Journal of Experimental Psychology:
Learning, Memory, and Cognition
1995, Vol. 21, No. 5, 1322–1338

0278-7393/95/$3.00

Unconscious Knowledge of Artificial Grammars Is Applied Strategically

Zoltán Dienes, Gerry T. M. Altmann, Liam Kwan, and Alastair Goode
University of Sussex

The criteria by which incidentally acquired knowledge of an artificial grammar (A. S. Reber, 1967) could be unconscious was explored in 5 experiments. Participants trained on an artificial grammar lacked metaknowledge of their knowledge: Participants classified substantially above chance even when they believed that they were literally guessing, and, under some conditions, participants' confidence in incorrect decisions was just as great as their confidence in correct decisions. However, participants had a large degree of strategic control over their knowledge: Participants trained on 2 grammars could decide which grammar to apply in a test phase, and there was no detectable tendency for participants to apply the other grammar.

How to decide whether knowledge is unconscious or not is a problem that has intrigued psychologists for many years (e.g., Berry & Dienes, 1993; Cheesman & Merikle, 1984; Dulany, Carlson, & Dewey, 1984; Holender, 1986; Jacoby, 1991; Reber, 1967, 1989; Reingold & Merikle, 1988). Several criteria of the divide between the conscious and the unconscious have been proposed as distinguishing between qualitatively different types of knowledge. We focus on two criteria. First, knowledge might be unconscious in the sense that participants do not know that they have it. That is, participants might lack metaknowledge about their knowledge. This criterion has been investigated in the case of subliminal perception by Cheesman and Merikle (1984, 1986) and in the case of artificial grammar learning by Chan (1992). Second, knowledge might be unconscious in the sense that it may be applied regardless of the participants' intentions to use it. Intriguing methodology for investigating this criterion in the case of implicit memory has been introduced by Jacoby (1991). We apply both criteria to the knowledge acquired about an artificial grammar (Reber, 1989). Initially, we describe the methodology for investigating the two criteria. Then we describe the artificial grammar learning paradigm and discuss existing evidence on the sense in which knowledge of artificial grammars might be unconscious. Finally, we indicate how the methodology for investigating the two criteria can be applied to artificial grammar learning.

Zoltán Dienes, Gerry T. M. Altmann, Liam Kwan, and Alastair Goode, Laboratory of Experimental Psychology, University of Sussex, Brighton, England.

Preparation of this article was supported by a Medical Research Council grant. We thank Lee Brooks, Axel Buchner, Al Parkin, Josef Perner, and Bruce Whittlesea for valuable discussion; Mark Round for preparing the stimuli for Experiment 1 and running a pilot experiment on which it was based; and Chris Hall and Sarah Stockton for testing participants in Experiment 5.

Correspondence concerning this article should be addressed to Zoltán Dienes, Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex BN1 9QG England. Electronic mail may be sent via Internet to dienes@epunix.biols.susx.ac.uk.

Unconscious Means Lacking Metaknowledge

Cheesman and Merikle (1984, 1986) argued that a threshold of consciousness defined subjectively in terms of confidence may be psychologically interesting. They initially gave participants a sequence of trials in which one of four stimuli were presented. For each trial participants said which of the stimuli were presented on that trial. Participants also provided an estimate of their detection performance that varied from claimed random guessing to complete confidence. The subjective threshold was defined as occurring at the level of discriminative responding for which participants claim not to be able to detect perceptual information (i.e., they claim to be literally guessing). Cheesman and Merikle found that when participants believed they were guessing, they were actually discriminating significantly above chance. Thus, participants did not know that they knew what the stimulus was. Further, stimuli presented under these conditions also produced reliable semantic priming. That is, when participants believed that they did not know what a stimulus was, the identity of the stimulus nonetheless affected their behavior, providing an example of unconscious knowledge according to this criterion.

Theoretically, however, it can be argued that the finding of discriminative responding and priming below subjective threshold is, by itself, of no more value than a curiosity to psychologists unless perception below as compared with above subjective threshold has qualitatively different consequences. In that case, theoretical work needs to be done to explain the differences. In fact, Merikle (1992) summarized evidence that the subjective threshold is a psychologically real one in the case of subliminal perception; perception above and below the threshold is qualitatively different. For example, Cheesman and Merikle (1986), using a Stroop-priming task, found that only information above rather than below subjective threshold was spontaneously used by participants for making predictions.

Unconscious Means Lacking Intentional Control

Jacoby (1991) argued that unconscious knowledge was knowledge that was applied regardless of, or contrary to, the participants' intentions; conscious knowledge was applied according to the participants' intentions. Thus, if knowledge is applied contrary to the participants' intentions, it must have been unconscious. Jacoby showed that according to this framework, memory for specific episodes could be unconscious. In one study (Jacoby, Woloshyn, & Kelley, 1989, Experiment 2), participants were asked to study a list of names and they were correctly told that these names were of nonfamous people. Participants were later given a list of famous and nonfamous names, including some from the study phase, and they had to indicate which names were famous. Recognizing a name as earlier read allowed participants to be certain that it was nonfamous. Jacoby et al. argued that in this situation unconscious memory leading to a sense of familiarity (which would tend to make a participant say "famous" for an old study list name) was put in opposition to conscious recognition (which would enable the participant to correctly say "nonfamous"). When participants could give full attention to the study and test phases, old as compared with new nonfamous names were less likely to be mistaken as famous (presumably because of conscious recognition). If participants divided attention during the study or test phases, however, old rather than new nonfamous names were more likely to be mistaken as famous (presumably because of unconscious familiarity). That is, the knowledge that they had seen the name before was applied even though this was contrary to their intentions.

Jacoby's (1991) criterion is a useful one because it appears to separate qualitatively different types of knowledge. For example, on the one hand, conscious recognition but not familiarity (as measured by Jacobys' techniques) is reduced by divided rather than full attention (Jacoby, 1991; Jacoby, Toth, & Yonelinas, 1993; Jacoby et al., 1989), as described above; on the other hand, familiarity but not recollection in a stem-completion task is increased by reading rather than by solving anagrams in the study phase (Jacoby et al., 1993). Note that the conscious and unconscious processes demarcated by Jacoby's criterion are affected by variables (divided attention and perceptual similarity) traditionally associated with the distinction between conscious and unconscious (or automatic).

Both metaknowledge and intentional control capture some aspect of the everyday notion of consciousness, but we do not claim that either fully captures our notion of consciousness. Further, the everyday notion might not be picking out a single natural kind; the criteria of metaknowledge and intentional control might demarcate different subsets of knowledge. For example, Schacter, Bowers, and Booker (1989) argued that whereas simple priming in stem completion involved both unintentional retrieval and unawareness of remembering, priming of new associations in stem completion involved unintentional retrieval but awareness of remembering (see also Bowers & Schacter, 1990; Richardson-Klavehn, Lee, Joubran, & Bjork, 1994). In this article we explore whether criteria for consciousness involving metaknowledge and intentional control give the same or different answers when applied to artificial grammar learning.

Learning Artificial Grammars

Artificial grammar learning is a paradigm that has generated considerable controversy over the degree to which the acquired knowledge is unconscious (e.g., Dulany et al., 1984; Perruchet & Pacteau, 1990; Reber, 1967, 1989; Shanks & St. John, 1994). Two examples of an artificial finite state grammar are shown in Figure 1. For each network, every time that a transition is made between nodes, a letter is produced. Strings of letters that can be made by traversing the network in this way are grammatical according to that network; all other strings are nongrammatical. In a typical experiment, participants are initially shown a set of grammatical strings and asked to memorize them. Participants are then told that the strings of letters obeyed a complex set of rules and are shown novel strings, only half of which obey the rules. Participants' classification of these strings is above chance (typically 60%–70%), even though participants have difficulty describing in free report how they classified (Dienes, Broadbent, & Berry, 1991; Mathews et al., 1989; Reber, 1967). The difficulty participants have in free report indicates that the knowledge about the grammar might be unconscious. However, free report is hardly an exhaustive measure of what the participants might consciously know, so it cannot provide compelling evidence in itself. Participants can say quite a lot about the grammar when probed with various forced-choice tests. For example, Dulany et al. showed that participants could underline that part of a test string that made it grammatical or nongrammatical. Perruchet and Pacteau showed that participants could indicate which bigrams are allowed by the grammar and which are not. Finally, Dienes et al. showed that participants could say which letters are allowed in any part of a string. Both Dulany et al. and Perruchet and Pacteau interpreted their results as indicat-

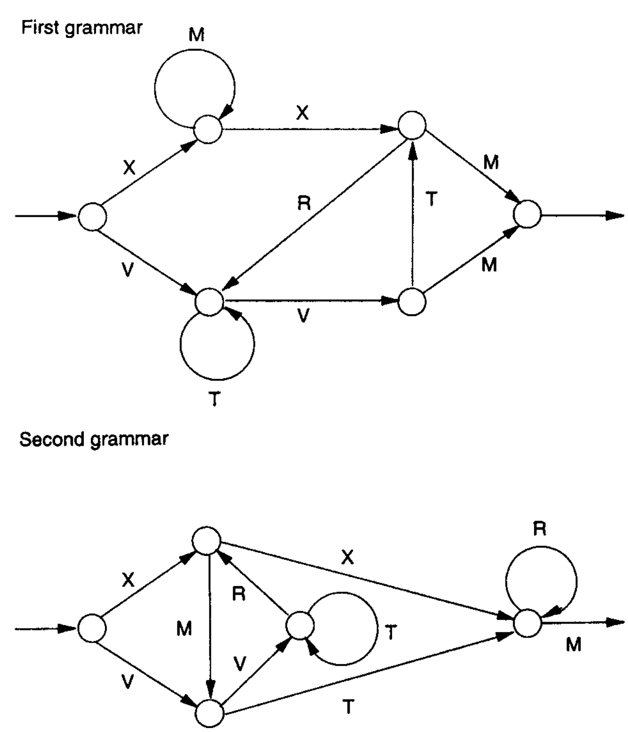

Figure 1. The two grammars used in Experiments 1–4. The top grammar was always presented to participants first, and the bottom grammar second.

ing that the knowledge underlying classification performance was conscious. The results of Dulany et al. and Perruchet and Pacteau certainly help to characterize what it is that participants know, but they leave open the question of whether the knowledge is unconscious according to the criteria of lack of metaknowledge and lack of intentional control. It is these two aspects of our notion of unconscious that we investigate in this article.

Metaknowledge in Artificial Grammar Learning

The level of participants' metaknowledge in artificial grammar learning was first investigated by Chan (1992). After each classification decision, Chan asked participants (N = 28) to rate how confident they were that the decision was correct. He found that when participants were trained in the normal incidental way, they were just as confident in incorrect decisions as in correct decisions. The correlation between confidence and accuracy was .20 (Chan's Experiments 6 and 7). When participants (N = 28) were asked to look for rules in the training phase, the overall level of accuracy was the same and the variances for accuracy and confidence were the same. However, they were now more confident in correct rather than in incorrect decisions, and the correlation between confidence and accuracy was .54 (significantly greater than the value for incidental participants). That is, incidentally trained participants lacked metaknowledge about their knowledge, but explicitly trained participants did not. Chan also found that the magnitude of the correlation depended on the type of stimuli used. For example, if incidentally trained participants were asked to rate bigrams for grammaticality, then confidence was correlated with accuracy (r = .47). Chan concluded that participants acquired explicit knowledge of bigrams. Manza and Reber (1994), using stimuli different from Chan's, found that confidence was reliably higher for correct rather than incorrect decisions. The conditions that gave rise to metaknowledge in the Manza and Reber case are not clear. In the experiments reported in this article we explored some of the conditions that either do or do not give rise to metaknowledge according to Chan's criterion in artificial grammar learning tasks.

Note that the criterion introduced by Chan (1992; we call this the *zero correlation criterion*) is different from the one used by Cheesman and Merikle (1984) for subliminal perception (we call this the *guessing criterion*). A lack of relationship between confidence and accuracy (the zero correlation criterion) implies that participants should be accurate when they believe they are guessing (the guessing criterion), but this has not explicitly been tested. The two criteria also divide knowledge into different subsets: According to the guessing criterion, it is only when participants say "guess" that the knowledge is unconscious; according to the zero correlation criterion, even knowledge leading to confident responses might be unconscious if confidence is not related to accuracy. The way to resolve the question of which criterion is the most useful is to see which criterion divides knowledge into subsets that behave in qualitatively different ways. We attempt to make a step in that direction within this article.

Both the guessing and zero correlation criteria tap only one of two possible forms of metaknowledge. Dienes and Perner (in press) argued that full metaknowledge requires both representing oneself as being in the possession of certain propositional content (they called this type of representation *content explicit*) and also representing an appropriate propositional attitude toward that content—namely, representing the content as knowledge and not, for example, just as confabulation (they called the second type of representation *attitude explicit*). Knowledge could be unconscious according to the metaknowledge criterion either because content explicitness was lacking or because attitude explicitness was lacking. Finding that confidence is unrelated to accuracy, or that participants are accurate when they believe that they are guessing, shows that participants' representations lack attitude explicitness. However, they may have content explicitness, as in the possible case of accurate verbal reports offered as guesses. Reber (1993, p. 136) suggested that participants in the Manza and Reber (1994) experiments showed the converse; that is, they seemed to know that they knew something, even though they did not know what it was that they knew. Both types of lack of metaknowledge are possible, and even plausible, in different circumstances. We only investigated whether participants' representations can lack attitude explicitness.

Intentional Control in Artificial Grammar Learning

Knowledge of an artificial grammar might be unconscious in the different sense that the knowledge is applied in an obligatory way that makes grammatical items look "good" or familiar, regardless of whether the participant wants to apply the knowledge (cf. Buchner, 1994). If this were the case, then a participant trained on two grammars, A and B, might have difficulty distinguishing them at test. If the participant wanted only to say yes to the items of Grammar A, the familiarity of the B items might also elicit yes responses, indicating knowledge application contrary to the participants' intentions. Note that the previous findings of lack of metaknowledge in artificial grammar learning experiments (Chan, 1992) do not rule out participants' strategic control of the knowledge. For example, one could control which particular grammar to use by remembering some of the training strings or the context in which they were learned, without knowing the structure of that knowledge or even that one had the knowledge.

In two previous studies, researchers have investigated participants' ability to selectively apply different grammars (Brooks, 1978; Whittlesea & Dorken, 1993). Brooks generated strings from two different grammars (Grammar A and Grammar B) and used them in a paired-associate learning paradigm in which strings were paired with English words. The strings of the two grammars could be distinguished in a nonobvious way: whether they were paired with a word describing a New World or Old World entity. Participants subsequently informed of the distinction could successfully classify strings as belonging to Grammar A, Grammar B, or neither. Participants not informed of the distinction could not classify the test strings. These results indicate that under these conditions participants do have at least partial intentional control. A test item did not just feel familiar or feel good regardless of the grammar; it could be placed in a relevant category according to the participants' intentions. However, the results did not indicate whether there was also some knowledge that applied regardless of intentions.

Whittlesea and Dorken (1993) also trained participants on exemplars from two grammars. The exemplars were distinguished by the task that participants performed on them: For one grammar, participants pronounced the exemplars; for the other grammar, participants spelled them. Participants were later asked to pronounce or spell test items and then classify these strings as either belonging to the pronouncing grammar or to the spelling grammar. Participants could discriminate the grammars reliably above chance. Although this result provides prima facie evidence for intentional control, Whittlesea and Dorken argued that discrimination was achieved simply by feelings of familiarity induced by task context. Because processing is more fluent when a test experience perceptually matches representations of prior experiences (Jacoby, Kelley, & Dywan, 1989), fluency or familiarity could be used to discriminate the grammars. A test item belonging to the spelling grammar if spelt at test would seem familiar, and thus the participant could correctly infer that it did belong to the spelling grammar; however, if pronounced at test it would seem unfamiliar, and the participant could correctly infer that it did not belong to the pronouncing grammar. That is, externally controlled task demands may have controlled which grammar applied and not the participants' intentions. There was evidence that which grammar applied was partly insensitive to participants' intentions, depending instead on the type of processing engaged in: When test strings were common to both grammars, participants tended to classify them as belonging to the spelling grammar if they spelt them and as belonging to the pronouncing grammar if they pronounced them.

Just as Whittlesea and Dorken's (1993) results do not indicate that participants had strategic control, they also do not indicate that knowledge applied, regardless of participants' intentions. Two types of lack of intentional control can be distinguished: failing to apply knowledge that one wants to and failing to stop the application of knowledge that one does not want applied. Contextual mismatch presumably makes it more difficult to retrieve knowledge, which implies lack of complete strategic control over the retrieval process. This is illustrated by the numerous examples of encoding specificity (e.g., Tulving, 1983). However, if it were to be argued that knowledge applies regardless of intentions, it has to be shown that when only intentions are manipulated the knowledge is still applied (see Dienes and Perner, in press, for further discussion).

Overview of Experiments

Because intentional control logically does not imply metaknowledge, nor vice versa, both criteria need to be explored. Table 1 summarizes the definitions of the criteria and the main results of the research reported in this article. We report four experiments in which we had investigated the extent of participants' metaknowledge according to both the guessing and zero correlation criteria. We investigated the criterion of intentional control in two different ways. In Experiments 1–4 we investigated the extent of participants' intentional control by training participants on two grammars and asking them later to distinguish them. In Experiment 5 we investigated the extent of participants' intentional control by testing participants in an apparently different experiment to the study phase. If participants do not realize that knowledge is relevant, they should not apply knowledge that is conscious according to the intentional control criterion.

In Experiments 1 and 2 we also used a secondary task methodology inspired by Jacoby's (1991) work. As described above, Jacoby found that a previous study episode made a word seem familiar even when this was contrary to the participants' intentions. The performance of a secondary task at study did not influence familiarity, but it did harm explicit recollection. These results suggest that a qualitative difference between knowledge applied according to intentions and knowledge applied regardless of intentions is that only the former is affected by secondary tasks. Thus, in Experiments 1 and 2, some participants performed a secondary task at testing, and

Table 1
Definition Summaries for Criteria of Consciousness and the Results of Experiments 1–5

Criterion	Description	Experiment(s): Finding evidence for such knowledge	Experiment(s): Failing to find evidence
	Lack of metaknowledge[a]		
Guessing	Accurate responding when participants believe they are literally guessing	Exps. 1, 2, 3, 5	None
Zero correlation	Confidence in incorrect decisions as high as confidence in correct decisions	Exps. 1, 3, 5	Exp. 2
	Obligatory application[b]		
Method			
Two grammar	Participants unable to choose which of two grammars to apply	None	Exps. 1–4
Relevance	Participants apply knowledge even though they do not know it is relevant	None	Exp. 5

[a]Participants lack knowledge that they have knowledge. [b]Knowledge applies regardless of intentions.

other participants did not. Dienes et al. (1991) showed that random number generation during learning of an artificial grammar interfered with later classification performance, so random number generation was used as the secondary task in Experiments 1 and 2.

The use of a secondary task at testing allows one to see which of the criteria for lack of metaknowledge carve nature at interesting joints. Consider a case in which under the conditions of full attention during testing, confidence is not related to accuracy. If the zero correlation criterion does the interesting carving, then a secondary task might have an equal influence on low- and high-confidence responses because they belong to the same category. If the guessing criterion does the interesting carving, then a secondary task might have different effects on low- and high-confidence knowledge because they belong to different categories.

Experiment 1

Method

Design. We used a 2 × 2 between-subjects design. Participants were trained first on one grammar and then on another grammar. The first variable, grammar, refers to whether participants were asked to check the first grammar or the second grammar in the subsequent test phase. The second variable, attention, refers to whether participants had full attention in the test phase or divided attention between classifying the grammars and generating random numbers. In addition to these four groups, there was a control group that did not have a learning phase but was given the test items and asked to check all items they thought were grammatical.

Participants. Fifty volunteers from the University of Sussex were randomly assigned to one of the five groups such that each group contained 10 participants.

Stimuli. The two grammars used were taken from Reber (1969) and are shown in Figure 1. The top grammar in Figure 1 was always the first grammar that participants were exposed to, and the bottom grammar was always the second grammar. Both grammars used the same letter set, *M, T, V, R,* and *X*. Also, the legitimate starting bigrams and final letter were the same for both grammars. Each grammar produced 52 strings between five and nine letters in length. Thirty-two representative sequences were selected from each grammar to form the training sets shown in the Appendix. Ten ungrammatical strings were constructed for each grammar by beginning with a legal bigram for each string, but then jumping between nodes at two points through the finite state grammar, and finishing with the only legitimate final letter. These 20 ungrammatical strings were randomly mixed with the remaining 20 strings from each of the grammars, giving a test set of 60 strings that are shown in the Appendix.

Procedure. All participants, except those in the control group, received identical training phases. Participants were initially told that they would be shown two sheets of paper containing strings of letters and that they were to study the strings as carefully as possible. Participants were then presented with the training strings for the first grammar for 7 min. This was then replaced with a sheet containing the training strings for the second grammar, which participants looked at for a further 7 min. In the subsequent test phase, participants were informed that the order of letters in each string was determined by a complex set of rules: one set of rules for the first sheet, and another set of rules for the second sheet. Participants were then given the test sheet with 60 strings on it and informed that a third of the strings were like the strings on the first sheet, a third were like the strings on the second sheet, and a third were like neither, breaking both sets of rules. Half of the participants were asked to check only the strings that were like those on the first sheet, and half of the participants were asked to check only the strings that were like those on the second sheet. In addition, half of the participants gave full attention to the test phase, and these participants were given no further instructions. The other half were asked to generate random digits while performing the test phase. They were asked to produce a digit every 1½ s according to the beat of a metronome. They were told that the sequence of digits should be completely random so that each digit should occur equally often and follow every digit equally often (see Dienes et al., 1991, for complete instructions). All participants were told to write down a confidence rating with each classification response on a scale ranging from 50 (*complete guessing*) to 100 (*complete certainty*). Participants could use any integer in the specified range. Only exactly 50 indicated a literal guess.

Control participants received no training phase. They were given the test sheet and told that the order of letters in some of the strings obeyed a complex set of rules and the order of letters in other strings broke these rules. They were asked to check only the strings that obeyed the rules.

Results

Strategic knowledge. If participants have some intentional control over the application of their knowledge, then they should be able to check more strings from the grammar they were asked to check (the consistent grammar) than from the other grammar (the inconsistent grammar). The proportion of inconsistent strings checked (i.e., number of checks for the inconsistent grammar divided by 20) was subtracted from the proportion of consistent strings checked to give a measure of strategic knowledge. A score of zero would indicate no strategic knowledge, and a score of one would indicate complete strategic knowledge (i.e., perfect knowledge and complete strategic control over it). Table 2 displays the means for the different groups. A 2 × 2 (Grammar [first vs. second] × Attention [full vs. divided]) between-subjects analysis of variance (ANOVA; not including the control participants) indicated only a significant effect of attention, $F(1, 36) = 5.10, p = .03$, $MSE = 0.030$. That is, participants had less strategic knowledge in divided (.25) rather than full attention (.37) conditions. The t tests for each of the four groups indicated that all means were significantly greater than zero (all $ps < .005$). That is, all groups had a significant amount of strategic knowledge.

Obligatory knowledge. If participants had some knowledge not under intentional control, then they may tend to check the inconsistent grammatical strings to a greater extent than the ungrammatical strings. As indicated in Table 3, the total number of checks for the control groups differed from the experimental groups, $t(20) = 3.62, SE = 2.62, p < .01$, in which degrees of freedom were calculated with the Welch–Satterthwaite procedure. For this reason, we report a proportional measure for obligatory knowledge: the number of checks that participants gave to inconsistent grammatical strings was divided by the number of checks given to inconsistent grammatical strings plus the number of checks given to ungrammatical strings. The expected chance value of this measure of obligatory knowledge was .50. The means are displayed in Table 2. A 2 × 2 (Grammar [first vs. second] × Attention [full vs. divided]) between-subjects ANOVA (not including the control

Table 2
Performance in Experiment 1

Attention	Consistent		Inconsistent		Ungrammatical		Strategic		Obligatory	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Full										
First grammar	.59	.15	.16	.12	.21	.08	.43	.25	.46	.12
Second grammar	.53	.08	.21	.08	.22	.07	.32	.13	.48	.09
Divided										
First grammar	.46	.08	.20	.08	.35	.07	.26	.12	.35	.10
Second grammar	.48	.07	.24	.11	.27	.05	.24	.16	.44	.14
Control group										
First grammar									.43	.15
Second grammar									.50	.07

Note. Attention refers to whether participants gave full attention to the test phase or divided attention during test. Grammar refers to whether participants were asked to respond to the first or the second grammar in testing. Consistent = the proportions of consistent grammatical strings checked; Inconsistent = the proportions of inconsistent grammatical strings checked; Ungrammatical = the proportions of ungrammatical strings checked; Strategic = Consistent − Inconsistent; Obligatory = number of checks to the inconsistent grammar divided by number of checks to the inconsistent grammar plus number of checks to the ungrammatical strings.

participants) indicated no significant effects, $MSE = 0.013$. A baseline measure was calculated from the control group for each grammar: the number of checks to grammatical strings of one grammar divided by (number of checks to grammatical strings of that grammar plus number of checks to ungrammatical strings). The means for the control groups are also shown in Table 2. The *t* tests for each of the four groups comparing their obligatory knowledge to the relevant baseline revealed no significant differences ($ps > .10$). Also, the overall mean of the experimental groups (.43, $SD = .12$) was not significantly different from the mean of the control group (.46, $SD = .09$; $t < 1$). The upper limit of the 95% confidence interval for the difference between the overall mean of the experimental groups and the mean of the control group was .04. That is, if participants had some obligatory knowledge, the amount was small compared with their strategic knowledge.

The guessing criterion. A strategic knowledge score was calculated (a) only for responses for which the participant claimed were literal guesses (low-confidence knowledge) and (b) only for responses for which the participant had at least some confidence (high-confidence knowledge). Let C_l be the number of checks to consistent strings that the participant thought were literal guesses; let I_l be the number of checks to inconsistent strings that the participant thought were literal guesses. The measure of low-confidence knowledge was then $C_l - I_l$. A score of zero would be expected if the participant was literally responding randomly. High-confidence knowledge was calculated in the same way but by using only those responses given a confidence greater than guessing. That is, high-confidence knowledge was $C_h - I_h$. Table 4 shows the means for low- and high-confidence knowledge. The *t* tests indicated that the mean for low-confidence knowledge for each group was significantly different from zero (all $ps < .05$). That is, even when participants believed that they were literally guessing, they were still performing significantly above chance, indicating that some knowledge was unconscious in the sense of the guessing criterion. We used 2 × 2 × 2 (Grammar [first vs. second] × Attention [full vs. divided] × Knowledge Type [low-confidence vs. high-confidence knowledge]) mixed-model ANOVA (not including the control participants) to test the interaction between knowledge type and attention. The interaction was significant, $F(1, 36) = 7.96$, $p = .008$, $MSE = 18.7$, and there were no higher order interactions. The *t* tests indicated that the difference between full and divided attention was not significant for low-confidence knowledge (from 2.4 to 3.9), $p > .05$, but it was significant for high-confidence knowledge (from 5.0 to 1.0), $t(38) = 3.10$, $SE = 1.29$, $p < .005$. That is, a secondary task interfered with high- but not low-confidence knowledge. This interaction was

Table 3
Total Number of Checks in Experiment 1

Grammar	Attention: Full *M*	Full *SD*	Divided *M*	Divided *SD*	Control *M*	Control *SD*
First	19	3	20	3		
Second	19	2	20	1	29	11

Note. Grammar refers to whether participants were asked to respond to the first or the second grammar in testing. Attention refers to whether participants gave full attention to the test phase or divided attention during test.

Table 4
Low- and High-Confidence Knowledge in Experiment 1

Attention	Low confidence		High confidence	
	M	*SD*	*M*	*SD*
Full				
First grammar	2	3	6	6
Second grammar	3	3	4	4
Divided				
First grammar	4	3	1	3
Second grammar	4	3	1	2

Note. The scores are number of checks to consistent strings minus number of checks to inconsistent strings. Attention refers to whether participants gave full attention to the test phase or divided attention during test. Grammar refers to whether participants were asked to respond to the first or the second grammar in testing.

of the cross-over type and cannot be explained by scaling effects.

The number of high-confidence responses given to consistent strings was higher in the full-attention (6.5, $SD = 5.3$) rather than the divided-attention condition (1.6, $SD = 2.7$), $t(38) = 3.68$, $SE = 1.33$, $p < .001$. Thus, low-confidence knowledge may have seemed unaffected by the secondary task simply because there were more opportunities for it to apply under dual- rather than single-task conditions. We conducted a further analysis to see if the secondary task interfered with low-confidence knowledge when it was assessed as a proportion of the number of strings not given a high-confidence response. That is, the proportional measure of low-confidence knowledge was $C_l/(20 - C_h) - I_l/(20 - I_h)$. The proportional measure of low-confidence knowledge in the full-attention condition (.17, $SD = .17$) did not significantly differ from that in the divided-attention condition (.21, $SD = .14$), $p > .10$. Proportional low-confidence knowledge was greater than chance (i.e., zero) in both conditions ($ps < .001$).

The zero correlation criterion. Confidence ratings (for the full 50-point scale) were averaged separately for checks for the consistent and inconsistent grammars in which the former are correct decisions and the latter are incorrect decisions. The difference between confidence for consistent and inconsistent checks was used as a measure of metaknowledge. A complete lack of metaknowledge would give a score of zero. Means are shown in Table 5. A 2 × 2 (Grammar [first vs. second] × Attention [full vs. divided]) between-subjects ANOVA (not including the control participants) indicated no significant effects, $MSE = 155.5$. None of the group means were significantly different from zero. The mean value of metaknowledge was one ($SD = 12$) and also not significantly different from zero ($t < 1$). The upper limit of the 95% confidence interval for this difference was 4.5.[1]

Discussion

Our goals for Experiment 1 were, first, to investigate the extent of participants' metaknowledge in classifying artificial grammars, second, to investigate the extent of participants' intentional control, and, third, to investigate the effect of a secondary task during the test phase on high- and low-confidence knowledge and on strategic and obligatory knowledge.

In terms of participants' metaknowledge, we found the application of knowledge of an artificial grammar to satisfy the zero correlation criterion of unconsciousness in that participants were just as confident in incorrect responses as in correct responses. Some of the knowledge also satisfied the guessing criterion in that when participants believed that they were guessing, they were still performing significantly above chance. The use of a secondary task methodology indicated that the guessing criterion might be the more interesting: A secondary task interfered with the application of high-confidence knowledge but not with the application of low-confidence knowledge, indicating that the criterion might separate qualitatively different types of knowledge.[2] An alternative explanation is that the secondary task has two independent effects: One is to reduce performance, but the other is to independently reduce confidence. That is, in terms of the second effect, however the participant distributes responses to consistent and inconsistent strings, the participant may also simply give a lower confidence for each response. As confidence goes down, and the participant gives fewer high-confidence responses, there would be more opportunities for the participant to produce correct low-confidence responses. However, this explanation predicts a decrease in the amount of low-confidence knowledge expressed as a proportion of the number of opportunities (i.e., the number of strings minus the number of high-confidence responses). In fact, there was no detectable tendency for a proportional measure of low-confidence knowledge to decrease. That is, the data were not consistent with independent effects on confidence and performance, but with a specific effect on high-confidence knowledge.[3]

In terms of strategic control, we found the application of knowledge of an artificial grammar to be largely under intentional control. In fact, there was no evidence of any knowledge that applied contrary to the participants' intentions. The amount of strategic knowledge can be analyzed by using Jacoby's (1991) process-dissociation framework. Assume that

Table 5
Confidence in Consistent and Inconsistent Responses in Experiment 1

Attention	Consistent		Inconsistent		Difference	
	M	*SD*	*M*	*SD*	*M*	*SD*
Full						
First grammar	73	19	68	24	5	16
Second grammar	66	12	69	16	−3	11
Divided						
First grammar	56	10	56	12	0	7
Second grammar	55	12	55	16	1	14

Note. The minimum confidence rating is 50, and the maximum is 100. Attention refers to whether participants gave full attention to the test phase or divided attention during test. Grammar refers to whether participants were asked to respond to the first or the second grammar in testing. Consistent = the mean confidence rating for checks to consistent strings. Inconsistent = the mean confidence rating for checks to inconsistent strings; Difference = Consistent − Inconsistent.

[1] The mean confidence for checks to nongrammatical items was 60 ($SD = 19$). This was also nonsignificantly different from the confidence for checks to consistent items, $t(38) = 1.05$.

[2] Note that this interaction was not due to a floor effect preventing further decrements in low-confidence knowledge: Under divided-attention conditions, low-confidence knowledge was significantly greater than high-confidence knowledge, $t(38) = 3.58$, $p < .001$.

[3] This argument requires that in the full-attention condition there is no relation between confidence and accuracy. This appears to be the case as indicated by the analyses relevant to the zero correlation criterion. When participants say a response is a guess, they are just as likely to be correct as when participants say they have some confidence in a response. If the criterion for saying "guess" was simply shifted by the secondary task, the probability of being correct would still be the same for high-confidence as for low-confidence responses. Thus, if the secondary task impaired all knowledge equally, the probability of being correct would be impaired for low-confidence responses. The proportional measure of low-confidence knowledge indicates that, in fact, low-confidence knowledge was unimpaired. That is, the secondary task does not impair all knowledge equally—it spares low-confidence knowledge.

strategic and obligatory knowledge apply independently of each other, with probabilities S and O, respectively. Then a participant will check a consistent string if (a) the strategic knowledge of that string is applied; (b) the strategic knowledge is not applied but obligatory knowledge is; or (c) neither is applied, but the participant responds on some other nondiscriminating basis (call this the *guessing probability,* G). The participant will check an inconsistent string only under Conditions (b) and (c). Algebraically:

$$p(\text{check/consistent string}) = S + (1 - S)O + (1 - S)(1 - O)G. \quad (1)$$

$$p(\text{check/inconsistent string}) = (1 - S)O + (1 - S)(1 - O)G. \quad (2)$$

Thus, the amount of strategic knowledge, S, is given by

$$p(\text{check/consistent string}) - p(\text{check/inconsistent string}). \quad (3)$$

This is estimated by the strategic knowledge measure given in Table 2. The reduction in strategic knowledge by dividing attention at test conceptually replicated Jacoby's (1991) finding that strategic memory for specific episodes is influenced by a secondary task at test. The results of Experiment 1 further suggest that this decrement might only occur for knowledge associated with high confidence (as discussed above).

We found no evidence in Experiment 1 for any obligatory knowledge; that is, O was not distinguishable from zero. So by the criterion of intentional control, application of artificial grammar knowledge was conscious. The implications of this finding will be more fully explored in the General Discussion section.

Experiment 2

Our goal for Experiment 2 was to provide a replication of the results of Experiment 1. A few minor procedural changes were introduced: In the study phase, strings were presented one at a time on a computer monitor rather than all at once on a sheet of paper and participants were asked to "memorize" the strings rather than to "study them as carefully as possible." We designed the new procedures to minimize more carefully the possibility of participants elaborating nonrepresentative rules during learning (Reber, Kassin, Lewis, & Cantor, 1980). Further, in the test phase, participants were tested with a forced-choice procedure and gave confidence on a 5-point rather than on a 50-point scale. Because the control group in Experiment 1 performed at expected chance values, no control group was tested for Experiment 2.

Method

Design. We used a 2 × 2 between-subjects design: Grammar (first vs. second) × Attention (full vs. divided).

Participants. Forty volunteers from the University of Sussex were randomly assigned to one of the four groups such that each group contained 10 participants. No participant had participated in Experiment 1.

Stimuli. The same learning and test strings were used as in Experiment 1. The 60 test strings were arranged into triplets. Each triplet contained one string from the first grammar, one string from the second grammar, and one ungrammatical string. Sixty triplets were constructed such that each test string appeared in three different triplets, but no two strings appeared in the same triplet more than once. Other than these constraints, allocation to triplets was random.

A Turbo Pascal program controlled the display of instructions, stimuli, and the collection of responses.

Procedure. The training strings for the first grammar were presented one at a time on a computer monitor. Each string was displayed for 3.5 s, and the entire set of strings was shown three times in a different random order each time. The procedure was then repeated with the strings from the second grammar. Thus, the total time of exposure to each grammar was 5.6 min. In the subsequent test phase, participants were informed that the order of letters in each string was determined by a complex set of rules: one set of rules for the first set of strings, and another set of rules for the second set. Participants were then shown each triplet of test strings on the monitor. First grammar, second grammar, and ungrammatical strings appeared equally often in the top, middle, and bottom positions. To choose the top, middle, or bottom string, participants entered one, two, or three on the keyboard, respectively. Half of the participants were asked to choose the string from each triplet that was generated by the same rules as the first set of strings, and half of the participants were asked to choose the string that was generated by the same rules as the second set of strings. In addition, half of the participants gave full attention to the test phase, and these participants were given no further instructions. The other half were asked to generate random digits while performing the test phase, as in Experiment 1. After choosing a string, participants entered a confidence rating on a scale ranging from 1 (*complete guessing*) to 5 (*complete certainty*). Participants could use any integer in the specified range.

Results

Strategic knowledge. The measure of strategic knowledge was the proportion of consistent strings chosen minus the proportion of inconsistent strings chosen. Table 6 displays the means for the different groups. A 2 × 2 (Grammar [first vs. second] × Attention [full vs. divided]) between-subjects ANOVA indicated only a significant effect of attention, $F(1, 36) = 10.93, p = .002, MSE = 0.085$. That is, participants had less strategic knowledge in divided- (.12) rather than full-attention (.42) conditions. The t tests indicated that both full- and divided-attention participants had strategic knowledge significantly greater than zero (both $ps < .05$, one tailed).

Obligatory knowledge. The proportional measure for obligatory knowledge was used, as it had been for Experiment 1: the number of inconsistent strings chosen was divided by (the number of inconsistent strings chosen plus the number of ungrammatical strings chosen). The expected chance value of this measure of obligatory knowledge was .5. The means are displayed in Table 6. A 2 × 2 (Grammar [first vs. second] × Attention [full vs. divided]) between-subjects ANOVA indicated no significant effects, $MSE = 0.023$. The t tests for each of the four groups comparing their obligatory knowledge to chance expected revealed no significant differences ($ps > .10$). Also the overall mean (.49, SD = .15) was

Table 6
Performance in Experiment 2

Attention	Consistent		Inconsistent		Ungrammatical		Strategic		Obligatory	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Full										
First grammar	.58	.17	.22	.13	.20	.07	.36	.30	.49	.14
Second grammar	.63	.17	.18	.15	.19	.07	.45	.30	.43	.20
Divided										
First grammar	.48	.20	.29	.15	.23	.08	.18	.33	.52	.15
Second grammar	.37	.12	.33	.12	.30	.08	.05	.22	.53	.11

Note. Attention refers to whether participants gave full attention to the test phase or divided attention during test. Grammar refers to whether participants were asked to respond to the first or the second grammar in testing. Consistent = the proportion of consistent strings chosen; Inconsistent = the proportion of inconsistent strings chosen; Ungrammatical = the proportion of ungrammatical strings chosen; Strategic = consistent minus inconsistent; Obligatory = inconsistent divided by consistent plus ungrammatical.

not significantly different from chance. The upper limit of the 95% confidence interval for overall mean was .54. That is, if participants had some obligatory knowledge, the amount was small compared with their strategic knowledge.

The guessing criterion. We calculated a strategic knowledge score (a) only for responses for which the participant claimed were literal guesses (low-confidence knowledge) and (b) only for responses for which the participant had at least some confidence (high-confidence knowledge) in the same way as in Experiment 1. For example, if C_1 is the number of consistent strings chosen that the participant thought were literal guesses, I_1 the number of inconsistent strings chosen that the participant thought were literal guesses, then the measure of low-confidence knowledge was $C_1 - I_1$. Table 7 shows the means for low- and high-confidence knowledge. The overall level of low-confidence knowledge (2.0, $SD = 5.3$) was significantly above chance, $t(36) = 2.31$, $SE = 0.87$, $p = .03$. We used a 2 × 2 × 2 (Grammar [first vs. second] × Attention [full vs. divided] × Knowledge Type [low-confidence vs. high-confidence knowledge]) mixed-model ANOVA to test the interaction between knowledge type and attention. The interaction was significant, $F(1, 36) = 19.13$, $p = .0001$, $MSE = 151.8$, and there were no higher order interactions. The t tests indicated that the difference between full- and divided-attention was not significant for low-confidence knowledge (from 2.6 to 1.3), $p > .10$, but it was significant for high-confidence knowledge (from 22.5 to 5.5), $t(36) = 3.21$, $SE = 5.30$, $p = .003$. That is, a secondary task interfered with high- but not low-confidence knowledge.

Table 7
Low- and High-Confidence Knowledge in Experiment 2

Attention	Low confidence		High confidence	
	No.	*SD*	No.	*SD*
Full				
First grammar	2	4	21	18
Second grammar	3	6	24	17
Divided				
First grammar	1	4	10	20
Second grammar	1	6	1	11

Note. The scores are number of consistent choices minus number of inconsistent choices. Attention refers to whether participants gave full attention to the test phase or divided attention during test. Grammar refers to whether participants were asked to respond to the first or the second grammar in testing.

The number of high-confidence responses was not significantly higher in the full-attention (47, $SD = 12$) rather than the divided-attention condition (41, $SD = 18$), $p > .10$.

The zero correlation criterion. Confidence ratings were averaged separately for consistent and inconsistent choices in which the former were correct decisions and the latter were incorrect decisions. The difference between confidence for consistent and inconsistent choices was used as a measure of metaknowledge. A complete lack of metaknowledge would give a score of zero. Means are shown in Table 8. A 2 × 2 (Grammar [first vs. second] × Attention [full vs. divided]) between-subjects ANOVA (not including the control participants) indicated no significant effects, $MSE = 0.497$. The mean value of metaknowledge was .40 ($SD = .70$), which was significantly different from zero, $t(36) = 4.02$, $p = .0003$.

Discussion

Our goal for Experiment 2 was to determine the replicability of the results of Experiment 1. As for Experiment 1, there were three crucial issues: first, the extent of participants' metaknowledge in classifying artificial grammars; second, the extent of

Table 8
Confidence in Consistent and Inconsistent Responses in Experiment 2

Attention	Consistent		Inconsistent		Difference	
	M	*SD*	*M*	*SD*	*M*	*SD*
Full						
First grammar	2.9	0.8	2.2	0.8	0.7	1.0
Second grammar	2.5	0.6	2.0	0.6	0.5	0.6
Divided						
First grammar	2.9	1.2	2.3	1.0	0.6	0.7
Second grammar	2.4	0.9	2.4	0.6	0.0	0.4

Note. The minimum confidence rating is one, and the maximum is five. Attention refers to whether participants gave full attention to the test phase or divided attention during test. Grammar refers to whether participants were asked to respond to the first or the second grammar in testing. Consistent = the mean confidence rating for choices of consistent strings; Inconsistent = the mean confidence rating for choices of inconsistent strings; Difference = Consistent − Inconsistent.

participants' strategic control; and, third, the effect of a secondary task during the test phase on high- and low-confidence knowledge and on strategic and obligatory knowledge.

In terms of strategic control, the results of Experiment 1 were replicated perfectly. Participants could choose which grammar to apply, a secondary task interfered with this capacity, and there was no tendency to apply the knowledge from the inconsistent grammar. In terms of participants' metaknowledge, once again we found some knowledge that satisfied the guessing criterion for unconscious knowledge, and this low-confidence knowledge, unlike high-confidence knowledge, appeared to be unaffected by the secondary task. However, we should note that in Experiment 2, unlike Experiment 1, a floor effect may have prevented a clear effect of secondary tasks on low-confidence knowledge from emerging.

In terms of the zero correlation criterion there was a major discrepancy with Experiment 1: In Experiment 2, participants were more confident in correct rather than in incorrect decisions.[4] Chan (1992) argued that participants may have metaknowledge about their knowledge for some subsets of strings and not others, but in Experiments 1 and 2 we used exactly the same set of strings. One procedural difference between Experiments 1 and 2 was the use of forced-choice tests in Experiment 2 but a response to each string in Experiment 1. Imagine that in Experiment 1 we found no relation between confidence and accuracy because when a participant looks at a string, a response (e.g., grammatical according to the target grammar) comes to mind, but the participant does not know the basis of this response. Thus, confidence would have to be reconstructed by some other means. In Experiment 2, the participant needed to look at each of the three strings in turn before making a decision. What would the participant do if two or more of the strings elicited the response "grammatical" (according to the target grammar)? However participants made their final decision in this case, they may have less confidence in their answer than if only one string elicited a grammatical response. However, on average, participants would also be less accurate when more than one string rather than just one string elicits a grammatical response because in the former case, at least one of the answers must be wrong, but in the latter case, there is no such necessity. That is, according to this argument, the forced-choice procedure may have allowed participants to infer metaknowledge, even though the participants may have been completely ignorant of the basis of the process that produced the grammatical response.

Experiments 1 and 2 also differed in the details of the learning phase and this may have also produced the different results. In Experiment 3 we tested the idea that the test phase was important by testing all participants on the same learning phase as in Experiment 2, but half of the participants responded to test strings one at a time (similarly to Experiment 1), and the other half choose one string from three (as in Experiment 2).

Experiment 3

Method

Design. We used a 2 × 2 between-subjects design: Grammar (first vs. second) × Test (forced choice vs. yes–no). All participants gave full attention to the test phase.

Participants. Forty volunteers from the University of Sussex were randomly assigned to one of the four groups such that each group contained 10 participants. No participant had participated in any previous experiment.

Procedure. Half of the participants received the identical procedure as in Experiment 2. The other half were trained in the same way as in Experiment 2, but in the test phase they saw only one string at a time and had to respond yes (it is like the target grammar) or press the space bar to move onto the next test string. If the participant gave a yes response, they then gave a confidence rating on a 5-point scale.

Results

Because of a computer error there was missing data for 1 participant in the yes–no group.

Strategic knowledge. The amount of strategic knowledge did not differ between forced-choice (.35, $SD = .30$) and yes–no (.34, $SD = .36$) test conditions ($t < 1$). Further, strategic knowledge within each group was significant, $t(19) = 5.24$, $p < .001$, and $t(18) = 4.20, p < .001$, respectively.

Obligatory knowledge. The amount of obligatory knowledge did not differ between forced-choice (.45, $SD = .19$) and yes–no (.47, $SD = .22$) test conditions ($t < 1$). Further, the overall level of obligatory knowledge (.46, $SD = .21$) was not significantly different from chance ($t < 1$). The upper limit of the 95% confidence interval was .53.

The guessing criterion. A strategic knowledge score was calculated only for responses for which the participant claimed were literal guesses. The amount of low-confidence knowledge did not differ between forced-choice (2.00, $SD = 4.26$) and yes–no (1.11, $SD = 2.13$) test conditions ($p > .10$). Further, low-confidence knowledge within each group was significant, $t(19) = 2.11, p < .05$, and $t(18) = 2.27, p < .05$, respectively.

The zero correlation criterion. Confidence ratings were averaged separately for consistent and inconsistent choices. The difference between confidence for consistent and inconsistent choices was used as a measure of metaknowledge. Two participants in the yes–no condition did not make any inconsistent choices, so they had missing data for this metaknowledge measure. The amount of metaknowledge in the forced-choice condition (0.53, $SD = 0.64$) was significantly higher than the amount in the yes–no condition (0.15, $SD = 0.62$), $t(35) = 1.81, SE = 0.21, p < .05$, one tailed. Furthermore, the amount of metaknowledge in the forced-choice condition was significant, $t(19) = 3.79, p < .01$, but the amount of metaknowledge in the yes–no condition was insignificant, $t(16) = 1.01, p > .10$.

Discussion

Experiment 3 replicated both Experiments 1 and 2 in finding substantial strategic knowledge, significant low-confidence

[4] This did not seem to be because Experiment 1 lacked the sensitivity to detect the size of effect found in Experiment 2. The experiments differed in the scale used: In Experiment 1 we used a 50-point scale, and in Experiment 2 we used a 5-point scale. Thus, the scales differed by a factor of 10. The ratio of the standard deviations of total confidence for the two experiments (15.7:.8) was no less than the ratio of 10:1. Thus, to test the difference in metaknowledge between the two experiments, the scores from Experiment 1 were divided by 10. The difference was significant, $t(78) = 2.67, p < .01$.

knowledge, and insignificant obligatory knowledge. Of importance, Experiment 3 illustrated how testing procedures can affect Chan's (1992) measure of metaknowledge. Forced-choice testing procedures may enable participants to infer partially the validity of their knowledge in cases in which they would not have otherwise been able to do so.

Experiment 4

In Experiments 1–3 we had failed to find any evidence for obligatory knowledge. Obligatory knowledge might more readily be detected if participants had more training on the grammar that they were asked to ignore. In this case, it might be more difficult for participants to reject strings from the inconsistent grammar. In Experiment 4 we tested this possibility.[5]

Method

Design. There was only one between-subjects variable: grammar (first vs. second).

Participants. Twenty volunteers from the University of Sussex were randomly assigned to one of the two groups such that each group contained 10 participants. No participant had participated in any previous experiment.

Procedure. The same procedure as in Experiment 2 was followed, with the exception that if participants were asked to choose strings from the first grammar, they were shown each of the learning strings of the first grammar twice (total time exposed to the first grammar was 3.7 min) and each of the learning strings of the second grammar four times (total time exposed to the second grammar was 7.5 min). Conversely, if participants were asked to choose strings from the second grammar, in the training phase they received twice the exposure to the first grammar. Confidence ratings were not taken.

Results

Strategic knowledge. The amount of strategic knowledge did not differ between first-grammar (.23, SD = .37) and second-grammar (.20, SD = .43) conditions ($t < 1$). Further, the overall level of strategic control was significant (.22, SD = .40), $t(18) = 2.42, p < .05$.

Obligatory knowledge. The amount of obligatory knowledge did not differ between first-grammar (.49, SD = .18) and second-grammar (.49, SD = .18) conditions ($t < 1$). Further, the overall level of obligatory knowledge (.49, SD = .18) was not significantly different from chance ($t < 1$). The upper limit of the 95% confidence interval limit was .575, indicating that whatever obligatory knowledge participants might have had, it was not substantial.

Discussion

Despite the fact that participants were trained twice as long on the inconsistent as the consistent grammar, there was no detectable tendency for participants to choose an inconsistent rather than a nongrammatical string. That is, participants had an impressive degree of strategic control over their knowledge.

In Experiment 5 we investigated the issue of strategic control in another way. It might be that knowledge will not apply contrary to the participants' intentions—that is, the application of the knowledge can be effectively suppressed—but it may still be obligatory in the sense that it might apply in the absence of any intention. One might have thought that Experiments 1–4 would pick up on any knowledge of this sort because suppression is most likely to operate in only a probabilistic way (e.g., as assumed in Jacoby's, 1991, equations). However, in Experiment 5 we more directly addressed the issue of obligatory knowledge operating in the absence of intention by testing participants in an apparently completely different experiment to the one in which learning took place. In this case, strategic and obligatory knowledge were not put in opposition to each other. As Jacoby has argued, it can be difficult to disentangle obligatory from strategic knowledge when they are not put into opposition because it is not clear whether the measure of obligatory knowledge is contaminated with strategic knowledge. However, there is one clear outcome: If no knowledge application can be detected, then one can be sure that there was little contamination with strategic knowledge.

In the training phase of Experiment 5, participants were shown a sheet of paper with strings of letters on it and were later asked to recall them. This, apparently to the participants, was the entire experiment. Another experimenter brought the participants to a different room (for the test phase) where they were first asked to classify letter sequences that were in fact generated from the same grammar as the training set. Some of the participants were informed of the connection between the experiments before the test phase (the informed participants), and other participants were not informed until the end of the experiment (the uninformed participants). A control group completed the test phase without experiencing a learning phase. The difference in classification performance between the uninformed and control groups was taken as an indication of obligatory knowledge, and the difference between the informed and uninformed groups was taken as an indication of strategic knowledge.

Experiment 5

Method

Design. There was one between-subjects variable, training group, with three levels: informed, uninformed, and control.

Participants. Participants were undergraduate volunteers from the University of Sussex. Twelve were randomly assigned to the informed group, 12 to the uninformed group, and 10 to the control group. No participants had participated in any of the previous experiments.

Stimuli. The artificial grammar used and the training and test exemplars were the same ones used before by Dienes et al. (1991) and also by Dulany et al. (1984) and Perruchet and Pacteau (1990, Experiments 1 and 3). The strings varied in length between three to six letters. The letters used were *M, T, V, R,* and *X.* Twenty representative grammatical strings were used for the training set. Twenty of the remaining grammatical strings were used in the test set, and 5 of the training grammatical strings were repeated in the test set. Twenty-two of the 25 ungrammatical strings were made by substituting an inappropriate for an appropriate letter in an otherwise grammatical string; the remaining three nongrammatical strings contained multiple violations of the grammar. The position of violation covered letter positions one to six over the 25 strings.

A Turbo Pascal program controlled the display of instructions, stimuli, and collection of responses.

[5] We thank B. Whittlesea for suggesting this experiment.

Procedure. All participants were individually tested. Participants in the informed and uninformed groups were initially recruited to take part in a 15-min memory experiment. They were taken to an office and given a sheet of the 20 training strings. They were told to try to memorize as many of the strings as they could. After 10 min, the sheet was removed, and participants were given 2 min to freely recall as many strings as possible. They were then thanked for their time and told that the experiment was over. As participants were leaving, the second experimenter, passing by the office, asked them if they would like to take part in a 10-min experiment. Every participant agreed to this request. Participants were taken to an experimental cubicle in a different building and sat in front of a PC. All participants initially engaged in a distractor task for 2 min that involved classifying geometric shapes that appeared on the PC monitor. Next, participants were told that they would be shown strings containing the letters *M, T, V, R*, or *X*. They were told that in half of these strings the order of letters obeyed a complex set of rules and that in the other half the rules were broken in some way. Their task was to type *Y* if the string obeyed the rules and *N* if it did not. The informed participants were additionally told that the rules were the same as those used to create the letter stimuli in the first experiment they had participated in. All participants were asked to enter a confidence rating after each judgment on a scale ranging from 50 (*complete guessing*) to 100 (*complete confidence*). Participants were presented with the 50 test letter strings twice in a different random order each time. Control participants were treated exactly like the uninformed participants, but they were not given a learning phase. At the end of the experiment, all participants were debriefed and thanked for their time.

Results

Classification. The mean classification performance of the letter strings for informed, uninformed, and control groups was .62 (SD = .08), .54 (SD = .05), and .52 (SD = .04). A one-way ANOVA was significant, $F(2, 31) = 8.92, p = .0008$, $MSE = 0.0036$. The difference between the uninformed and control groups was not significant, $t(20) = 1.02$. That is, there was no evidence of obligatory knowledge. The upper limit of the 95% confidence interval for this difference was .06. The difference between the informed and uninformed group was significant, $t(22) = 2.93$, $SE = .03$, $p = .008$. That is, participants did have strategic knowledge.

Pattern of classification. Following the analysis of Reber (1989), we calculated (for the informed group) the percentages of responses that were correct twice in a row (CC), correct and then in error (CE), in error and then correct (EC), and in error twice in a row (EE) to the same item. The values for CE and EC were averaged (AV). The means were .47 (SD = .13) for CC, .29 (SD = .12) for AV, and .24 (SD = .05) for EE. A repeated measures one-way ANOVA was significant, $F(2, 20) = 9.56$, $p = .001$, $MSE = 0.017$ (there were missing data for 1 participant). The difference between AV and EE was not significant, $t(10) = 1.42$, but the difference between CC and the average of EE and AV was $t(10) = 3.42$, $SE = 0.02$, $p = .007$. Both of these results replicated those reported by Reber and are consistent with the idea that the informed participants behaved as most participants do in artificial grammar learning experiments.

The guessing criterion. Considering only the informed group, we calculated a classification performance by using only those responses that the participant rated as a literal guess. The mean for the informed group was .56 (SD = .12; there were missing data for 1 participant), and the mean for the control group was .49 (SD = .07). The variability was large because sometimes only a few observations were relevant. In the informed group, the percentage of responses rated as guesses varied from 2% to 100% (M = 30%); in the case of 2%, for example, classification performance was .00. Nonetheless, a Mann–Whitney test indicated a significant difference between the groups, $p = .04$, one tailed, as did a trimmed t test, $t(15) = 2.09$, $p = .027$, one tailed.

Table 9
Confidence in Experiment 5 for the Informed Group

String length	Correct		Incorrect	
	M	*SD*	*M*	*SD*
Three letters	73	12	66	11
More than three letters	66	11	65	12

Note. The minimum score is 50 (guessing), and the maximum is 100 (complete confidence).

The zero correlation criterion. Confidence scores were averaged separately for correct and incorrect responses for the informed group. Participants were consistently more confident when they were correct (68) rather than when they were incorrect (65), $t(11) = 3.39$, $SE = 0.88$, $p = .006$. This result contrasts with the repeated findings of Chan (1992, Experiments 5, 6, 7, and 9), and also of Experiments 1 and 3 (yes–no response condition) of no detectable relationship between confidence and accuracy. In Experiments 1–4 we used different grammars and materials compared with those in Experiment 5, so it may be that some of the stimuli in Experiment 5 could consciously be classified by Chan's criterion. In reports filled out after the experiment, participants often said that they "noticed the string *MTV* and were highly confident of its grammaticality." It may be that three-letter strings were classified in a conscious way by this criterion. (Similarly, Chan, 1992, found that classification of another type of short string, bigrams, produced a strong relationship between confidence and accuracy). In neither Chan's experiments nor in Experiment 1 were three-letter strings used. Thus, we conducted a post hoc analysis separating out performance on three-letter strings and all other strings. The means are shown in Table 9. A 2 × 2 (Response Type [correct vs. incorrect] × String Length [three letters vs. more than three letters]) ANOVA indicated a significant interaction between string length and response type, $F(1, 10) = 5.71$, $p = .038$, $MSE = 17.33$. The t tests indicated that the difference between correct and incorrect responses was significant for strings of three letters, $t(10) = 3.40$, $SE = 2.06$, $p = .007$, but not for strings of more than three letters, $t(10) = 1.58$. That is, participants appeared to have metaknowledge only for short strings. Note that participants had substantial knowledge of both types of strings: Mean classification performance for strings of length three was .64 (SD = .18), and for longer strings it was .61 (SD = .07). The difference in classification performance was not significant ($t < 1$).

We conducted a more sensitive test of metaknowledge for strings greater than a three-letter length.[6] If participants had some metaknowledge, they may have been more confident of

[6] We thank A. Reber for suggesting this analysis.

CC strings than of AV or EE strings. The mean confidence ratings for CC, AV, and EE were 69 (12), 64 (11), and 66 (12), respectively. A one-way repeated measures ANOVA was significant, $F(2, 20) = 6.56$, $p = .010$ (with Huyn–Feldt correction), $MSE = 10.371$. The difference between confidence for AV and confidence for EE was not significant, $t(10) = 1.26$, $SE = 1.59$, $p > .10$, but the difference between confidence for CC and the average confidence for AV and EE was significant, $t(10) = 4.07$, $SE = 1.23$, $p = .002$. That is, strings that participants consistently classified correctly were associated with higher confidence ratings than other types of strings, indicating the presence of some metaknowledge.

Discussion

Our goal for Experiment 5 was, first, to investigate the extent of participants' intentional control of artificial grammar knowledge when learning and testing occur in apparently unrelated contexts, and, second, to investigate the extent of participants' metaknowledge. In terms of intentional control, the main finding was that when testing occurred in a different context to learning, it was only those participants that were informed of the connection between testing and learning that showed above-chance classification performance. Participants who were not informed of the connection performed no better than a control group. This finding conceptually replicated the results of Experiments 1–4, which showed that participants can choose which grammar to apply in a test phase.[7]

In terms of metaknowledge, qualified support was obtained for both the guessing criterion and the zero correlation criterion. In the case of the guessing criterion, the results of Experiments 1, 2, and 3 were replicated: Participants could perform above chance when they believed that they were literally guessing. If this criterion were adopted, however, then most of participants' responses were in fact based on conscious knowledge (70% of responses had a confidence greater than guessing). The application of the zero correlation criterion led to more complicated results. A post hoc analysis showed that participants had considerable metaknowledge of strings consisting of only three letters but less metaknowledge of longer strings. The result has fully been replicated with the same materials (Dienes & Altmann, 1993), so some confidence can be placed in it. However, a more detailed probing of the data indicated that even for the longer strings there was some metaknowledge: Participants were more confident in strings that they were consistently correct to than in other sorts of strings. Taken as a whole, the results indicate that level of metaknowledge can vary for different types of stimuli, but it may be difficult to isolate a set of stimuli for which metaknowledge is literally zero (cf. Reber, 1993). There only needs to be one string or part of a string that participants know is (for example) nongrammatical, and also know that they know, for the metaknowledge to be above zero for the set of stimuli as a whole. The reliable variation in metaknowledge was consistent with there being either a continuum with no sharp divisions for different types of knowledge or two distinct types of knowledge that apply in varying proportions for different stimuli sets.

We have isolated a set of strings (i.e., those of length four or greater) for which confidence bears only a small relation to accuracy, so it might be concluded that by the zero correlation criterion, knowledge of the grammar is largely unconscious for these strings (i.e., either close to the unconscious end of a continuum or containing only a small proportion of conscious knowledge). However, what if further subsets of this subset could be taken in such a way so as to produce a substantial correlation? For example, a selection of strings could be taken that participants tended to get correct and also tended to give high-confidence ratings and combine this with a selection of strings for which participants were less likely to get correct and were likely to give low-confidence ratings (cf. Hintzman & Hartry, 1990). In this scenario, a given string could partake in a set of test strings constructed to give no relationship between confidence and accuracy and in another test set constructed to give a relationship. In the former case, the zero correlation criterion would suggest that the string is unconsciously processed, but in the latter case the criterion would suggest that it is consciously processed. If this scenario were possible, one would doubt that the criterion was telling us anything about mental processes.

Dienes (1992) showed that these strings did have a reliable rank ordering in terms of participants' tendencies to give correct responses to them. Thus, for the zero correlation criterion to be upheld, it must not be the case that certain strings reliably elicit low- or high-confidence ratings when a correct response is given to them. If strings do not reliably elicit a certain confidence rating, then sets of strings cannot be constructed to give zero or nonzero correlations between confidence and accuracy at will. Thus, an intraclass correlation, r_i, was calculated on the confidence ratings for the correct responses given by participants to the strings greater than three letters. Because sometimes participants did not give a correct response to a particular string, 25% of the entries in the Subject × String Matrix were missing. We adopted the solution to insert the average confidence for each string in which there were missing values: This solution would overestimate the consistency with which participants gave confidence ratings to particular strings. The obtained value for the correlation was small, $r_i = .16$. This correlation in itself indicates a lack of metaknowledge: Whatever information led participants to consistently answer correctly to some strings rather than others was not being transmitted through their confidence ratings. The correlation also suggests that it would be difficult to construct sets of strings that did give a large correlation between confidence and accuracy because confidence itself has close to zero reliability.

[7] Some proportion of participants in the uninformed group might have noticed the connection with the learning phase, and these participants should have performed like informed participants. A replication of the uninformed group was tested, and participants were asked afterward what they thought the experiment was about and if it had anything to do with any other experiment they had been in. If the learning phase was mentioned, participants were asked to say when they noticed the connection. Out of 18 participants, 10 did not claim to notice a connection at all and their classification performance was .54 (SD = .04), 5 noticed a connection more than half way through testing and their classification performance was .53 (SD = .04), and 3 participants noticed early in the testing and their classification performance was .63 (SD = .03).

General Discussion

In this article we have reported the results of five experiments in which we investigated the extent to which participants are conscious of their incidentally acquired knowledge of artificial grammars. The criteria examined were, first, do participants have metaknowledge about their knowledge (Chan, 1992; Cheesman & Merikle, 1984)? Second, do participants have intentional control over their knowledge (Jacoby, 1991)? In terms of the first criterion, in Experiments 1 and 3 we found conditions under which participants' confidence was not detectably related to their accuracy. Experiment 3 showed that the relationship between confidence and accuracy depended on testing conditions. Specifically, when we asked participants to respond to each string individually, this did not lead to detectable levels of metaknowledge; when we asked participants to choose between a set of strings, this seemed to allow participants to induce metaknowledge in the testing phase. In Experiment 5 we found that confidence was related to accuracy to a greater extent for short rather than long strings, consistent with Chan's previous results, but there remained a small amount of metaknowledge for the long strings. The experiments also showed that a substantial proportion of correct responses occurred when participants believed that they were guessing, satisfying the guessing criterion of unconscious knowledge. Further, this low-confidence knowledge, unlike high-confidence knowledge, was not harmed by a secondary task, suggesting that the guessing criterion really demarcated two different knowledge types. In summary, in terms of the first criterion, there was reliable knowledge about which participants lacked metaknowledge. It might be concluded from Experiments 1–5 that at least some knowledge was unconscious.

In terms of the second criterion, in all of the experiments we found that participants could control which grammar they wished to apply or whether they would apply a grammar. That is, it might be concluded on the basis of this criterion that the knowledge in these experiments was almost entirely conscious.

One of the striking results of these experiments was that different criteria for consciousness gave different results. Presumably, this is because these criteria (concerning metaknowledge and intentional control) indicate different properties of knowledge and its application. As such, the different criteria reveal different things about what it is that the participant is "conscious of." We consider the three criteria—knowledge applied against intentions, confidence unrelated to accuracy, and having no confidence in the response, in that order—to see what properties or contents of participants' knowledge they specify.

Unconscious as Lacking Intentional Control

Taking first the criterion of knowledge applied, regardless of participants' intentions (e.g., Jacoby, 1991), what knowledge might Jacoby's participants be unconscious of? Consider the experiment described earlier in which an episode of seeing a name on a list led to a famous response at test even though if the participant had remembered that the name was on the list she would not have wanted to say "famous" at test. Jacoby argued that unconscious processes were at work in this fame judgment task. However, this result in itself gives no compelling reason to believe that there was any knowledge being applied that the participant was unconscious of. Not being conscious of the fact that the name was on the list needs to be distinguished from applying unconscious knowledge that the name was on the list. It could be assumed that reading the name in the list leads to storing in long-term memory two possible facts: the fact that the name has been seen and the context in which that event occurred. If the first fact is retrieved at test but not the second, then it would not be irrational to sometimes give a famous response. No knowledge or process need be fired off against the participant's intentions; there may only have been some inevitable forgetting. Jacoby's measure of automatic processes (labeled O in the discussion to Experiment 1) could instead be taken as a measure of the extent to which the fact of having experienced names (or other objects) can be retrieved without retrieving the context (specifically, when the name was seen), and the measure of conscious processes (labeled S in Experiment 1) could be taken as a measure of the extent to which the fact of having experienced an object is retrieved with the context. The processes involved in storing and retrieving these facts could be identical.

Part of the power of Jacoby's (1991) approach (and the reason why he called it a process-dissociation framework) is that different experimental variables have been found to affect O and S, respectively. Although some of these dissociations could simply be explained in terms of the fact that familiarity requires retrieval of one type of fact, and recollection requires retrieval of both types, many of the dissociations indicate that the storage or retrieval of context does rely on different principles from the storage or retrieval of other types of information (e.g., Jacoby et al., 1993). Whether there is any evidence that this second type of memory will give an answer, regardless of whether the participant wants it to or not, should be known. This would be the crucial evidence that there is unconscious knowledge according to the criterion of applying regardless of intentions.

Jacoby, Ste-Marie, and Toth (1993) provided relevant evidence. They found that if during test the target item to be responded to is flanked by other items that are to be ignored, participants' reaction times are influenced by the congruence between the flanker and target items. That is, if both the target and flankers were old, or if both were new, then participants responded more quickly than if one was old and the other was new. The fact that the participants processed whether the flankers were old supports the claim that knowledge applies regardless of intentions, if there is confidence that it was not the participants' intention to process the flankers. If knowledge applies regardless of intentions in this case, one suspects that it does in other cases and that other experiments giving nonzero zero values were revealing such knowledge. In any case, this would seem to be an issue in need of further converging evidence.

In the experiments reported in this article we found no evidence for any obligatory knowledge but a large degree of strategic knowledge. In this case, application of Jacoby's (1991) techniques are not problematic: Showing nonzero S clearly shows knowledge under intentional control. So by the criterion of intentional control, application of artificial grammar knowledge was conscious. However, what is it that participants were conscious of? Intentional control only indicates that one is conscious of the highest order chunk controlled by the intentions. In the case of Experiment 1, it can only be said that participants were conscious of which grammar they wished to apply and not that they were conscious of any further aspect of their knowledge (e.g., if it was exemplar or rule based or what rules or what exemplars were used). Testing for the level of detail of participants' conscious knowledge in terms of this criterion would entail testing for application of intentional control over specified details of the participants' knowledge. Shanks and St. John (1994) suggested one way of doing this by asking participants to provide letter continuations but to avoid any continuations that they had seen in the study phase.

In other areas of memory research it is known that reinstating the context of study can increase the tendency of information studied in that context to be retrieved and used (e.g., Baddeley, 1990; Smith, 1979; cf. Whittlesea & Dorken, 1993). In the same way, it might be that the participants in Experiments 1–5 could mentally reinstate the specific context of one grammar to ensure the application of that grammar.

Unconscious as Zero Correlation Between Confidence and Accuracy

Chan's (1992) zero correlation criterion is meant to be a methodology for indicating a lack of metaknowledge. A piece of knowledge can be taken to be unconscious by this criterion if that knowledge is not the object of further knowledge (cf. Rosenthal, 1993). What can be concluded about what participants are conscious of by the application of Chan's criterion? This depends on whether rules are applied deterministically or probabilistically. If a participant applies a set of partially correct rules deterministically, then there is no reason to suppose that the participant should give a more confident response when the rules make a correct classification than when they make an incorrect classification. Even if confidence does not relate to accuracy in this situation, there is no reason to suppose that the participant's rules are not or could not themselves be the object of further knowledge. However, participants do not apply deterministic rules in artificial grammar learning experiments. Reber (1989) pointed out that typically when participants are tested twice on the same strings, the probability of giving one correct and one incorrect answer, $p(\mathrm{AV})$, is very close to the probability of giving two incorrect answers, $p(\mathrm{EE})$. (This result was replicated in Experiment 5.) If the participant consistently applied incorrect rules, $p(\mathrm{EE})$ would exceed $p(\mathrm{AV})$. Reber argued that the actually obtained pattern (i.e., $p[\mathrm{AV}] = p[\mathrm{EE}]$) can be explained by assuming that participants know some exemplars perfectly and guess the rest of the time. A lack of relationship between confidence and accuracy indicates then that participants do not know when they are applying their knowledge (and when they are applying guesses). More generally, a lack of relationship between confidence and accuracy indicates that participants do not know the strengths of their stochastic rules (Dienes & Perner, in press). Thus, finding that artificial grammar knowledge is "conscious" by the criterion of intentional control does not contradict finding that the knowledge is "unconscious" by the zero correlation criterion. The first criterion shows that participants are conscious of (have intentional control over) which grammar to apply; the second shows that participants are not conscious of (lack metaknowledge of) the strengths of the rules of that grammar. Thus, the criteria do not literally contradict each other in terms of what it is they say the participant is conscious of.

Unconscious as Guessing Accurately

In terms of Cheesman and Merike's (1984) guessing criterion, finding that participants can discriminate above chance when they believe that they are guessing shows that participants do not know that they are able to discriminate. That is, participants do not know that they have knowledge. This could be because (a) participants do not know that they have rules that are being applied or (b) participants do not know that the rules are correct. The latter situation occurs when participants do not have insight into the learning process by which veridical knowledge is generated, but they may have insight into the contents of the product of the process (in the terminology of Dienes & Perner, in press, the knowledge is content explicit but attitude implicit). Future research needs to determine the conditions under which the criterion picks out qualitatively different types of knowledge and the sense in which participants lack metaknowledge (i.e., Is it Situation 1 or Situation 2?).

Conclusion

In summary, the results of the experiments reported in this article show that different criteria of consciousness give different answers because they are tapping different contents of knowledge. Thus, the experiments demonstrate that these different contents do not have to all occur at the same time: They can be dissociated.

One response to the exploration of the criteria in this article might be to wonder if the criteria really measured consciousness. However, our aim was not to provide a final definition of consciousness. It was to see if criteria inspired by (but not completely determined by) the everyday notions of consciousness are psychologically interesting; that is, if they divide knowledge into qualitatively different types. The criteria may then pick out examples of knowledge that, in the everyday use of the term, is said to be unconscious. Whether or not a criterion picks out all such examples or only those examples is not so important. What is important is whether the criteria enable the understanding of different knowledge and learning systems.

References

Baddeley, A. D. (1990). *Human memory: Theory and practice.* Hove, England: Erlbaum.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: empirical and theoretical issues.* Hove, England: Erlbaum.

Bowers, J., & Schacter, D. (1990). Implicit memory and awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 404–416.

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

Buchner, A. (1994). Indirect effects of synthetic grammar learning in an identification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 550–566.

Chan, C. (1992). *Implicit cognitive processes: Theoretical issues and applications in computer systems design.* Unpublished doctoral dissertation, University of Oxford, Oxford, England.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics, 36,* 387–395.

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology, 40,* 343–367.

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

Dienes, Z., & Altmann, G. T. M. (1993, July). *Transfer of implicit knowledge across domains.* Paper presented at the Joint meeting of Canadian Society for Brain, Behaviour and Cognitive Science and the Experimental Psychology Society, Toronto, Ontario, Canada.

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

Dienes, Z., & Perner, J. (in press). Implicit knowledge in people and connectionist networks. In G. Underwood (Ed.), *Implicit cognition.* Oxford, England: Oxford University Press.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General, 113,* 541–555.

Hintzman, D., & Hartry, A. (1990). Item effects in recognition and fragment completion: Contingency relations vary for different subsets of words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1029–1039.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parofoveal vision, and visual masking: A survey and reappraisal. *Behavioral and Brain Sciences, 9,* 1–23.

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

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., Ste-Marie, D., & Toth, J. P. (1993). Redefining automaticity: Unconscious influences, awareness, and control. In A. Baddeley & L. Weiskrantz (Eds.), *Attention, selection, awareness, and control: A tribute to Donald Broadbent* (pp. 261–282). Oxford, England: Oxford University Press.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122,* 139–154.

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

Manza, L., & Reber, A. S. (1994). *Representation of tacit knowledge: Transfer across stimulus forms and modalities.* Unpublished manuscript.

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, and Cognition, 15,* 1083–1100.

Merikle, P. M. (1992). Perception without awareness: Critical issues. *American Psychologist, 47,* 792–795.

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

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

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

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

Reber, A. S. (1993). *Implicit learning and tacit knowledge.* Oxford, England: Oxford University Press.

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 492–502.

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics, 44,* 563–575.

Richardson-Klavehn, A., Lee, M. G., Joubran, R., & Bjork, R. A. (1994). Intention and awareness in perceptual identification priming. *Memory & Cognition, 22,* 293–312.

Rosenthal, D. M. (1993). Thinking that one thinks. In M. Davies & G. W. Humphries (Eds.), *Consciousness: Psychological and philosophical essays* (pp. 191–223). Oxford, England: Basil Blackwell.

Schacter, D. L., Bowers, J. S., & Booker, J. (1989). Intention, awareness, and implicit memory: The retrieval intentionality criterion. In S. Lewandowsky, J. 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. *Behavioural and Brain Sciences, 17,* 367–448.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 460–471.

Tulving, E. (1983). *Elements of episodic memory.* Oxford, England: Clarendon Press.

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

(Appendix follows on next page)

Appendix
Stimuli Used in Experiment 1

Strings				
Training		Test		
First grammar	Second grammar	G1	G2	U
xmxrttvtm	xmtrrrm	xxrtvtm	xmvrxrm	xmtvvxrm
vttvtrvm	vvtrxrm	vvtrvtm	xmvrxrrrm	vvtrmm
xmmxrvm	xmtrm	xmxrtvtm	vvtrmtrm	vtxrm
vtvtm	xxrrm	xmmxrvtm	xmvtrmtm	xmtvm
xxrvtm	xmvrmtm	xxrvm	vvrxrrm	xmxvvxm
vtttvtm	vvttrxrrm	xxrvtrvtm	xmtrrm	vvrmtmvm
xxrvtrtvm	xmvrmtrm	xxrtvm	vvttrxm	vvtrmxmm
vvtrttvtm	vvtttrmtm	vtvtrttvm	vvrmvrmtm	xmvxrmvtm
xmxrtttvm	xmvrmvrxm	xmxrttvm	xmvtrxrm	xmtxm
xmmxrtvm	vvttrmtrm	vttvm	vvtrxm	vtrtvtvm
vtvtrtvtm	vvrmvrxrm	vvtrtttvm	vvttrmtm	xxvtxm
vtttvtrvm	vtrrm	xxrtttvm	xmvttrxrm	xmrmxm
xxrvtrvm	vvtrxrrm	vtvtrvtm	xmvtttrxm	xmmxtrmxm
vtvtrvm	xmvtrmtrm	vttvtrtvm	vvtttrxm	vtvmmxtvm
xxrtttvtm	xmvrmtrrm	xmmmxrvm	vvrmtrrm	vvmvtm
vttvtm	vvrxrrrm	vvtrtvtm	xxrrrm	xxrtrmn
vvtrttvtm	xmvtrxrrm	vtttvm	vvtrxrrrm	xxvtvxm
xmmmxrvtm	vtrrrm	xmxrvm	vvrmtm	vvxmm
xmxrvtm	xmvrxm	xmmxrttvm	vvrmvrxm	vvttxxtrm
xxrttvtm	xmvtrxm	xmmmxm	vvrxm	vtrrvvm
vvtrtvm	vvrmtrrm			
xmxrvtrvm	xmvrxrrm			
xmmmxrtvm	vvtrmtm			
vvtrvm	vvrxrm			
vttvtrvtm	xmvttrmtm			
xmxrtvm	vvrmvtrxm			
vvtrvtrvm	vvrmtrm			
xmmxm	vvtrmvrxm			
vtvtrtvm	vvtttrxrm			
xxrttvm	vvtrmtrrm			
xmmxrtvtm	vvttrxrm			
xxrtvtrvm	xmvttrxm			

Note. For the test strings, G1 means the string is grammatical according to the first grammar only, G2 means that the string is grammatical according to the second grammar, and U means that the string is ungrammatical to both grammars.

Received September 14, 1993
Revision received October 11, 1994
Accepted October 18, 1994 ■