combinatorial explosion, given that the number of possible associations to be scrutinized grows exponentially with the number of authorized intervening events. The problem, in principle, is avoided here because potential associations are preselected by attentional mechanisms.

However, a gap remains to be bridged between theory and the potential domains of application. The main problem concerns the question of the features that are able to capture attention in each domain. Certainly, attention may be guided by explicit instructions, as in the experiments described above. However, this cannot be considered a general solution. At best, explicit instructions can play a role in supervised learning, but they cannot account for the much more frequent situations in which people learn from their own experience, without instructional guidance. Of course, a number of environmental features are known to be naturally prone to capturing attention. This is the case of certain acoustic properties of language or music, or visual features such as movement. However, for the explanation to be viable, those features should have a high chance of guiding attention toward the relevant events. The deep issue is, then, Why should those features be a priori relevant? Why, for instance, should the acoustic and prosodic cues that are most predisposed to capturing attention in language utterances provide a guide toward the relevant aspects of the language, given that those aspects are presumably syntactic or semantic in nature?

Although this issue stands out of the scope of this article, we submit that two general, and nonexclusive, responses may be put forth. The first solution consists of invoking innateness: Evolution would have made us sensitive to some specific features of our environment because of their relevance for adaptive purposes. Another solution,

which applies only to those aspects that are a product of human culture, with language as the main target, relies on the concept of reversed evolution. Instead of positing that the human species has evolved to deal with complex linguistic structures, it is possible to conceive that language has evolved to be learned by humans, given the general abilities they have developed in response to adaptive constraints preceding language appearance (Newport, 1990). Further research is needed to deepen these issues.

Conclusion

The detection of remote dependencies represents an apparent challenge for a framework based on associative or statistical mechanisms. Indeed, most of the literature on associative learning has focused on adjacent dependencies, whereas there is evidence that higher cognitive activities, such as language, music, and object perception, also exploit the existence of nonadjacent dependencies in their material. We have provided experimental support for an associative learning model positing attention as a necessary and sufficient condition, which accounts for both adjacent and nonadjacent dependency learning in the very same terms.

References

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research, 36,* 3487–3500.

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science, 15,* 460–466.

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit–explicit distinction. *British Journal of Psychology, 79,* 251–272.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16,* 451–459.

Cohen, A., Ivry, R., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 17–30.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25,* 257–271.

Cowan, N. (1991). Recurrent speech patterns as cues to the segmentation of multisyllabic sequences. *Acta Psychologica, 77,* 121–135.

Crabb, B. T., & Dark, V. (1999). Perceptual implicit memory requires attentional encoding. *Memory & Cognition, 27,* 267–275.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1119–1130.

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

Driver, J., & Baylis, G. (1998). Attention and visual object segmentation. In R. Parasuraman (Ed.), *The attentive brain* (pp. 299–325). Cambridge, MA: MIT Press.

Dulany, D. E. (1997). Consciousness in the explicit (deliberative) and implicit (evocative). In J. Cohen & J. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 179–211). Mahwah, NJ: Erlbaum.

Fisk, A. D., & Schneider, W. (1984). Memory as a function of attention, level of processing, and automatization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 181–197.

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, S. C. (1994). Effects of presentation rate and of individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory and Cognition, 22,* 95–110.

Ganor-Stern, D., Seamon, J. G., & Carrasco, M. (1998). The role of attention and study time in explicit and implicit memory for unfamiliar visual stimuli. *Memory & Cognition, 26,* 1187–1195.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436.

Gómez, R. L. (2006). Dynamically guided learning. In Y. Munakata & M. H. Johnson (Eds.), *Attention & Performance XXI: Processes of change in brain and cognitive development* (pp. 87–110). Oxford, England: Oxford University Press.

Green, R. E. A., & Shanks, D. R. (1993). On the existence of independent learning systems: An examination of some evidence. *Memory & Cognition, 21,* 304–317.

Guthrie, E. R. (1930). Conditioning as a principle of learning. *Psychological Review, 37,* 412–428.

Hoffmann, J., & Sebald, A. (2005). When obvious covariations are not even learned implicitly. *European Journal of Cognitive Psychology, 17,* 449–480.

Hsiao, A. T., & Reber, A. (1998). The role of attention in implicit sequence learning: Exploring the limits of the cognitive unconscious. In M. Stadler & P. Frensch (Eds.), *Handbook of implicit learning* (pp. 471–494). Thousand Oaks, CA: Sage.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *Quarterly Journal of Experimental Psychology, 54A,* 1105–1124.

Jiang, Y., & Leung, A.-W. (2005). Implicit learning of ignored visual context. *Psychonomic Bulletin & Review, 12,* 100–106.

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

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a 'preattentive' feature search task. *Nature, 387,* 805–807.

Jungé, J. A., Turk-Browne, N. B., & Scholl, B. J. (2005). Visual statistical learning through intervening noise. *Journal of Vision, 5,* 421a.

Kamin, L. J. (1968). 'Attention-like' processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9–33). Coral Gables, FL: University of Miami Press.

Kruschke, J. K. (2001). The inverse base rate effect is not explained by eliminative interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1385–1400.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science, 12,* 171–175.

Kruschke, J. K. (2005). Learning involves attention. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 113–140). Hove, United Kingdom: Psychology Press.

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

Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 1022–1050.

Lubow, R. E. (1989). *Latent inhibition and conditioned attention theory.* Cambridge, England: Cambridge University Press.

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica, 98,* 291–310.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associa-

bility of stimuli with reinforcement. *Psychological Review, 82,* 276–298.

McGeorge, P., & Burton, A. M. (1990). Semantic processing in an incidental learning task. *Quarterly Journal of Experimental Psychology, 42A,* 597–609.

Newell, B. R., & Bright, J. E. H. (2002). Evidence against hyperspecificity in implicit invariant learning. *Quarterly Journal of Experimental Psychology, 55A,* 1109–1126.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14,* 11–28.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127–162.

Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. (2003, August). *Reduction of uncertainty in human sequential learning: Preliminary evidence from artificial grammar learning.* Paper presented at the 25th annual meeting of the Cognitive Science Society, Boston.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language, 53,* 225–237.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298,* 604–607.

Perruchet, P. (2005). Statistical approaches to language acquisition and the self-organizing consciousness: A reversal of perspective. *Psychological Research, 69,* 316–329.

Perruchet, P. (in press). Implicit learning. In J. Byrne (Series Ed.) & H. L. Roediger III (Vol. Ed.), *Learning and memory: A comprehensive reference: Vol. 2. Cognitive psychology of memory*. Oxford, England: Elsevier.

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

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics, 17,* 97–119.

Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review, 7,* 307–313.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General, 133,* 573–583.

Perruchet, P., & Vinter, A. (2002). The self-organized consciousness. *Behavioral and Brain Sciences, 25,* 297–388.

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

Perruchet, P., Vinter, A., Pacteau, C., & Gallego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *Quarterly Journal of Experimental Psychology, 55A,* 485–503.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13,* 129–191.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22,* 425–469.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66,* 1–5.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64–69). New York: Appleton-Century-Crofts.

Roediger, H. L., III (1990). Implicit memory: A commentary. *Bulletin of the Psychonomic Society, 28,* 373–380.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606–621.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23,* 569–588.

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

Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202–220). London: Sage.

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

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

Stadler, M. A., & Frensch, P. A. (1998). *Handbook of implicit learning.* Thousand Oaks, CA: Sage.

Stadler, M. A., Warren, J. L., & Lesch, S. L. (2000). Is there cross-format transfer in implicit invariance learning? *Quarterly Journal of Experimental Psychology, 52A,* 235–245.

Stanback, M. L. (1992). Syllable and rime patterns for teaching reading: Analysis of a frequency-based vocabulary of 17,602 words. *Annals of Dyslexia, 42,* 196–221.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97,* B25–B34.

Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564.

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

Received April 28, 2006
Revision received August 6, 2007
Accepted August 18, 2007 ■