Effects of divided attention and speeded responding on implicit and explicit retrieval of artificial grammar knowledge

SHAUN HELMAN
University of Wales Institute, Cardiff, Wales

and

DIANNE C. BERRY
University of Reading, Reading, England

The artificial grammar (AG) learning literature (see, e.g., Mathews et al., 1989; Reber, 1967) has relied heavily on a single measure of implicitly acquired knowledge. Recent work comparing this measure (string classification) with a more indirect measure in which participants make *liking* ratings of novel stimuli (e.g., Manza & Bornstein, 1995; Newell & Bright, 2001) has shown that string classification (which we argue can be thought of as an explicit, rather than an implicit, measure of memory) gives rise to more explicit knowledge of the grammatical structure in learning strings and is more resilient to changes in surface features and processing between encoding and retrieval. We report data from two experiments that extend these findings. In Experiment 1, we showed that a divided attention manipulation (at retrieval) interfered with explicit retrieval of AG knowledge but did not interfere with implicit retrieval. In Experiment 2, we showed that forcing participants to respond within a very tight deadline resulted in the same asymmetric interference pattern between the tasks. In both experiments, we also showed that the type of information being retrieved influenced whether interference was observed. The results are discussed in terms of the relatively automatic nature of implicit retrieval and also with respect to the differences between analytic and nonanalytic processing (Whittlesea & Price, 2001).

Implicit learning is said to occur when a person "learns about the structure of a fairly complex stimulus environment, without necessarily intending to do so, and in such a way that the resulting knowledge is difficult to express" (Dienes & Berry, 1997, p. 3). One methodology for investigating implicit learning is the artificial grammar (AG) learning task, developed by Reber (1967). In this task, participants are exposed to stimuli (usually sequences, or *strings*, of letters) that follow an underlying grammar that dictates the order in which string elements may occur (see Figure 1). After exposure, the participants are able to classify novel strings (some grammatical, some ungrammatical) at above-chance levels, despite possessing little verbalizable knowledge of the rules of the grammar underlying the strings with which learning took place (e.g., Mathews et al., 1989; Reber, 1967). Over the last 35 years, there has been considerable

debate as to what kind of knowledge participants actually acquire in the AG learning task and how accessible this knowledge is to awareness (Dienes & Berry, 1997; Dulany, Carlson, & Dewey, 1984; Perruchet & Pacteau, 1990; Reber, Allen, & Regan, 1985; Shanks & St. John, 1994). Research has led to a good understanding of these issues (Cleeremans, Destrebecqz, & Boyer, 1998). It is now generally accepted that at least some of the knowledge acquired in AG learning is available to awareness (Dienes & Berry, 1997) and that the knowledge acquired includes letter pair information (Perruchet & Pacteau, 1990), various other types of *chunk strength* information (Meulemans & Van der Linden, 1997; Pothos & Bailey, 2000), similarity information (Pothos & Bailey, 2000), and knowledge about repetition patterns in study stimuli (Gomez, Gerken, & Schvaneveldt, 2000).

One limitation of this base of knowledge is that almost all published experiments in the AG learning literature have used string classification as the measure of what people "know." Although this task was originally intended as a subtle *implicit* way of measuring implicitly acquired knowledge, some researchers—most notably, Manza and his colleagues (see Manza & Bornstein, 1995; Manza et al., 1999; Manza, Zizak, & Reber, 1998)—have pointed out that string classification is not ideal for this role, since participants are explicitly asked to refer back to the learning phase in order to make their
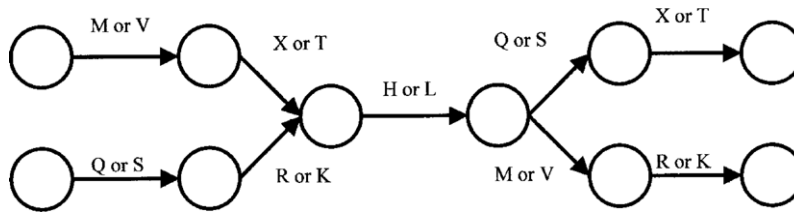
Figure 1. The artificial grammar used in Experiments 1 and 2. Strings are generated by starting at one of the two leftmost states and traversing between states in the direction of the arrows, until exiting the grammar from one of the two rightmost states. The letters picked up along any given route constitute a grammatical string (e.g., MXHQX, VTLVK).

classification decisions (Graf & Schacter, 1985). Strictly speaking, string classification is, at least partly, an explicit measure. Research findings support this view. For example, a concurrent secondary task has been found to interfere with classification performance, both when presented at study (Dienes, Broadbent, & Berry, 1991) and when presented at test (Dienes & Altmann, 1997; Dienes, Altmann, Kwan, & Goode, 1995). Such findings are usually observed with explicit, but not with implicit, tests of memory (e.g., Parkin & Russo, 1990; see Roediger & McDermott, 1993, for a review). Furthermore, a change between study and test in the surface features of strings (but not in the underlying structure) does not reduce, or only partly reduces, classification performance (e.g., Mathews et al., 1989; Whittlesea & Wright, 1997). Surface changes roughly analogous to this are known to influence implicit tests of memory, but not explicit tests of memory (e.g., Berry, Banbury, & Henry, 1997). If we characterize string classification as an explicit measure, we must conclude that we know a great deal about the explicit retrieval of AG knowledge but very little about the implicit retrieval of AG knowledge (Newell & Bright, 2001). In the present article, we examine existing evidence and present evidence from two new studies, all of which suggests that implicit and explicit retrieval of AG knowledge differ in a number of important ways.

Liking Ratings as an Implicit Measure of Memory[1]

Liking ratings have been suggested as an implicit measure of memory in the AG learning task (Gordon & Holyoak, 1983; Manza & Bornstein, 1995; Newell & Bright, 2001). Gordon and Holyoak were the first to show that liking ratings could be used to discriminate a *class* of stimuli (grammatical strings) from another *class* of stimuli (nongrammatical strings). Following up on this early work, Manza and Bornstein directly compared classification and liking ratings in an AG learning task. After exposure to grammatical strings, participants rated novel grammatical and ungrammatical strings, on a 6-point scale, either for whether they followed or violated the grammar or for how much they were liked or disliked. Discrimination performance was equivalent on both scales: Novel grammatical strings were rated as both more grammatical and more liked than ungram-

matical strings. However, the participants who carried out classification ratings demonstrated more explicit knowledge on a later explicit fragment completion measure (Dienes et al., 1991) than did the participants who carried out liking ratings. Manza and Bornstein suggested that classification ratings primed explicit knowledge of the grammar. Liking ratings, which did not draw participants' attention to the earlier study phase, were suggested to be less open to the influence of explicit processing and, therefore, a more appropriate implicit measure. The liking-based discrimination of grammatical and ungrammatical strings has been replicated a number of times (Helman, 2000; Manza et al., 1998; Newell & Bright, 2001; see also Whittlesea & Wright, 1997).

Recent research has focused on other differences between the liking and the classification tasks. Newell and Bright (2001) showed that liking-based discrimination is eliminated when test stimuli are represented in a different letter set from learning stimuli and when learning stimuli are fragments of whole strings and test stimuli are whole strings. These are both manipulations that classification-based discrimination withstands to some degree (Newell & Bright, 2001; Perruchet & Pacteau, 1990). Newell and Bright suggested that classification judgments rely on deliberate searches through memory to explicitly recall information about learning stimuli (explicit retrieval), whereas liking-based discrimination can rely on the unintentional attribution of processing fluency arising from previous exposure to grammatical learning strings (implicit retrieval). A characterization of the two tasks in terms of implicit and explicit retrieval explains why changing the physical characteristics of stimuli, such as letter set, might interfere with liking-based discrimination. It is a well-known finding in implicit memory research that physical changes in stimuli adversely affect performance on implicit measures of memory, while leaving performance on explicit measures largely unaffected (Berry et al., 1997; Roediger & McDermott, 1993).

In the present article, we report results from two experiments in which we sought to explore other potential differences between explicit and implicit retrieval of AG knowledge. In Experiment 1, we examined the effect of dividing attention on implicit and explicit retrieval. In Experiment 2, we examined whether or not there is a dif-

ference in the speed of the processing underlying implicit and explicit retrieval. Although both manipulations have previously been used in the AG learning literature (Dienes & Altmann, 1997; Dienes et al., 1995; Dienes et al., 1991; Higham, Vokey, & Pritchard, 2000; Turner & Fischler, 1993), neither has yet been used when the liking-based discrimination of AG strings has been examined. Using such manipulations is important if we are to find out more about the processing underlying liking ratings in the AG learning task.

Retrieval Intentionality and Analytic Versus Nonanalytic Processing

Two theoretical frameworks are especially relevant to the present article. First, the retrieval intentionality criterion, developed by Schacter and his co-workers (Schacter, Bowers, & Booker, 1989), has relevance to the implicit and explicit tasks used. Second, the distinction between analytic and nonanalytic processing, recently introduced by Whittlesea and Price (2001), provides a good basis for some of the fine-grained predictions about the dissociations we can expect between our implicit and explicit tasks. These two theoretical frameworks are discussed separately below.

Previous work in the AG learning task has tended not to match implicit and explicit measures closely enough, in terms of task demands and sensitivity, to permit firm conclusions as to the crucial factor in any observed dissociations (see Merikle & Reingold, 1991, and Shanks & St. John, 1994, for discussions of this issue). In the present study, we attempted to match the implicit and explicit tasks so as to be able to avoid this criticism. The retrieval intentionality criterion (Schacter et al., 1989; see Butler & Berry, 2001, for a review; see Reingold & Toth, 1996, for a critique) states that implicit and explicit tasks should differ only in terms of whether their instructions make reference to the previous learning episode (see also Merikle & Reingold, 1991). This ensures that when dissociations are observed between tasks, they can be attributed to differences in the degree to which instructions encourage the use of past experience, rather than to any other differences, such as task difficulty. With this in mind, we chose not to use classification as an explicit measure of memory in the present study, since it differs in more than one way from the liking task (our proposed implicit measure). Most important, in addition to encouraging people to explicitly search for and retrieve information from the earlier study phase, classification task instructions typically mention *rule structure* as the information to be retrieved. This instruction may alter the effectiveness of retrieval independently of the explicit focus on past experience and is especially problematic given findings in the literature that suggest that rule structure is not something people learn in AG learning experiments (e.g., Kinder & Assmann, 2000). We therefore decided to develop a version of the liking-rating task that encouraged the use of past experience but did not differ in any other way from its implicit counterpart. In this task, which was

developed through pilot work in Helman (2000), participants were required to rate test strings for liking (as in our implicit test condition) but were asked to base their liking ratings on how similar the test strings were to the study strings seen earlier.

The distinction between analytic and nonanalytic processing was introduced by Whittlesea and Price (2001), partly as a theoretical explanation of a particular feature of the *mere-exposure effect* (Zajonc, 1968; see Bornstein, 1989, for a review). The mere-exposure effect occurs when exposure to stimuli leads to an increased liking rating for those stimuli, relative to nonexposed stimuli.[2] One feature of the mere-exposure effect is that it can be demonstrated in the absence of stimulus recognition (e.g., Kunst-Wilson & Zajonc, 1980). That is, people can be shown to be able to discriminate exposed and nonexposed stimuli on the basis of liking, even when they fail to discriminate the same stimuli on the basis of recognition. On first inspection, this result seems at odds with the dominant theoretical explanation of the mere-exposure effect itself, which is that when making liking ratings to stimuli that have been seen before, people experience a fluency of processing related to the previous exposure and attribute this to liking the stimulus (Bornstein & D'Agostino, 1992). Although this account is widely accepted in the literature, it cannot explain why people are unable to attribute the same fluency of processing to stimulus recognition. Whittlesea and Price (2001) suggested that the answer lies in the types of stimuli used in mere-exposure studies and how these stimuli encourage a particular type of processing when task instructions ask for recognition. Specifically, the stimuli used are always from the same class (e.g., pictures or nonsense words), which leads to high perceptual resemblance between exposed and nonexposed stimuli. Whittlesea and Price suggested that this resemblance encourages people to adopt an *analytic* style of processing when asked to try and recognize stimuli they have seen, and that this strategy, which involves analyzing stimuli in parts (rather than as wholes) is doomed to fail if crucial *part-based* details have not been encoded in exposed stimuli or are not noticed during retrieval. According to Whittlesea and Price, adopting a part-based analytic style of processing also stops people from experiencing the fluency that would result from the processing of stimuli as wholes—that is, nonanalytic processing. The distinction between analytic and nonanalytic processing explains why liking-based discrimination of exposed and nonexposed stimuli can occur in the absence of recognition of exposed stimuli. When people make liking decisions, they naturally adopt a nonanalytic processing style (processing stimuli as wholes), experience fluency for exposed stimuli, and attribute this fluency to liking the stimuli more than they like the nonexposed stimuli. On the other hand, when people make recognition decisions, they adopt a part-based analytic processing style, fail to experience the fluency that would help them make the discrimination, and thus perform at chance on the recog-

nition task because the stimuli are all perceptually similar and the relevant part-based information is not available to enable successful analytic discrimination.

Whittlesea and Price (2001) showed that when participants were encouraged to adopt a nonanalytic processing style during recognition (by use of a pool of stimuli that did not show high perceptual resemblance), recognition performance improved from the chance levels. On the basis of this finding and on the way in which differences between the two types of processing are conceptualized, it follows that another implication of the distinction between analytic and nonanalytic processing is that stimuli that have part-based information that is sufficiently salient will be amenable to successful analytic processing (since the parts needed for discrimination will be noticed), whereas stimuli that have only nonsalient part-based information will not be amenable to analytic processing (since the parts needed for discrimination will not be noticed). We contend that the explicit retrieval task used in the present study is consistent with an analytic processing style, given that it encourages the use of previous experience in the same way that a recognition task does. Therefore, a further aim of the present study was to examine whether or not analytic processing could proceed on stimulus sets that were homogenous (as sets of AG strings are), but only when sufficiently salient part-based information was available. To do this, we used two different classes of nongrammatical string during the tests: one with salient violations of grammatical structure and one with very subtle, nonsalient violations of grammatical structure. The details of these two types of string can be found in the Method section of Experiment 1.

To summarize, our aim was to demonstrate implicit and explicit retrieval in an AG learning task by using implicit and explicit versions of the liking ratings task used by Manza and Bornstein (1995). We also examined the influences of divided attention and speeded responses on the implicit and explicit tasks. Finally, we asked whether the type of violation of grammatical structure (salient vs. nonsalient) influences whether explicit retrieval (analytic processing) can be successfully used to discriminate grammatical from nongrammatical strings.

EXPERIMENT 1

In Experiment 1, we examined whether retrieval of AG knowledge under implicit and explicit retrieval instructions differs in terms of the amount of attentional resources required for its successful execution (Mulligan, 1997; Parkin & Russo, 1990). Previous studies in the AG learning literature have shown that a concurrent secondary task interferes with *overall* classification performance, when present either during learning (Dienes et al., 1991) or at test (Dienes & Altmann, 1997; Dienes et al., 1995). Interestingly, Dienes et al. (1995) and Dienes and Altmann (1997) showed that when only trials on which participants claimed to be completely guessing were considered,

classification performance was not affected by a secondary task at test. This raises the possibility that retrieval of AG knowledge sometimes requires the allocation of attentional resources but sometimes does not. Dienes et al. (1995) interpreted their findings in terms of a subjective threshold of awareness (see also Dienes & Berry, 1997), which separates qualitatively different implicit and explicit knowledge bases. Another interpretation is that trials on which the participants believed they were guessing were trials that did not involve the explicit retrieval of AG knowledge but, rather, involved falling back on an implicit, or nonanalytic, processing style (subjectively experienced by the participants as *guessing*). Therefore, we asked whether a secondary task at *test* interferes differentially with explicit and implicit retrieval of AG knowledge (i.e., under instructions encouraging and instructions not encouraging the use of past experience). We also examined the role of salience of violations of grammatical structure present in ungrammatical strings. We used two types of ungrammatical string: ungrammatical positional (UP; ungrammatical due to a *positional* violation—a legal bigram in an illegal position) and ungrammatical bigram (UB; ungrammatical due to an illegal bigram). It was expected that the increased salience of the violations in UB strings would facilitate their discrimination from grammatical strings via explicit (analytic) processing, although we did not make any clear predictions of exactly how this might interact with the other variable of interest (concurrent task).

Method

Participants

Sixty-four undergraduate and postgraduate psychology students, between 18 and 45 years of age, acted as participants. They received either course credit or £3 as a reward for their participation. A further 16 students from the same population were tested as baseline participants.

Design

The independent variables were test string type (within subjects; grammatical, UP, and UB), task (between subjects; single- or dual-task conditions), and instructions (between subjects; implicit and explicit). The participants were assigned randomly to the four resulting groups (single-task implicit, dual-task implicit, single-task explicit, and dual-task explicit). The dependent variable was the mean liking rating given to the test strings.

Materials

A finite state AG was developed (see Figure 1). With this grammar, 80 strings were created (see the Appendix for all the strings): 16 grammatical learning strings, 16 grammatical distractors for the recognition phase, 16 grammatical test strings, and 32 ungrammatical test strings. Learning strings were created in such a way that the eight possible beginning bigrams (MX, MT, VX, VT, QR, QK, SR, and SK) and the eight possible end bigrams (QX, QT, SX, ST, MR, MK, VR, and VK) were each seen twice in the whole learning set.

The learning strings were created in the following way. The middle letters (L and H) were each used eight times in the learning set, and each beginning or end bigram was seen before and after each middle letter only once. This ensured that there was no information contained in the middle letters except their own identity and position. No beginning bigram was present in the same string with any

particular end bigram more than once in the whole learning set. This ensured that, as for the middle letters, there was no information in the beginning or the end bigrams that went beyond their own identity and position. The grammatical distractors used in the recognition phase were created in exactly the same way as the learning strings.

Each grammatical test string was created by taking one of the 16 learning strings and changing one or two letters in the first, second, fourth, or fifth position within that string, in such a way as to keep the string grammatical. Ungrammatical test strings were created in the same way as grammatical test strings, except that the change to learning strings rendered them ungrammatical. There were two different kinds of ungrammatical test strings. The 16 UP strings contained a single *positional* violation, meaning that it was rendered ungrammatical by a legal bigram in an illegal position. This was achieved by using either one of the 8 legal beginning bigrams as an end bigram or one of the 8 legal end bigrams as a beginning bigram, in each UP string. Each bigram from this pool of 16 was used only once. The 16 UB strings were created in the same way as the UP strings, except that a pool of 16 illegal bigrams was used. These were bigrams that had never been seen before, in any position, in the learning and recognition distractor sets.

Procedure

Learning phase. The participants were told that they would be taking part in a recognition memory test for meaningless strings of letters and that the first phase involved simply studying the strings on which they would later be tested. The participants were then exposed to the learning strings, one at a time, on a computer screen. Each presentation consisted of a learning string that remained on screen for 3 sec, followed by a 3-sec interstimulus interval. Each of the 16 learning strings was presented five times in all, with order being completely randomized for each participant.

Recognition phase. The participants were instructed that they would be shown more strings and would be required to try to recognize which strings they had seen before and which strings were new. Learning strings and recognition distractor strings were then displayed, one at a time, on the computer screen, with order of presentation completely randomized for each participant. Each string remained on screen until the participants had indicated, on a 6-point scale, how confident they were that the string was old or new. The scale consisted of six buttons, presented on the screen, and the participants made their recognition ratings by clicking the relevant button. From left to right on the screen, the buttons were labeled *really sure–new*, *sure–new*, *sort of sure–new*, *sort of sure–old*, *sure–old*, and *really sure–old*. For the purpose of analysis, the numerical values 1 to 6 were applied to the response buttons, from left to right.

Test phase. After the recognition phase, the participants were informed that there was a final phase to the experiment. The test phase procedure differed according to instructions and task conditions. The participants in the implicit conditions were informed that they would be required to rate some new strings for liking, on a 6-point scale, and were told to base their liking ratings solely on their "first impression" or "gut feeling." Baseline participants carried out the test phase without any prior phases, in order to check that the three sets of test strings were equally likeable without prior exposure to learning strings. The participants in explicit conditions were told to base their liking ratings solely on how similar the strings were to the strings they had seen in the learning phase (the more similar the test strings were to the learning strings, the more liking for them the participants should express).[3] The participants in the single-task conditions simply carried out the test task in isolation, as per their instructions. Those in the dual-task conditions verbally generated random digits, between zero and nine (inclusive), at a rate of one every 1.5 sec, while making their liking ratings. A metronome was set to click every 1.5 sec to aid the participants in this secondary task.

For all the groups, test strings were presented one at a time and remained on the screen until the participants made their liking rat-

ings. The participants rated strings via a 6-point scale, on screen, as they had in the recognition phase. Buttons were labeled, from left to right, on the screen, *really dislike*, *dislike*, *sort of dislike*, *sort of like*, *like*, and *really like*. For the purpose of analysis, numerical values were assigned as in the recognition phase.

Results

Recognition Data

The participants gave higher recognition ratings to learning strings than to distractors (mean ratings and standard deviations for learning and distractor strings were 4.47 [0.48] and 3.59 [0.54] respectively). A $2 \times 2 \times 2$ (string type $\times$ instructions $\times$ task) mixed analysis of variance (ANOVA) revealed a significant main effect of string type [$F(1,60) = 89.59$, $MS_e = 0.274$, $p = .0001$]. All other main effects and all interactions were nonsignificant. Thus, the participants in all the conditions performed equally well at discriminating old from new strings in the recognition phase.

Liking Rating Data

Data from the 16 baseline participants were analyzed using a one-way repeated measures ANOVA, which revealed no significant effect of string type [$F(2,30) < 1$]. This confirmed that all three test string lists were equally likeable without prior exposure to learning strings (means and standard deviations for novel grammatical, UP, and UB strings, 3.66 [0.42], 3.61 [0.47], and 3.66 [0.33], respectively).

The baseline and recognition data suggest that any effects of string type found in any of the experimental groups could be attributed to prior exposure to learning strings, not to differing attention at learning or preexisting preferences for any of the string lists. Figure 2 shows the mean liking ratings given to grammatical, UP, and UB test strings under single- and dual-task conditions by the experimental participants.

These data were entered into a $3 \times 2 \times 2$ (string type $\times$ instructions $\times$ task) mixed ANOVA. This revealed significant main effects of string type [$F(2,120) = 53.32$, $MS_e = 0.114$, $p = .0001$] and of task [$F(1,60) = 6.89$, $MS_e = 0.52$, $p = .011$]. The interactions between string type and instructions and between string type and task were also significant [$F(2,120) = 4.05$, $MS_e = 0.114$, $p = .02$ and $F(2,120) = 4.35$, $MS_e = 0.114$, $p = .015$, respectively], as was the three-way interaction of string type, instructions, and task [$F(2,120) = 6.56$, $MS_e = 0.114$, $p = .002$].

Planned 3×2 (string type $\times$ task) ANOVAs were carried out for implicit and explicit instruction groups separately. An analysis of the implicit participants' data revealed that the main effect of string type was significant [$F(2,60) = 24.3$, $MS_e = 0.078$, $p = .0001$] but that the string type $\times$ task interaction was nonsignificant [$F(2,60) < 1$]. Tukey's HSD tests on the effect of string type at each level of task confirmed that grammatical test strings were rated more highly than both UP and UB strings, under both single- and dual-task conditions. An analysis of the explicit participants' data revealed a significant main effect of
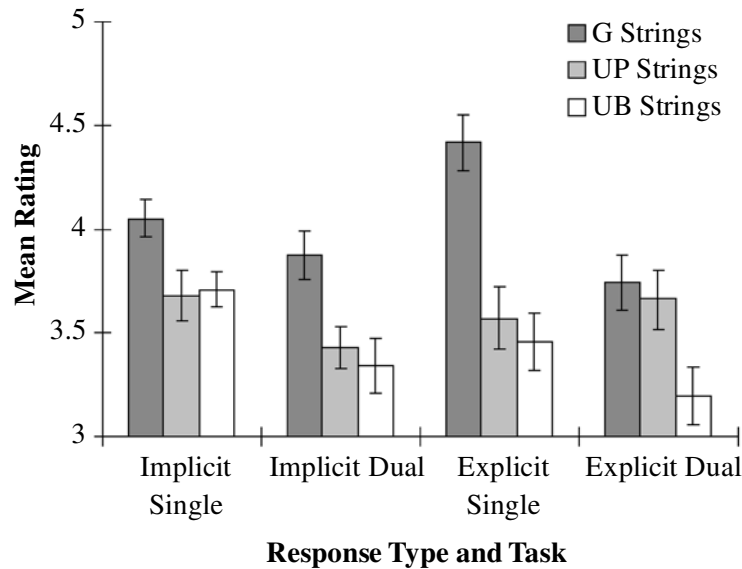
Figure 2. Mean liking rating by string type, response type (implicit or explicit), and task in Experiment 1 (standard errors shown). G, grammatical; UP, ungrammatical positional; UB, ungrammatical bigram.

string type $[F(2,60) = 30.99, MS_e = 0.15, p = .0001]$ and a significant interaction between string type and task $[F(2,60) = 7.86, MS_e = 0.15, p = .0009]$. Tukey's HSD tests on the effect of string type at each level of task revealed that, under single-task conditions, grammatical test strings were rated more highly than both UP and UB strings. In contrast, under dual-task conditions, grammatical test strings were rated more highly than UB strings but were not rated more highly than UP strings.

Analyses of the effect of condition on the ratings given to different string types were also run separately for implicit and explicit data. For explicit data, only grammatical strings varied significantly with condition, being rated more highly under single-task conditions $[t(30) = 3.618, p < .001]$. For implicit data, only UB strings varied significantly, being rated more highly under single-task conditions $[t(30) = 2.35, p = .026]$.

Random Number Generation

The random numbers generated by the participants in the dual-task conditions were analyzed, and the following measures of departure from randomness were calculated: redundancy (R), adjacency (A), and second- and fifth-order phi indexes (Towse & Neil, 1998). Between-subjects t tests confirmed that the random number sequences generated by the explicit and the implicit participants did not differ on any of these measures $[t(30) = -0.1716, -0.5419, -0.9772,$ and -0.7811, respectively; all ps > .33]. This suggests that the discrimination of grammatical and UP strings under dual-task conditions by the implicit instructions participants is unlikely to have been due to less attention's being paid to the random number generation task.

Discussion

When measured implicitly, discrimination of grammatical and UP strings (strings made ungrammatical by the presence of a legal bigram in an illegal position) was unaffected by a concurrent secondary task. Conversely, when measured explicitly, grammatical and UP strings were discriminated only under single-task conditions. These results are consistent with the idea that explicit, but not implicit, retrieval of AG knowledge requires the allocation of substantial attentional resources. The fact that grammatical and UB strings (strings made ungrammatical by the presence of an illegal bigram) were discriminated using implicit and explicit retrieval equally well under single- and dual-task conditions is consistent with the idea that the type of knowledge being retrieved can influence the amount of attentional resources required for successful retrieval, independently of the retrieval mode being used. One possibility is that some types of knowledge (in this case, the knowledge that some bigrams in a given string are novel) are sufficiently salient to allow relatively automatic retrieval, even when such retrieval is carried out under a relatively explicit, or analytic, processing mode. This point will be returned to in the General Discussion section.

EXPERIMENT 2

Another manipulation that has been used to dissociate implicit and explicit processes in the AG learning task is that of speeded responses (Higham et al., 2000; Turner & Fischler, 1993). Higham et al. used a variant of the process dissociation procedure (PDP), introduced by Jacoby (1991). Participants learned strings from two grammars

(*A* and *B*) and were then split into two test conditions. Some participants were told at test to endorse only strings from Grammar B as *grammatical* (opposition condition), whereas some were told to endorse strings from both Grammars A and B (in-concert condition). Higham et al. found that the participants in the opposition condition, despite being told to classify them as ungrammatical, still endorsed strings from Grammar A as grammatical more often than they rejected them as ungrammatical. Higham et al. argued that this reflected an unconscious, automatic influence on responding. Supporting this, Higham et al. showed that imposing a very short response deadline (1 sec) did not alter this influence. Grammar A strings were discriminated from ungrammatical strings equally well under no deadline and a 1-sec deadline in the opposition condition. In contrast, the ability of participants to reject Grammar A strings (effectively, the difference between Grammar A endorsements in the in-concert condition and the opposition condition), which Higham et al. argued reflected a controlled influence on responding, was altered by the introduction of the response deadline. The participants were hindered in their ability to correctly reject Grammar A strings under a short response deadline.

As we have mentioned, Higham et al. (2000) used an opposition paradigm based on the PDP (Jacoby, 1991). In Experiment 2, we adopted the more usual dissociation paradigm, again using the differences in task instructions (encouraging the use of past experience vs. not encouraging the use of past experience) as the manipulation designed to encourage explicit and implicit retrieval, respectively. We examined the effect of imposing a very short response deadline at test on implicit and explicit retrieval of AG knowledge. Following Higham et al., if the liking-based discrimination of strings at test (implicit retrieval) is underlain by automatic influences on responding, a short response deadline should have little or no effect on the discrimination performance. On the other hand, it should have an adverse effect on discrimination performance in the explicit condition. On the basis of the findings from Experiment 1, we might also expect the salience of violations in ungrammatical strings to influence the effect of a deadline. Specifically, we might expect to find grammatical and UP strings being discriminated in the explicit condition only under no deadline. On the other hand, if the UB strings contained information that was in itself salient enough to allow automatic retrieval independently of retrieval mode, we would expect UB and grammatical strings to be discriminated in the explicit condition under both deadlines. Such a finding would provide more evidence consistent with the idea that the nature of retrieval of AG knowledge can depend on the nature of the knowledge itself, as well as on the processing mode (implicit/nonanalytic vs. explicit/analytic) adopted.

Method

Participants

Sixty-four undergraduate and postgraduate students acted as participants. They received either course credit or £3 as a reward for their participation. None of the participants had taken part in Ex-

periment 1 or in any implicit learning experiment before. As in Experiment 1, a further 16 students from the same population were run as baseline participants. Eight of these baseline participants were run with no response deadline, whereas the remainder were run with a 2.5-sec deadline.

Design

The independent variables were string type (within subjects; grammatical, UP, and UB), deadline (between subjects; no deadline or 2.5-sec deadline), and instruction (between subjects; implicit and explicit). The participants were assigned randomly to the four resulting groups (no-deadline explicit, 2.5-sec deadline explicit, no-deadline implicit, and 2.5-sec deadline implicit). The dependent variable was the mean liking rating given to strings.

Materials

The strings in Experiment 2 were exactly the same as those used in Experiment 1.

Procedure

Learning and recognition phases. The learning and recognition phases were identical to those in Experiment 1.

Test phase. Before the actual test phase, the participants were given practice at using the liking rating scale. The labels from the rating buttons were presented on screen, and the participants were required to click on the relevant button with the mouse. This practice was really necessary only for the 2.5-sec deadline participants (to familiarize them with responding under a deadline), but all the participants took part in the practice session for the sake of methodological balancing. After the practice phase, the participants in the no-deadline conditions gave their liking ratings in exactly the same way as the participants had in the single-task condition in Experiment 1. The participants in the 2.5-sec deadline conditions did the same but were required to produce each rating within 2.5-sec, measured from the time each string appeared on screen. Whenever no rating button was pressed within the deadline, a message appeared on screen, saying "You missed the deadline!—Please try to respond within the time given!—Click on OK to rate next string." The participants clicked on the OK button to resume rating. After the test phase, the participants were debriefed and rewarded for their participation. The baseline participants carried out the test phase without any prior phases, in order to check that the three sets of test strings were equally likeable at both deadlines, without prior exposure to learning strings.

Results

Recognition Data

The participants gave higher recognition ratings to learning strings than to distractors (mean ratings and standard deviations for learning and distractor strings were 4.55 [0.54] and 3.62 [0.69], respectively). A $2 \times 2 \times 2$ (string type $\times$ deadline $\times$ response type) mixed ANOVA revealed that the main effect of string type was significant [$F(1,60) = 93.86$, $MS_e = 0.3$, $p = .0001$], but that all other main effects and interactions were nonsignificant. Thus, the participants in all the conditions performed equally well at discriminating old from new strings in the recognition phase.

Liking Rating Data

The data from the 16 baseline participants were entered into a 3×2 (string type $\times$ deadline) mixed ANOVA, which revealed no significant main effect of deadline [$F(1,14) < 1$], no main effect of string type [$F(2,28) = 1.47$, $MS_e = 0.076$, $p = .25$], and no interaction between

these two variables [$F(2,28) < 1$]. This confirmed that all three test string lists were equally likeable, without prior exposure to learning strings, at both deadlines (means and standard deviations for grammatical, UP, and UB strings, 3.36 [0.36], 3.34 [0.37], and 3.51 [0.29] for the no-deadline condition, and 3.29 [0.53], 3.46 [0.48], and 3.48 [0.38] for the 2.5-sec deadline condition).

As in Experiment 1, the baseline and recognition data suggest that any effects of string type found in the experimental groups can be attributed to prior exposure to learning strings, not to differing attention at learning or to preexisting preferences for any of the string lists.

Figure 3 shows the mean liking ratings given to grammatical, UP, and UB strings under no-deadline and 2.5-sec deadline conditions by the experimental participants. The number of missed deadlines in the 2.5-sec deadline conditions constituted 5.3% and 3.5% of the responses for the implicit and the explicit instruction participants, respectively. These data were excluded from analyses, since they were never actually recorded by the computer. The proportions were considered sufficiently low so as to not have any noticeable effect upon the reliability of the liking rating measure under the 2.5-sec deadline.

The data from the experimental participants were entered into a $3 \times 2 \times 2$ (string type $\times$ deadline $\times$ response type) mixed ANOVA, which revealed a significant main effect of string type [$F(2,120) = 33.73, MS_e = 0.128, p = .0001$], a marginally significant string type $\times$ deadline interaction [$F(2,120) = 3.01, MS_e = 0.128, p = .053$], and a significant string type $\times$ deadline $\times$ response type interaction [$F(2,120) = 3.44, MS_e = 0.128, p = .035$].

Planned 3×2 (string type $\times$ deadline) ANOVAs were carried out for the implicit and the explicit participants'

data separately. An analysis of the data from the implicit conditions revealed that the main effect of string type was significant [$F(2,60) = 9.94, MS_e = 0.127, p = .0002$], but that the string type $\times$ deadline interaction was nonsignificant [$F(2,60) = 2.70, MS_e = 0.127, p = .0751$].[4] An analysis of the data from the explicit conditions revealed a significant main effect of string type [$F(2,60) = 25.74, MS_e = 0.13, p = .0001$] and a significant interaction between string type and deadline [$F(2,60) = 3.73, MS_e = 0.13, p = .0297$]. Tukey's HSD tests on the effect of string type at each deadline revealed that, under no deadline, grammatical strings were rated more highly than both UP and UB strings. In contrast, under the 2.5-sec deadline, grammatical strings were rated more highly than UB strings but were not rated more highly than UP strings.

Analyses of the effect of condition on the ratings given to different string types were also run separately for implicit and explicit data. For the explicit data, UP strings approached a significant variation with condition, being rated more highly under the deadline condition [$t(30) = -1.895, p = .068$]. For implicit data, grammatical strings approached significant variation, being rated more highly under the deadline condition [$t(30) = -1.843, p = .075$; see note 3].

Discussion

The results from Experiment 2 are consistent with the view that the liking-based discrimination of grammatical and ungrammatical strings, which seems to involve a retrieval mode that is implicit, or nonanalytic, involves processing that is relatively fast when compared with a retrieval mode that encourages explicit retrieval of past experience. The participants with implicit instructions
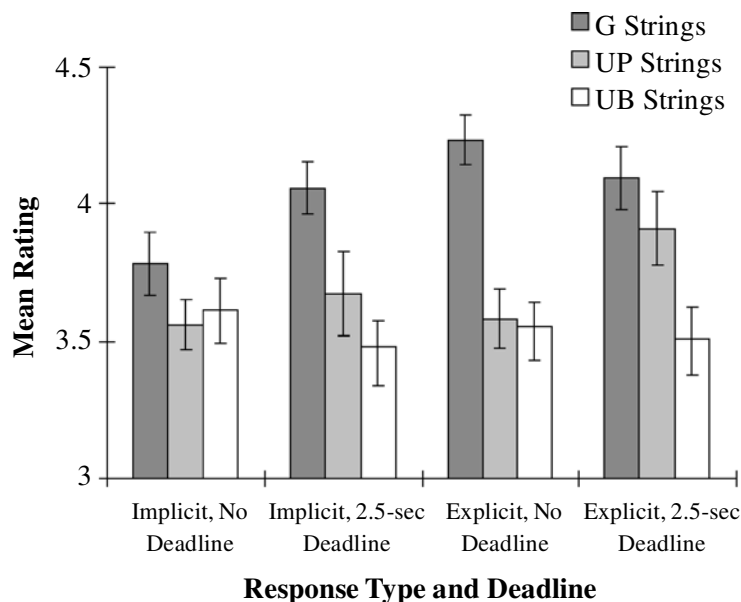

Figure 3. Mean liking rating by string type, response type (implicit or explicit), and deadline in Experiment 2 (standard errors shown). G, grammatical; UP, ungrammatical positional; UB, ungrammatical bigram.

were able to retrieve and use this information equally well under no-deadline and 2.5-sec deadline conditions. The participants given explicit instructions, conversely, were unable to successfully retrieve positional information under the 2.5-sec response deadline. Again, we found evidence that the knowledge of illegal bigrams, needed to discriminate grammatical from UB strings, was accessed more easily than the knowledge of the positions of legal bigrams. Grammatical and UB strings were again discriminated by all the participants at all response deadlines. This point is returned to in the General Discussion section below.

GENERAL DISCUSSION

The experiments presented in this article have provided evidence consistent with the idea that when an implicit measure is used, discrimination of grammatical and ungrammatical strings in AG learning is based on unintentional, automatic retrieval of AG knowledge. Such discrimination was not abolished by a concurrent secondary task or by a forced speeded response, in the way that explicit discrimination of grammatical and ungrammatical (UP) strings was. These findings add to a growing body of evidence suggesting that there are important differences between implicit and explicit tests of AG knowledge. Implicit tests give rise to less explicit knowledge of the grammar underlying performance than do explicit tests (Manza & Bornstein, 1995) and are more tied to the surface features of learning stimuli and to a match in processing contexts at learning and test (Newell & Bright, 2001; Whittlesea & Wright, 1997). Implicit tests are also, typically, a less sensitive measure of AG knowledge than is the intentional comparison of test to learning strings (Newell & Bright, 2001; but see Manza & Bornstein, 1995). However, Experiments 1 and 2 showed an implicit test to be *more* sensitive than a well-matched explicit test under dual-task conditions and when responses were required quickly.

Newell and Bright (2001) suggested that the liking-based task is underlain by perceptual fluency, as has been suggested for the standard mere-exposure effect and implicit measures of memory (see, e.g., Bornstein & D'Agostino, 1992; Roediger & McDermott, 1993). If we accept this characterization, the results from the present study suggest that, in the AG learning task, the processes that attribute this fluency to increased liking (Jacoby & Dallas, 1981) can operate with minimal processing resources being available and can operate more quickly than the processes underlying explicit retrieval of AG knowledge.[5]

Our findings may also offer an alternative explanation for the finding of Dienes et al. (1995) and Dienes and Altmann (1997) that, although a secondary task at test interfered with overall discrimination (by classification) of grammatical and ungrammatical strings, it did not interfere when only trials on which participants claimed they were "completely guessing" were considered. Dienes et al. (1995) explained this finding in terms of separate knowledge bases—one above and one below a phenomenologically distinct *subjective* threshold of awareness. If we make the assumption that trials on which participants claimed to be guessing were those trials on which they were using implicit retrieval processing, the results can be explained in terms of the retrieval style they adopted on any given trial (implicit/nonanalytic vs. explicit/analytic). Future work might concentrate on trying to resolve these alternate explanations of such findings.

A further issue that arises from the experiments in the present study is the fact that there were independent influences of instructions and of the types of information being retrieved on the amount of attention and time required for successful retrieval. This worked in two ways. First, regardless of instructions, UB strings were always discriminated from grammatical strings. An intuitive way of thinking about this is that, in the AG task, some violations of grammaticality in ungrammatical strings (in this case, illegal, never seen before, bigrams) are salient enough to *pop out* automatically as novel (and therefore, in this case, also less liked).[6] Second, although the knowledge required for discrimination of UP and grammatical strings could be accessed under dual-task and speeded response conditions when instructions encouraged implicit retrieval, when instructions encouraged explicit retrieval, this knowledge was not retrieved (under dual-task or speeded response conditions).

These findings are broadly compatible with the distinction between analytic and nonanalytic processing introduced by Whittlesea and Price (2001). With an analytic processing style, the participants were able to focus on part-based information and successfully use it for the salient UB strings, no matter what the other demands of the task (secondary task or speeded response). On the other hand, UP strings possessed information with insufficient salience to be used when task demands were high, although it could be used when task demands were low (the single-task condition in Experiment 1 and the no-deadline condition in Experiment 2). However, even when it could not be used for UP strings, the analytic processing style adopted by the participants blocked the experience of processing fluency that would have enabled discrimination of grammatical and UP strings (as evidenced by the fact that the participants in the implicit, or nonanalytic, condition were able to discriminate grammatical and UP strings under all levels of task demands). Whether this effect of knowledge salience is a general feature of encoding and retrieval in human memory or is specific to the AG learning paradigm remains to be examined in future work. For example, most previous work outside of the implicit learning literature has indicated that *retrieval* processes are unaffected by secondary task manipulations (e.g., Baddeley, Lewis, Eldridge, & Thomson, 1984; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Jacoby, Woloshyn, & Kelley, 1989; Naveh-Benjamin, Craik, Gavrilescu, & Anderson, 2000). The findings from the present study and those of Dienes et al. (1995) and Dienes and Altmann (1997) suggest that this may not always be the case and that it may rely on the

type of the material being retrieved and the instructional manipulations employed.

Another feature of the data was that ratings given to UP and UB strings sometimes varied with condition.[7] By examining how string ratings varied across conditions for explicit and implicit tasks, we might be able to begin to explain, at a more detailed level, what is happening during interference from the dual tasks or speeded responses. In Experiment 1, for implicit data, UB strings were rated significantly lower under dual-task conditions than under single-task conditions. This finding is interesting since, like the finding of *increased* ratings for grammatical strings (and *increased* discrimination of grammatical and UP strings) under a deadline in Experiment 2, it suggests that the implicit task tends to be performed more successfully when there is some additional load on processing. We do not wish to advance this as a serious argument on the basis of these data alone. However, we would venture to suggest that this can be tested with stronger manipulations in future work. Also, as with all implicit tasks, there is a danger that some of the implicit performance in the present study was due to some explicit contamination. Future work that attempts to minimize contamination, while examining more specifically the influence of concurrent load on implicit retrieval performance in the AG learning task, is needed to further our understanding of these data.

Another interesting pattern of data emerges when one considers how ratings given to string types varied with condition in the explicit task. Ratings given to UB strings did not vary significantly with condition in either Experiment 1 or Experiment 2. However, when grammatical and UP strings are considered, a more complex story emerges. We did not specify how any interference in the explicit task would manifest itself in the data. We predicted only that discrimination in the explicit task would be successful in some conditions and not in others. With regard to the discrimination of grammatical and UP strings, there are at least two possibilities. Given higher ratings to grammatical strings than to UP strings under ideal conditions (single task/no deadline), this discrimination could disappear under dual-task or deadline conditions either by a drop in the ratings given to grammatical strings to the level of those given to UP strings or by a rise in the ratings given to UP strings to the level of those given to grammatical strings. The data are interesting because *both* these possibilities occur, depending on the nature of the interfering manipulation. In Experiment 1, with a dual-task manipulation, ratings to grammatical strings drop to the level of those given to UP strings. In Experiment 2, with a deadline on responding, ratings to UP strings rise to the level of those given to grammatical strings. We have no theoretically based explanation for this pattern of data, only speculation based on possible differences in subjective quality that may be experienced by participants between carrying out an explicit task under dual-task or under deadline conditions and in the different rating strategies that might result. If

we assume that participants carry out an explicit task by adopting an *analytic*, or *part-based*, style of processing, nongrammatical and grammatical strings must be discriminated on the basis of finding violations of grammar in those strings (UP strings, legal bigram in an illegal position; UB strings, illegal bigram) and using this information to identify the string as relatively dissimilar to the strings studied previously (and thus rating them lower on the scale). It seems intuitive that when carrying out an explicit task under a deadline, participants may notice that some strings (UB) are very easily discriminated (on the basis of their salient violations of grammatical structure). Under the extreme time pressure involved, participants may simply decide to rate the UB strings low on the scale and all other strings high on the scale. The very act of deciding on such a strategy may result in a strategic bias in responses, since every time a string is noticed as being *not obviously dissimilar (as the UB strings are)*, it is logically labeled as *similar*, and this may increase the rating given to it. Hence, in this situation, UP strings are given higher ratings than normal. On the other hand, under a dual-task manipulation, there is no "pressure" to respond at all. Participants may find the task difficult, but there is less motivation, in the absence of any pressure to respond, to develop such a strategy. Indeed the dual-task manipulation was almost universally experienced as "very boring." Such a subjective quality may well have manifested itself as lower ratings to strings generally (there is, in fact, quite a marked, although nonsignificant, trend for UB strings to be rated lower in explicit dual-task than in explicit single-task conditions). Hence, in this situation, grammatical strings are given lower ratings than normal. This speculation could be tested in future work. On a subject-by-subject basis, it should be possible to measure the urgency that participants felt to respond (in the deadline condition) or the "boredom" they felt (in the dual-task condition), to see whether these predict the pattern of ratings given to grammatical and UP strings. Such data would begin to expand our understanding of how different task manipulations influence different retrieval modes, beyond just labeling the manipulations with the catch-all term *interference*.

In the present article we have reported results suggesting that implicit retrieval of AG knowledge is based on processing that is fast and relatively undemanding of attentional resources, when compared with the explicit retrieval of AG knowledge. This evidence adds to a growing literature (e.g., Manza & Bornstein, 1995; Newell & Bright, 2001) that is consistent with the idea that the type of task used to test for AG knowledge influences what knowledge people reveal and how they reveal it. Further work is needed to expand upon this general theme. For example, we know that when string classification is used as the test of AG knowledge, people are sensitive to a number of sources of information in the AG learning task (see, e.g., Meulemans & Van der Linden, 1997; Pothos & Bailey, 2000). We do not yet

know, however, whether people are sensitive to the same sources of information when other tests of AG knowledge are used. The degree to which the test used to measure knowledge in the AG learning task influences such factors will have implications for how we think about implicit learning and will also address the wider question of how precisely we can ever measure the knowledge acquired by any learning system.

REFERENCES

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long term memory. *Journal of Experimental Psychology: General*, **113**, 518-540.

Berry, D. C., Banbury, S., & Henry, L. (1997). Transfer across form and modality in implicit and explicit memory. *Quarterly Journal of Experimental Psychology*, **50A**, 1-24.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, **106**, 265-289.

Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality & Social Psychology*, **63**, 545-552.

Butler, L. T., & Berry, D. C. (2001). Implicit memory: Intention and awareness revisited. *Trends in Cognitive Sciences*, **5**, 192-197.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, **2**, 406-416.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, **125**, 159-180.

Dienes, Z., & Altmann, G. T. M. (1997). Transfer of implicit knowledge across domains: How implicit and how abstract? In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 107-123). Oxford: Oxford University Press.

Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1322-1338.

Dienes, Z., & Berry, D. {C.} (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review*, **4**, 3-23.

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

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgement: How conscious and how abstract? *Journal of Experimental Psychology: General*, **113**, 541-551.

Gomez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition*, **28**, 253-263.

Gordon, P. C., & Holyoak, K. J. (1983). Implicit learning and the generalization of the mere exposure effect. *Journal of Personality & Social Psychology*, **45**, 492-500.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 501-518.

Helman, S. (2000). *The knowledge and processing underlying the structural mere exposure effect.* Unpublished PhD thesis, University of Reading.

Higham, P. A., Vokey, J. R., & Pritchard, J. L. (2000). Beyond dissociation logic: Evidence for controlled and automatic influences in artificial grammar learning. *Journal of Experimental Psychology: General*, **129**, 457-470.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, **118**, 115-125.

Kinder, A., & Assmann, A. (2000). Learning artificial grammars: No evidence for the acquisition of rules. *Memory & Cognition*, **28**, 1321-1332.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, **207**(1), 557-558.

Manza, L., & Bornstein, R. F. (1995). Affective discrimination and the implicit learning process. *Consciousness & Cognition*, **4**, 399-409.

Manza, L., Hallman, C., McCoy, A., Shumway, E., Ruczhak, M., Laepple, K., & Dangler, R. (1999, April). *Structural mere exposure: The influence of instructional set and stimulus complexity on implicit learning.* Paper presented at the Meeting of the Eastern Psychological Association, Providence, RI.

Manza, L., Zizak, D., & Reber, A. S. (1998). Artificial grammar learning and the mere exposure effect. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 201-222). Thousand Oaks, CA: Sage.

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

Merikle, P. M., & Reingold, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 224-233.

Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1007-1028.

Mulligan, N. W. (1997). Attention and implicit memory tests: The effects of varying attentional load on conceptual priming. *Memory & Cognition*, **25**, 11-17.

Naveh-Benjamin, M., Craik, F. I. M., Gavrilescu, D., & Anderson, N. D. (2000). Asymmetry between encoding and retrieval processes: Evidence from divided attention and a calibration analysis. *Memory & Cognition*, **28**, 965-976.

Newell, B. R., & Bright, J. E. H. (2001). The relationship between the structural mere exposure effect and the implicit learning process. *Quarterly Journal of Experimental Psychology*, **54A**, 1087-1104.

Parkin, A. J., & Russo, R. (1990). Implicit and explicit memory and the automatic/effortful distinction. *European Journal of Cognitive Psychology*, **2**, 71-80.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge. *Journal of Experimental Psychology: General*, **119**, 264-275.

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

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

Reber, A. S., Allen, R., & Regan, S. (1985). Syntactical learning and judgement, still unconscious and still abstract: Comment on Dulany, Carlson and Dewey. *Journal of Experimental Psychology: General*, **114**, 17-24.

Reingold, E. M., & Toth, J. P. (1996). Process dissociations versus task dissociations: A controversy in progress. In G. Underwood (Ed.), *Implicit cognition* (pp. 159-202). Oxford: Oxford University Press.

Roediger, H. L., III, & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63-130). Amsterdam: Elsevier.

Schacter, D. L., Bowers, J., & Booker, J. (1989). Intention, awareness and implicit memory: The retrieval intentionality criterion. In S. Lewandowsky, J. C. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 47-65). Hillsdale, NJ: Erlbaum.

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

Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments, & Computers*, **30**, 583-591.

Turner, C. W., & Fischler, I. S. (1993). Speeded tests of implicit knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1165-1177.

Whittlesea, B. W. A., & Price, J. R. (2001). Implicit/explicit memory versus analytic/nonanalytic processing: Rethinking the mere exposure effect. *Memory & Cognition*, **29**, 234-246.

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

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality & Social Psychology*, **9**(2, Pt. 2), 1-27.

NOTES

1. We use the term *liking* throughout the present article when referring to our indirect test. However, we wish to make it clear that we do not necessarily consider *liking* ratings to be based on any kind of affective responding. We simply consider liking ratings to be a useful indirect test of the effect of prior experience, given that there is no need to mention prior experience to the participants.

2. The link to the liking-based AG task is obvious here. Indeed the liking-based discrimination of grammatical and ungrammatical strings in the AG task has been termed the *structural mere-exposure effect*. We have chosen not to use this term, since it implies substantial overlap with the mere-exposure literature—a debate we do not wish to deal with in the present article. We thank Bruce Whittlesea for helpful comments regarding this issue during a review of an earlier draft of this manuscript.

3. The reason for the instruction to "base liking ratings on similarity," rather than just to rate similarity per se. was that it kept the instructions between the two retrieval conditions as similar as possible. We do not make any claims here as to how the participants in the explicit condition actually interpreted this instruction. We claim only that the instruction was designed to encourage the use of previous experience and to not differ in any other way from the implicit instruction.

4. The fact that this interaction was very close to being significant was due to the unusually low discrimination performance in the no-deadline implicit condition. A Tukey's HSD test on these participants' data indicated that grammatical strings, although given significantly higher liking ratings than UB strings, were not rated significantly higher than UP strings. Thus, in terms of grammatical/UP discrimination, we almost have a crossed dissociation between explicit and implicit retrieval in Experiment 2, with explicit retrieval suffering under a short deadline and implicit retrieval being enhanced. However, this result should be treated with caution. The liking-based discrimination of grammatical and ungrammatical strings has been replicated many times with many different grammars (e.g., Helman, 2000; Manza & Bornstein, 1995; Manza et al., 1999). It seems more likely that the failure, in the implicit condition in Experiment 2, to discriminate novel grammatical and UP strings under no deadline is an anomaly.

5. According to Merikle and Reingold (1991), the discovery that performance on an indirect test is greater than performance on a direct test can be used as evidence that unconscious processing is involved. We accept that some of our data are consistent with this view but do not feel that we have sufficient checks on the conscious/unconscious nature of processing in the present study to comment further.

6. Interestingly, there is previous work that has indicated that this might also be the case at encoding. Dienes et al. (1991) showed that knowledge of illegal bigrams was explicitly recalled equally well under single- and dual-task learning conditions, whereas explicit retrieval of positional knowledge about legal bigrams was lessened by a dual task during learning.

7. A reviewer pointed out that this might be problematic if we assume that the ratings given to UB and UP strings are effectively a *baseline* against which the ratings given to grammatical strings are to be compared. However, we do not see this as a necessary assumption. In the present article, we are interested only in the discrimination of grammatical, UP, and UB strings and how this varies with task and conditions. We have not made any prediction as to how any dissociation in discrimination might manifest itself in the data.

APPENDIX
Learning, Recognition, and Test Strings for Experiments 1 and 2

Learning Strings	Recognition Distractor Strings	Grammatical Test Strings	UP Strings	UB Strings
VTHQX	QRHQX	MTHQX	QTHQX	QMHQX
MXHSX	QRLMR	MXHST	MXHSR	MXHSL
VTLSX	MXHMR	VXLSX	VRLSX	VHLSX
MTHST	SRLVR	MTHQT	MTHVT	MTHSQ
MXLMR	MTHVR	VXLMR	VKLMR	MVLMR
MTLVR	MTLST	MTLVK	MTLVX	MTLVQ
VXHQT	QKHST	VXHST	VXHMT	VXHQS
VXLVK	QKLVK	VTLVK	SXLVK	VMLVK
SRLQX	SKHVK	QRLQX	MRLQX	SHLQX
QRLQT	SKLQT	QKLQT	MKLQT	QHLQT
SRHVR	VTHQT	SRHMR	SRHQK	SRHVL
SKLST	VTLMK	SKLSX	SKLMX	SKLMH
QRHMR	VXHMK	QKHMR	QXHMR	QVHMR
QKHVK	MXLSX	QKHMK	QKHQR	QKHQL
SKHMK	SRHSX	SRHMK	STHMK	MLHMK
QKLMK	VXLQX	QKLMR	QKLSK	QKLMS