

Subjective measures of awareness and implicit cognition

RICHARD J. TUNNEY and DAVID R. SHANKS
University College London, London, England

In this article, we examine whether artificial grammar learning is implicit according to a subjective criterion of awareness based on confidence ratings. In four experiments, participants discriminated between grammatical and ungrammatical sequences in both the same (Experiment 1) and a novel (Experiments 2–4) vocabulary and indicated their confidence in each decision. Replicating earlier studies, confidence judgments reported on a continuous scale (50%–100%) were only weakly related to accuracy, suggesting that learning was implicit. In contrast, confidence judgments reported on a binary scale (high vs. low) revealed that confidence was related to accuracy. We show that participants are better able to place their phenomenal states on a binary scale, as compared with a continuous scale.

The question of whether studies of human learning provide evidence for unconscious cognition is controversial. The issue is important, since many theories whose aim is to unify consciousness and cognition require a degree of sophisticated unconscious processing. Indeed, the introduction to a recent special issue of *Cognition* on the subject of consciousness (Dehaene & Naccache, 2001) lists unconscious processing as one of three fundamental findings that any theory of consciousness must explain. Yet despite over a century of research, unequivocal evidence for unconscious, or implicit, cognition remains elusive. This is partly due to the fact that criteria by which awareness can be *objectively* distinguished from the absence of awareness have failed to meet with universal acceptance. The issue has led a number of workers (e.g., Cheesman & Merikle, 1984; Dienes, Altmann, Kwan, & Goode, 1995; Kunimoto, Miller, & Pashler, 2001) to explore *subjective* measures of awareness instead. Here, we compare two subjective measures of awareness based on reports of confidence and ask whether they provide evidence of implicit cognition in the paradigm that defined implicit learning—artificial grammar learning.

Implicit Learning of Artificial Grammars

Much of the recent interest in unconscious cognition can be attributed to Reber (1967), who coined the term *implicit learning* to describe a form of learning that appears to occur without any awareness of what is being learned. In Reber's (1967) original experiment, participants were asked to study sequences of letters that, un-

known to them, were generated by an artificial grammar (see Figure 1). Later, the participants were asked to discriminate between previously unseen grammatical and ungrammatical sequences. When the test sequences were composed of the same vocabulary elements as the training sequences, classification performance could be as high as 70% correct (Reber, 1967), against a chance level of 50%. When the participants were asked to transfer their knowledge to sequences composed of novel vocabulary elements, classification was reduced but still could be as high as 65% correct (Reber, 1969). Reber observed that, although the participants were able to classify sequences correctly, they were unable to verbally report the basis on which their decisions had been made and, when asked, could not describe the rules of the grammar. The basic form of this experiment has been repeated many times, using a variety of different measures of awareness. However, the claim that learning can proceed without awareness continues to generate controversy because, as with other paradigms, such as subliminal perception, none of the criteria used to measure awareness has met with universal acceptance.

The criteria used to measure awareness fall into *subjective* and *objective* categories. Both sets of criteria are predicated on assumptions about the relationship between a measure of performance, on the one hand, and a measure of awareness for the information that influences performance, on the other. A dissociation between performance and awareness is taken to indicate unconscious cognition. A number of previous studies have demonstrated that artificial grammar learning occurs above objective thresholds of awareness, such as recognition and generation (e.g., Dulany, Carlson, & Dewey, 1984; Gomez & Schvaneveldt, 1994; Meulemans & Van der Linden, 2002; Perruchet & Pacteau, 1990), which rely on measures of discrimination. We do not consider such measures in the present article, since they reveal little evidence of implicit learning.

The research was funded by the United Kingdom Economic and Social Research Council. The work is part of the program of the ESRC Centre for Economic Learning and Social Evolution, University College London. We thank Zoltán Dienes for his helpful comments. Correspondence concerning this article should be addressed to R. J. Tunney, Department of Psychology, Keele University, Keele, Staffordshire ST5 5BG, England (e-mail: r.tunney@keele.ac.uk).

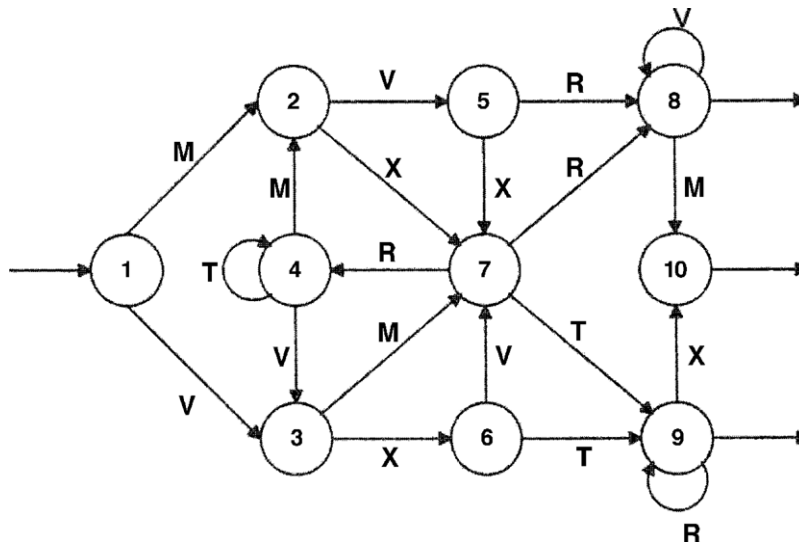


Figure 1. Artificial grammar originally used by Brooks and Vokey. From "Abstract Analogies and Abstracted Grammars: Comments on Reber (1989) and Mathews et al. (1989)," by L. R. Brooks and J. R. Vokey, 1991, *Journal of Experimental Psychology: General*, 120, p. 317. Copyright 1991 by the American Psychological Association. Adapted with permission.

In the case of a subjective measure, participants are asked specifically to respond according to their own internal state of awareness. Reber (1967) employed a subjective measure to assess awareness—namely, verbal introspection. Participants clearly show evidence of learning as expressed by discrimination between grammatical and ungrammatical sequences, but according to this criterion, their knowledge falls below the threshold of consciousness, because it cannot be expressed verbally.

Verbal reports of awareness are frequently criticized because they are susceptible to influence by response bias (e.g., Erdelyi, 1974; Eriksen, 1960; Holender, 1986; Shanks & St. John, 1994). Participants are asked to set their own criterion for awareness and may systematically claim to be unaware when, in fact, they possess a degree of awareness, albeit small, of the knowledge they are using. A participant may set a conservative response criterion and choose to report only knowledge held with high confidence. Knowledge with low confidence may, therefore, not be reported even if it is, in fact, conscious.

An alternative to verbal report as a subjective measure of awareness—and a potential solution to the attendant problem of response bias—is to ask participants to report their phenomenal states by means of confidence ratings. In this case, participants first make a discrimination and then are asked to rate how confident they are that their judgment was correct or, alternately, to estimate their overall level of accuracy. Two measures of awareness based on subjective confidence have been used in studies of unconscious cognition. The first defines knowledge as unconscious when participants believe they are guessing. Dienes et al. (1995) called this the

guessing criterion of awareness. In the earliest reported experiment of this kind, Peirce and Jastrow (1884) suspended weights from each others' fingers and attempted to discriminate changes in weight. Throughout their observations, they noted the degree of confidence with which they made their discriminations on a 4-point scale in which "0 denoted absence of any preference for one answer over its opposite, so that it seemed nonsensical to answer at all; 1 denoted a distinct leaning to one alternative; 2 denoted some little confidence of being right; 3 denoted as strong a confidence as one would have about such sensations." Across several thousands of trials, they observed that they could discriminate changes in weight even when they reported an "absence of any preference for one answer over its opposite, so that it seemed nonsensical to answer at all" (Peirce & Jastrow, 1884, p. 78).

Cheesman and Merikle (1984) also have argued for the use of this criterion in the study of subliminal perception. They estimated participants' awareness thresholds for four color words by reducing the interval between the stimulus and a pattern mask. On each trial, one of the four words was presented, and the participants were asked to identify which word had been presented by pressing a corresponding button. At the end of each block of 24 trials, the participants estimated the percentage of trials on which they thought they had correctly identified the stimulus. Cheesman and Merikle observed that when the participants believed themselves to be guessing (i.e., estimated their percent correct to be 25%), they were in fact identifying 66% of the trials correctly. Perception was, therefore, subliminal according to a criterion based on guessing. However, no priming was ob-

served when identification was at chance. Therefore, perception was not subliminal according to an objective criterion. Cheesman and Merikle concluded that perception could occur below a subjective, but not an objective, threshold of awareness.

Dienes et al. (1995) applied this criterion to artificial grammar learning. In each of their five experiments, participants first studied sequences generated by an artificial grammar and, at test, classified new sequences according to their adherence to the grammar. In the test phase, after each discrimination, the participants gave a confidence rating that ranged from 50 to 100, where 50 was a *complete guess* and 100 was *absolutely sure*. The participants reported a proportion of their responses to be guesses (i.e., their confidence was 50). Dienes et al. found that reliably more guesses were correct than were incorrect or, to put it another way, that classification accuracy based only on guesses was greater than chance. Dienes et al. concluded that artificial grammar learning involves some unconscious knowledge according to a subjective criterion of awareness based on guessing. Similar results have been reported in transfer to novel vocabularies (Dienes & Altmann, 1997) and in amnesic populations (Channon et al., 2002).

As numerous commentators have pointed out (e.g., Erdelyi, 1974; Eriksen, 1960; Holender, 1986), criteria based on guessing, whether derived from verbal reports or confidence ratings, are potentially influenced by response bias: Participants may systematically claim to be guessing when, in fact, they possess a degree of awareness, albeit small, of the knowledge they are using. Indeed, participants are systematically underconfident across a variety of perceptual judgment tasks (see Juslin, Winman, & Olsson, 2000). Thus, in the case of subliminal perception, the experimenter cannot be sure that a stimulus was truly subliminal when participants claim to be guessing. Or in the case of artificial grammar learning, the experimenter cannot be sure that the participant was not aware of the information that he or she used to correctly classify a test item. Participants in these studies have to decide where to place their own criterion for report. The conservative may not report awareness unless absolutely sure, whereas the liberal may report awareness of the merest intuition.

An alternative to the guessing method is to examine whether participants know when they are correct and when they are incorrect. As with the guessing criterion described earlier, participants rate how confident they are that each decision was correct. Chan (1992) suggested that if participants are aware of the knowledge they use to classify sequences, they should be more confident when they make correct decisions than when they make incorrect ones. Thus, confidence should predict accuracy. In Peirce and Jastrow's (1884) experiments, the participants' level of certainty that a weight had changed should be related to the magnitude of change. Conversely, if discriminations can be made without awareness, participants should be as confident when they make incorrect

decisions as when they make correct ones. Dienes et al. (1995) called this the *zero-correlation criterion*. This criterion is related to the concept of *calibration* in probability estimation (Lichtenstein, Fischhoff, & Phillips, 1982; Yates, 1990). When participants can accurately estimate the probability of an event, they are said to be *well calibrated*. When participants over- or underestimate the probability of an event, they are said to be *poorly calibrated*. In studies of both implicit cognition and probability estimation, calibration can be expressed as a correlation between the estimated and the actual probability that events will occur but is more frequently expressed as a difference score, sometimes known as a calibration index (see Yates, 1990).

The data from experiments in which this method has been used suggest that learning can be implicit under some circumstances. For instance, Chan (1992, Experiments 6 and 7) found that when participants were instructed to look for rules in the training sequences, they subsequently showed a reliable correlation between confidence and accuracy, suggesting a degree of explicit learning. In contrast, no such correlation was observed when participants were instructed to memorize training sequences, suggesting that learning had been implicit. Newell and Bright (2002) reported similar results in an implicit invariant learning task. Kelly, Burton, Kato, and Akamatsu (2001) demonstrated that these effects extend beyond laboratory tasks to memory for real-world regularities, such as coins and stamps.

Dienes and Altmann (1997, Experiment 2) examined the relationship between subjective confidence and the ability to transfer knowledge to sequences composed of a novel vocabulary. The issue of transfer has predominantly been concerned with understanding how participants might represent artificial grammars. For some researchers, the ability to transfer knowledge to novel letter sets must indicate that learning involves a large explicit component, because it involves more complex operations than does the standard paradigm (e.g., Redington & Chater, 2002). For others, transfer reflects a larger implicit component than does the standard case, because implicit abstract (form-independent) representations are transferred and explicit concrete representations are not (e.g., Reber, 1969). In the study reported by Dienes and Altmann, when participants classified sequences in the same vocabulary as that used at learning, there was a small but reliable relationship between confidence and accuracy, suggesting a degree of explicit knowledge. In contrast, when the vocabulary of the test sequences was changed, the relationship between confidence and accuracy disappeared (see also Tunney & Altmann, 2001). Dienes and Altmann concluded that the knowledge used to classify sequences in a novel vocabulary was implicit.

Recently, Kunimoto et al. (2001) have proposed a measure of awareness conceptually similar to the zero-correlation criterion, except that instead of making a confidence rating on a continuous scale, after each discrimination participants made a binary confidence judg-

ment to indicate whether their confidence in the accuracy of their discrimination was high or low. If participants are aware of the information they are using to classify sequences, they should respond systematically by reporting high confidence for correct decisions and low confidence for incorrect decisions. Consequently, confidence should be related to accuracy. On the other hand, if participants have no awareness of the information they are using, high and low confidence responses will be randomly assigned to correct and incorrect discriminations. In this case, confidence will be unrelated to accuracy. These confidence judgments can then be categorized in terms of signal detection theory (SDT), according to the scheme shown in Table 1. An important aspect of Kunimoto et al.'s method is that the use of an SDT analysis ensures that the sensitivity of the measure is unaffected by the participants' own report criteria and so is independent of response bias.

When participants are aware of the knowledge used to make a discrimination, they presumably believe themselves to be correct and should respond with high confidence. Thus, correct discriminations made with high confidence may be categorized as hits. But when participants believe themselves to be correct and respond with high confidence when they are, in fact, incorrect, the response is categorized as a false alarm. Participants' awareness of their own performance can consequently be expressed as d' , a measure of sensitivity that is independent of bias or of where participants place the criterion for making high and low confidence judgments.

Kunimoto et al. (2001, Experiment 1) compared their binary rating procedure with the confidence rating procedure reported by Cheesman and Merikle (1984). On each trial, participants were asked to identify which of four briefly presented pattern-masked color words was shown. In one condition designed to replicate Cheesman and Merikle's study, the criterion for awareness was based on guessing. At the end of each block of 24 trials, the participants estimated on a free scale the percentage of trials on which they thought they had correctly identified the stimulus. The threshold of awareness was the stimulus duration at which the participants believed they had correctly identified 25% of the trials (i.e., believed they were guessing). In another condition, the participants estimated their confidence on a binary (high or low) scale following each perceptual discrimination. Kunimoto et al. found that the criterion based on guessing reported by Cheesman and Merikle underestimated awareness, relative to the criterion based on high and low

confidence. For this reason, they concluded that the lack of sensitivity of Cheesman and Merikle's procedure was due to contamination by response bias. Nevertheless, Kunimoto et al. went on to find evidence for subliminal perception even with the binary method.

The use of SDT in the analysis of metacognitive judgments was suggested, but not implemented, by Wellman (1977), who instead used a correlation. Nelson (1984) reviewed a variety of indices that measured the accuracy of binary metacognitive judgments but rejected SDT on the grounds that it was, at that time, untested. The method described by Kunimoto et al. (2001) is, to our knowledge, the first time that this method has been implemented and provides a useful new criterion for assessing awareness in artificial grammar learning. In the experiments that follow, we applied Kunimoto et al.'s technique to artificial grammar learning. Previous experiments in which rating techniques have been used have shown that artificial grammar learning can, under some circumstances, proceed without awareness. However, as Kunimoto et al. demonstrated with respect to subliminal perception, ratings can underestimate levels of awareness.

EXPERIMENT 1

Experiment 1 assessed levels of awareness in artificial grammar learning, using the binary confidence technique developed by Kunimoto et al. (2001). In Experiment 1A, grammaticality and confidence judgments were concurrent.

In earlier experiments that have shown dissociations between confidence and accuracy, the confidence ratings were elicited momentarily after the classification was made and the sequence about which the judgment was being made had disappeared. For example, in all but one of the experiments reported by Dienes et al. (1995), participants first made a decision about each test sequence. After this, the string disappeared, and the participants were asked to rate their confidence on a scale ranging from 50% to 100%. One possibility is that participants' phenomenal state is relatively transient and, during this delay period, decays so that, a few seconds after making an accurate discrimination, awareness becomes undetectable. Consistent with this idea, Eldridge, Sarfatti, and Knowlton (2002) found, in a recognition memory experiment, that remember-know judgments are affected by short delays between a recognition and a remember-know judgment. To examine this possibility, in our Experiment 1B, grammaticality and confidence judgments were consecutive: The participants classified sequences while they were visible, and after this, the sequence disappeared, to be followed 1 sec later by a confidence question.

If artificial grammar learning involves explicit knowledge, this should produce an index of awareness in Experiment 1A greater than chance. If confidence decays rapidly, we would expect the index of awareness in Experiment 1B to approach chance levels.

Table 1
Categorization of Participants' Confidence Responses for a Signal Detection Analysis

Discrimination Accuracy	Confidence	
	High	Low
Correct	hit	miss
Incorrect	false alarm	correct rejection

Method

Participants. Twenty-four members of the University College London community volunteered for this experiment in return for a payment of £3. None had taken part in any previous grammar-learning experiment. Twelve participants were assigned to Experiment 1A, and 12 were assigned to Experiment 1B.

Stimuli. There was one set of training stimuli and two sets of test stimuli. Thirty-two sequences generated by the grammar shown in Figure 1 were assigned to the training set. These were presented twice in a different randomized order for each participant. There were 32 grammatical and 32 ungrammatical test sequences. These were presented in a different randomized order to each participant. The stimuli were identical to those used by Brooks and Vokey (1991) and are shown in the Appendix.

Procedure. The training phase was identical in Experiments 1A and 1B. Each training sequence appeared in the center of a computer screen. After 5 sec, the sequence disappeared, and the participants were required to reproduce the sequence, using the keyboard. If an error was made, the participants were asked to try again, and the sequence would reappear for a further 5 sec; otherwise, the next sequence would appear.

After the training phase, the participants were informed that the sequences they had just seen followed some simple rules and that they would now see some new sequences, half of which obeyed the rules and half of which did not. The test phase differed in Experiments 1A and 1B. In Experiment 1A, the participants made their grammaticality and confidence judgments concurrently by pressing one of four buttons marked *yes conforms to rules – more confident*, *yes conforms to rules – less confident*, *no does not conform to rules – less confident*, and *no does not conform to rules – more confident*. The buttons were displayed horizontally in the spatial arrangement indicated and were positioned approximately 1 in. below the location of the test string. During this period, the test sequence remained on the screen.

In Experiment 1B, the participants made only their grammaticality decisions while the test sequences remained visible, by pressing one of two buttons marked *yes conforms to rules* and *no does not conform to rules*. After this, the sequence disappeared, and 1 sec later, a prompt asked “How confident are you of your response?” and two buttons marked *more confident* and *less confident* appeared.

Results and Discussion

For each participant, two d' values were calculated and are reported in Table 2. The first, d'_g , indexed the participants' ability to discriminate between grammatical and ungrammatical sequences. Hits were calculated from *yes* responses to grammatical sequences, and false alarms were calculated from *yes* responses to ungrammatical sequences. We used a criterion of $\alpha = .05$ for all

analyses. The participants reliably classified more sequences correctly than would be expected by chance in both Experiment 1A [mean $d'_g = 0.78$, $SE = 0.11$; $t(11) = 7.27$] and Experiment 1B [$d'_g = 0.70$, $SE = 0.14$; $t(11) = 5.03$].¹

The second value, d'_a , measured the participants' ability to discriminate between correct and incorrect responses. As was discussed earlier, if participants are aware of the information used to classify sequences as grammatical or ungrammatical, they should be more confident in correct than in incorrect classification decisions. The scheme for calculating hits and false alarms is shown in Table 1: Hits were defined as *more confident* responses to correct classifications (including both endorsements and rejections), and false alarms were defined as *more confident* responses to incorrect classifications (again including both endorsements and rejections). Thus, explicit knowledge would result in d'_a values greater than zero. On the other hand, if participants are unaware of the information used to classify sequences, then *more confident* and *less confident* responses should be distributed equally between correct and incorrect classifications, resulting in d'_a values close to zero. The obtained d'_a values are shown in Table 2. The data indicate that the participants had more explicit knowledge than would be expected by chance in Experiment 1A [mean $d'_a = 0.39$, $SE = 0.14$; $t(11) = 2.73$] and in Experiment 1B [$d'_a = 0.29$, $SE = 0.12$; $t(11) = 2.49$].

A cross-experiment comparison of d'_a revealed no reliable difference between Experiments 1A and 1B ($t < 1.0$), indicating that the 1-sec delay did not lead to a decay of confidence. We conclude from this that earlier dissociations between confidence and accuracy, using rating scales (e.g., Dienes et al., 1995), are unlikely to have been due to the delay between making grammaticality and confidence judgments.

Note that the discrimination of correct and incorrect responses is probably harder than the discrimination of grammatical and ungrammatical decisions, and for reasons described by Kunimoto et al. (2001), it follows that even for an ideal observer, d'_a must always be less than d'_g . This means that the measure of awareness (d'_a) for a participant who is perfectly conscious of all the information that he or she used to classify items need not equal the

Table 2
Measures of Discrimination Accuracy (d'_g) and Binary Awareness (d'_a) for Experiments 1 and 2

Measure	Experiment 1A				Experiment 1B				Experiment 2			
	Classification		Binary Awareness		Classification		Binary Awareness		Classification		Binary Awareness	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
<i>p</i> (H)	.67	.03	.47	.04	.66	.04	.63	.05	.60	.03	.44	.03
<i>p</i> (FA)	.38	.03	.35	.06	.40	.04	.52	.05	.53	.03	.37	.03
<i>C</i>	−0.08	0.06	0.29	0.16	−0.08	0.07	−0.24	0.15	−0.18	0.06	0.26	0.09
<i>d'</i>	0.78	0.11	0.39	0.14	0.70	0.14	0.29	0.12	0.17	0.07	0.19	0.07

Note—The *C* statistic is a measure of response bias. Values greater than zero indicate liberal bias, and those less than zero indicate conservatism; see Macmillan and Creelman (1991).

measure of grammaticality discrimination (d_g). The difference between the measure of awareness and the measure of discrimination is not an index of unconscious processing (see Kunimoto et al., 2001, Appendix B).

What is the source of participants' confidence judgments? So far, we have assumed that confidence is related to accuracy because participants are aware of the information that they use to classify sequences. In principle, however, participants might be able to discriminate between correct and incorrect responses without being aware of the information they used to make the responses, if they were based on an alternative source of information. Indeed Koriat (2000) has suggested that feelings of knowing may be based on both implicit and explicit sources. One possibility is that participants could monitor their own response times and use this information to estimate their accuracy. If participants responded more quickly when making correct decisions than when making incorrect ones and estimates of confidence were unconsciously based on this information, our measure of awareness would be affected by unconscious influences. Response times are known to influence confidence estimates in a variety of tasks, including face recognition (Shaw, McClure, & Wilkens, 2001), general knowledge (Wright & Ayton, 1988), perceptual judgments (Baranski & Petrusic, 2001), and memory (Kelley & Lindsay, 1993).

In fact, there was no reliable difference between response times for correct and incorrect decisions for either Experiment 1A (mean RT for correct decisions = 3,285 msec, $SE = 303$; mean RT for incorrect decisions = 3,342 msec, $SE = 317$; $t < 1$) or Experiment 1B [mean RT for correct decisions = 3,454 msec, $SE = 485$; mean RT for incorrect decisions = 3,681 msec, $SE = 462.61$; $t(11) = 1.46$]. Hence, response times could not have provided a basis for the participants to determine the accuracy of their grammaticality decisions. Nonetheless, response times could have had some influence on the participants' confidence estimates. The relationship between confidence response times, and discrimination accuracy was examined using regression. Discrimination accuracy and response time were both reliably, and independently, related to the participants' levels of confidence in both Experiment 1A [RT, $\beta = -0.14$, $t(10) = 3.93$; discrimination accuracy, $\beta = 0.11$, $t(10) = 3.03$] and Experiment 1B [RT, $\beta = -0.27$, $t(10) = 7.71$; discrimination accuracy, $\beta = 0.12$, $t(10) = 3.55$]. Hence, estimates of confidence are related partly to accuracy and partly to response times. However, response times were unrelated to discrimination accuracy in both Experiment 1A [$r(10) = -.04$, $p > .05$]² and Experiment 1B [$r(10) = -.06$, $p > .05$]. We conclude from this that although participants' confidence judgments are related to response times, the latter are unrelated to whether or not participants make a correct decision.

In summary, Experiment 1 yielded evidence that artificial grammar learning involves explicit knowledge. Other research also suggests that the standard artificial

grammar learning experiment involves a substantial degree of explicit knowledge, even when assessed with the more usual rating technique. For example, in their Experiment 2, Dienes and Altmann (1997) observed that participants were reliably more confident in correct than in incorrect decisions when classifying sequences composed of the same letters as those used in the training sequences. This was not true, however, when the participants were asked to transfer their knowledge to sequences instantiated with a different vocabulary. In Experiment 2, we ask whether transfer is implicit.

EXPERIMENT 2

Many dissociations between confidence and accuracy have been observed when participants have been asked to transfer their knowledge acquired in one vocabulary to another (e.g., Dienes & Altmann, 1997; Tunney & Altmann, 2001). In Experiment 2, we examined whether transfer is implicit, using the binary confidence rating technique.

The extent to which grammatical knowledge can be transferred to novel vocabularies gives an indication of how participants might represent artificial grammars. Reber (1969) has argued that the knowledge transferred to novel vocabularies consists of unconscious knowledge of grammatical rules. Others have argued that transfer involves consciously learning the correspondence between training and test vocabularies (e.g., Gomez & Schvaneveldt, 1994; Redington & Chater, 2002).

If transfer involves conscious knowledge, it should be detectable with the d_a measure.

Method

Participants. Twenty members of the University College London community volunteered for this experiment in return for a payment of £3. No participant had taken part in any previous grammar-learning experiment.

Stimuli. The stimuli were identical to Experiment 1, with the exception that the training and test sequences were composed of different letter sets. The letters of each sequence were transposed from Set 1 to Set 2, using the following rules: M → Q, X → J, R → L, V → Z, T → F. Half the participants were trained on sequences identical to those used in Experiment 1 (Set 1), but the letters of the test sequences were changed to Letter Set 2. The other group of participants was trained on sequences composed of Letter Set 2 and were tested on sequences composed of Letter Set 1.

Procedure. The procedure was identical to that in Experiment 1A, with the exception that the instructions were changed to allow for the transfer to a new letter set.

Results and Discussion

The data were analyzed in the same way as in Experiment 1. Table 2 shows the measures of discrimination and awareness. The participants classified more sequences correctly than would be expected by chance [$d_g = 0.17$, $SE = 0.07$; $t(19) = 2.61$], replicating the finding that participants can transfer knowledge acquired in one vocabulary to another. The critical question, however, is whether this knowledge is implicit. The data in-

dicating that, in fact, the participants' knowledge was explicit [$d_a = 0.19$, $SE = 0.07$; $t(19) = 2.72$].

Did the participants base their confidence judgments on the same kinds of information as the participants in Experiment 1? There was no difference between response times for correct ($M = 4,794$ msec, $SE = 579$) and incorrect ($M = 5,175$ msec, $SE = 606$) decisions [$t(19) = 1.84$]. To determine the basis of the participants' confidence estimates, response times and decision accuracy were entered into a regression. Discrimination accuracy reliably predicted confidence estimates [$\beta = .06$; $t(18) = 2.30$]. However, in contrast to Experiment 1, response times did not predict confidence estimates [$\beta = -.04$; $t(18) = 1.27$]. Response times were unrelated to discrimination accuracy [$r(18) = -.04$, $p > .05$], so the former could not have been used as an implicit cue to the accuracy of responses. The participants based their confidence judgments on explicit knowledge of the accuracy of their grammaticality judgments.

The finding that transfer was explicit in Experiment 2 is at odds with experiments reported by Dienes and Altmann (1997) and Tunney and Altmann (2001), who found transfer to be implicit. In those experiments, however, the confidence rating procedure was used, rather than the signal detection procedure used here. Experiment 3 therefore assessed participants' awareness using the rating procedure.

EXPERIMENT 3

Experiment 3 was identical to Experiment 2, with the exception that instead of making a high or low confidence judgment, the participants rated how confident they were in each grammaticality decision on a scale ranging from 50% to 100%.

Method

Participants. Twenty members of the University College London community volunteered for this experiment in return for a payment of £3. No participant had taken part in any previous grammar-learning experiment.

Stimuli. The stimuli were identical to those in Experiment 2.

Procedure. The procedure was identical to that in Experiment 1B, with the exception that the participants indicated their confidence on a continuous scale. The participants first made a grammaticality decision by pressing one of two buttons marked *yes conforms to rules* and *no does not conform to rules* while the se-

quence was visible on the screen. After this, the sequence disappeared, and 1 sec later, a prompt appeared that asked "How confident are you of your response? Enter a number between 50 and 100, where 50 is a complete guess and 100 is absolutely certain." The confidence ratings were entered via keyboard input and were displayed in a small box below the prompt. This rating was a free response that could be any number between 50 and 100.

Results and Discussion

The measures of discrimination accuracy and awareness are shown in Table 3. The participants classified more sequences correctly than would be expected by chance [$d_g = 0.19$, $SE = 0.06$; $t(19) = 3.06$], replicating this aspect of Experiment 2. The mean confidence for correct decisions was 73.62 ($SE = 2.72$) and for incorrect decisions was 72.01 ($SE = 2.69$). These values were not reliably different, which seems to suggest that transfer occurred without explicit knowledge [mean difference = 1.61, $SE = 1.00$; $t(19) = 1.60$]. The upper and lower bounds of the 95% confidence interval for this difference score were 3.70 and -0.50 , respectively, on a scale ranging from -50 to 50 . Figure 2 shows the accuracy of the participants' decisions (as a percentage) for each level of confidence. First, confidence ratings were allocated to bins (e.g., 50%–55% confidence), and the mean confidence ratings within that bin were calculated (plotted on the x -axis). Next, the mean accuracies (expressed as percentages) for grammaticality decisions for confidence ratings within each bin were calculated (plotted on the y -axis; cf. Allwood, Granhag, & Johansson, 2000; Yates, 1990). Figure 2 clearly indicates that the participants' confidence judgments were unrelated to the accuracy of their grammaticality decisions.

This conclusion is also supported by an analysis of guesses. On average, guesses (defined as trials on which confidence was rated as 50%) constituted 17.10% ($SE = 5.69\%$) of all the responses. A percent correct score based solely on guess responses was 61.93% ($SE = 5.46\%$), and this figure was reliably greater than an expected chance value of 50% correct [$t(17) = 2.19$], indicating some implicit knowledge (2 participants reported no guesses and were omitted from this analysis).

There was no difference between response times for correct ($M = 5,059$ msec, $SE = 503$) and incorrect ($M = 5,314$ msec, $SE = 525$) decisions [$t(19) = -1.37$]. Regression analysis indicated that, as in Experiment 2, dis-

Table 3
Measures of Discrimination Accuracy (d_g) and Rating Awareness for Experiment 3

Measure	Rating Classification		Awareness		Awareness		
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	Measure	<i>M</i>	<i>SE</i>
<i>p</i> (H)	.59	.02	.54	.02	Confidence correct	73.62	2.72
<i>p</i> (FA)	.52	.03	.50	.02	Confidence incorrect	72.01	2.69
<i>C</i>	-0.15	0.07	-0.05	.08			
<i>d</i>	0.19	0.06	0.11	.08	Difference	1.61	1.00

Note—The index d_r is derived from a median split of the continuous ratings in the column labeled awareness.

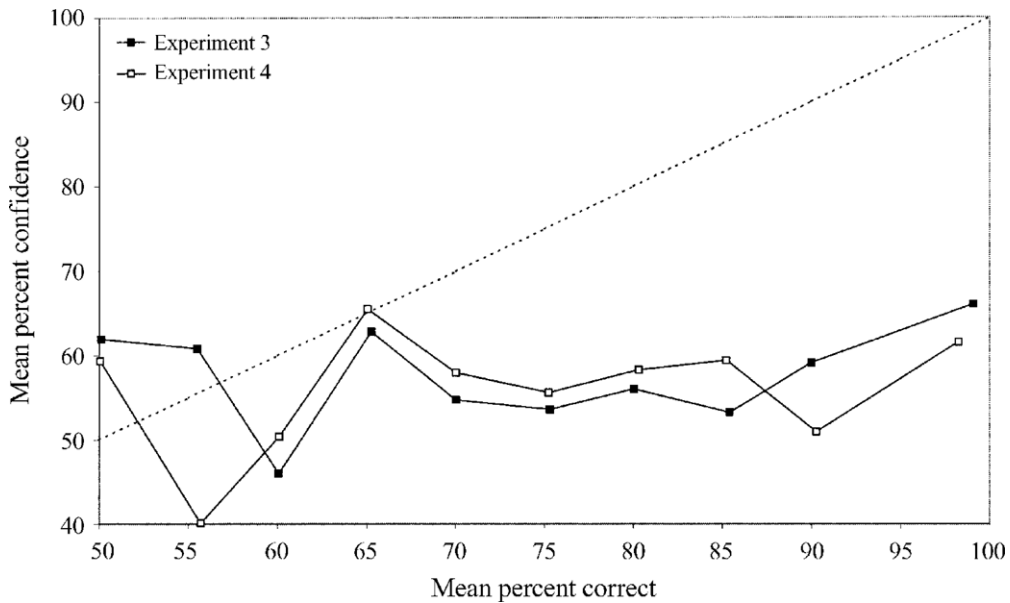


Figure 2. Calibration curves for Experiments 3 and 4: The numbers plotted in the graph give the mean percent correct for each level of confidence. The diagonal hatched line indicates hypothetical perfect calibration.

crimination accuracy predicted confidence estimates [$\beta = 0.06$; $t(18) = 2.15$]. Response times did not reliably predict confidence estimates [$\beta = 0.04$; $t(18) = 1.28$] and were unrelated to discrimination accuracy [$r(18) = -.04$, $p > .05$]. In contrast to the results of the analysis based on difference scores, the results of this regression analysis indicate that confidence ratings are determined by discrimination accuracy.

The results of Experiment 3 are consistent with previous reports in which rating scales have been used to indicate that transfer is implicit according to criteria based on difference scores (Dienes & Altmann, 1997; Tunney & Altmann, 2001) but contrast with our Experiment 2. The sole difference between Experiments 2 and 3 is the method used to assess awareness. One possibility is that the rating technique fails to detect the presence of explicit knowledge because of the way it is analyzed. To test this possibility, we calculated a third index of awareness, d_r , based on confidence ratings. To do this, the participants' ratings were divided according to whether they fell above or below their median rating: Ratings that fell above the median were categorized as *more confident* responses, and those that fell below were categorized as *less confident* responses. The d_r index was then calculated in the same way as in Experiments 1 and 2, according to the scheme shown in Table 1. Ties were omitted. Two participants were excluded because their median rating was 50%, and a 3rd was excluded whose median rating was 100%. The resulting d_r value was not reliably different from chance [$d_r = 0.11$, $SE = 0.08$; $t(16) = 1.38$; the lower and upper limits of the 95% confidence interval were -0.06 and 0.27 , respectively]. If taken at face value, this result would indicate that learning had been implicit. An

alternative hypothesis is that participants find it more difficult to place their phenomenal states on a continuous scale, relative to a binary scale. The consequence of this may be that the binary technique can detect lower levels of awareness for a given sample size than the rating technique can. To test this possibility, in Experiment 4, we used a within-subjects comparison of binary and scale confidence estimates.

EXPERIMENT 4

In Experiment 4, we compared the binary and rating techniques in a within-subjects design.

Method

Participants. Twenty members of the University College London community volunteered for this experiment in return for a payment of £3. No participant had taken part in any previous grammar-learning experiment.

Stimuli. The stimuli in Experiment 4 were identical to those in Experiments 2 and 3.

Procedure. The procedure was similar to those used before, with the exception that both binary and continuous confidence judgments were elicited. The participants first made their grammaticality and binary confidence decisions concurrently by pressing one of four buttons marked *yes conforms to rules – more confident*, *yes conforms to rules – less confident*, *no does not conform to rules – less confident*, and *no does not conform to rules – more confident*. After this, the sequence disappeared, and 1 sec later, the participants were prompted to enter a continuous confidence rating (from 50% to 100%) in the same way as in Experiment 3.

Results and Discussion

The measures of discrimination and awareness are shown in Table 4. The participants classified more se-

Table 4
Measures of Discrimination Accuracy (d_g) and Awareness for Experiment 4

Measure	Classification		Binary Awareness		Rating Awareness		Awareness		
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	Measure	<i>M</i>	<i>SE</i>
$p(H)$.67	.03	.47	.04	.54	.05	Confidence correct	73.37	1.97
$p(FA)$.55	.03	.40	.05	.52	.05	Confidence incorrect	72.32	1.82
<i>C</i>	-0.30	0.08	0.21	0.12	0.08	0.10			
d'	0.35	0.08	0.22	0.08	0.05	0.07	Difference	1.05	0.64

Note—The index d'_r is derived from a median split of the continuous ratings in the column labeled awareness. The index d'_a is derived from binary ratings.

quences correctly than would be expected by chance [$d'_g = 0.35$, $SE = 0.08$; $t(19) = 4.29$].

In terms of awareness, the binary technique indicated that the participants possessed significant explicit knowledge [$d'_a = 0.22$, $SE = 0.08$; $t(19) = 2.97$], thus replicating the results of Experiment 2. In contrast to this, but consistent with the results of Experiment 3, the rating technique did not reveal significant explicit knowledge [mean difference = 1.05, $SE = 0.64$; $t(19) = 1.65$]. The upper and lower bounds of the 95% confidence interval for this difference score were 2.40 and -0.29, respectively, on a scale ranging from -50 to 50. Figure 2 shows the accuracy of the participants' decisions (as a percentage) for each level of confidence and, as in Experiment 3, indicates that the participants' confidence was unrelated to decision accuracy.

On average, 16.8% ($SE = 3.93\%$) of the responses were rated by the participants to be guesses. A percent correct score calculated solely on guesses ($M = 59.2$, $SE = 4.9$) was reliably different from a chance value of 50% [$t(20) = 1.85$, one-tailed].

There was no evidence of a response time difference for correct ($M = 6,111$ msec, $SE = 799$) and incorrect ($M = 6,358$ msec, $SE = 831$) decisions [$t(19) = 1.22$]. As before, regression analyses were conducted to determine the source of the participants' confidence estimates. Separate analyses were conducted for binary and rating estimates. Binary confidence estimates were largely determined by discrimination accuracy [$\beta = 0.08$; $t(18) = 2.70$] and, in contrast to Experiments 2 and 3, were to some extent influenced by response times [$\beta = -0.07$; $t(18) = 2.60$]. Response times were unrelated to discrimination accuracy [$r(18) = -.11$, $p > .05$]. Rating estimates of confidence were marginally determined by discrimination accuracy, although this failed to reach significance on a two-tailed test [$\beta = 0.05$; $t(18) = 1.91$; this value was not reliably different from the value obtained from binary estimates ($t < 1$)]. Rating estimates of confidence were not determined by response time ($\beta = 0.02$, $t < 1$). Response times were unrelated to discrimination accuracy [$r(18) < -.02$, $p > .05$].

If the rating and binary measures are conceptually similar, why might one measure suggest that learning is implicit, whereas the other suggests that learning is explicit? Experiment 3 suggested that even when continuous

ratings are analyzed in the same way as binary ratings, they do not detect the presence of explicit knowledge. The continuous ratings obtained in Experiment 4 were also analyzed in binary form, using the median split procedure described in Experiment 3 (2 participants were omitted from the analysis because their median response was 50%). The subsequent value did not indicate significant explicit knowledge ($d'_r = 0.05$, $SE = 0.07$; $t < 1$; the lower and upper limits on the 95% confidence intervals were -0.09 and 0.20, respectively) and was significantly lower than the d'_a value of 0.22 derived from binary responses [$t(17) = 4.70$]. In contrast, when the binary ratings were transformed onto a scale similar to the continuous rating, they did indicate explicit knowledge. For this analysis, binary ratings were coded as 1 for *more confident* responses and 0 for *less confident* responses, thus enabling the binary data to be analyzed in the same way as the rating data. The mean confidence for correct and incorrect decisions was .47 ($SE = .04$) and .40 ($SE = .05$), respectively. Indicative of explicit knowledge, these scores differed significantly [mean difference = .07, $SE = .02$; $t(19) = 2.90$].

Taken together, these analyses suggest that the number of intervals participants use to indicate their state of awareness, rather than the properties of signal detection analysis, confer the advantage that the binary technique has over the rating technique in detecting explicit knowledge. One reason for this might be that participants find it relatively more difficult to indicate their states of awareness with continuous confidence ratings. Thus, although the binary and the rating procedures both measure awareness of discrimination accuracy, the latter might be less sensitive than the former and, because the technique relies on a null effect, may consequently lack power. To see whether this was the case, we pooled the difference scores for confidence across Experiments 3 and 4, increasing the amount of power we had to detect the participants' awareness in those experiments. Across both Experiments 3 and 4, we found learning to be explicit by the zero-correlation criterion [mean difference = 1.33, $SE = 0.59$; $t(39) = 2.26$]. We conclude that subjective measures of awareness, such as the zero-correlation criterion, can produce erroneous evidence of implicit learning when they are based on differences between confidence ratings made on a continuous scale or on an insufficiently

large sample. Instead, measures based on a binary scale provide a more sensitive measure of awareness.

GENERAL DISCUSSION

In the experiments reported here, we examined whether artificial grammar learning is implicit according to a subjective criterion of awareness. In Experiment 1, we applied a method derived from subliminal perception research (Kunimoto et al., 2001) to artificial grammar learning. Participants were asked to classify sequences composed of the same letters as those seen during training and, for each decision, to indicate whether their confidence was high or low. These binary confidence ratings revealed that the participants had higher confidence in correct than in incorrect decisions, indicating that artificial grammar learning was explicit. These data are consistent with an experiment reported by Dienes and Altmann (1997, Experiment 2), using an analogous measure based on confidence ratings that also showed learning to be explicit.

However, previous research has shown that when participants were asked to transfer their knowledge to sequences composed of different vocabulary elements, the participants were as confident in incorrect decisions as in correct ones (Dienes & Altmann, 1997; Tunney & Altmann, 2001), suggesting that transfer is mediated by implicit knowledge. In Experiment 2, we attempted to replicate this effect, using binary confidence judgments. In contrast to previous research we found transfer to be explicit. In Experiment 3, we did find evidence that transfer might be implicit, using confidence ratings. The only difference between Experiments 2 and 3 was that, in the first case, confidence estimates were elicited on a binary scale and, in the latter case, on a continuous scale. Experiment 4 confirmed directly, within the same participants, that the two methods differ in their sensitivity. The two indices produce conceptually similar measures of awareness; in each case, explicit knowledge is indicated when confidence is higher for correct than for incorrect decisions. We conclude that, although the binary and the rating scales measure the same aspect of awareness, the binary technique is relatively more sensitive, possibly because it minimizes scale end effects (Juslin et al., 2000; McClelland & Bolger, 1994) and because participants might find it easier to place phenomenal states on a binary than on a continuous scale. Allwood et al. (2000) reported that the relationship between confidence and accuracy increased as the number of training and test items increased, raising the possibility that participants' ability to use the rating scales might develop over time. The consequence of the relative lack of sensitivity of the rating method is that it may be statistically less powerful than the binary method. Indeed, we found learning to be explicit according to this criterion when the rating data from Experiments 3 and 4 were combined.

Criteria based on differences in confidence between correct and incorrect decisions, whether derived from

continuous or binary scales, cannot in themselves exclude the possibility that there may be some implicit contribution to grammaticality decisions. Rather, positive results, such as those obtained in our experiments with the binary technique, indicate a reliable explicit component in cases—in particular, in transfer experiments—in which less sensitive measures had previously indicated that behavior was largely under implicit control.

The term *implicit learning* was coined to describe the kind of learning supposed to occur when participants are asked to memorize sequences generated by an artificial grammar (Reber, 1967). In what sense can artificial grammar learning be said to be implicit?

Previous research has demonstrated artificial grammar learning to occur above objective thresholds of awareness (e.g., Dulany et al., 1984; Gomez & Schvaneveldt, 1994; Meulemans & Van der Linden, 2002; Peruchet & Pacteau, 1990)—that is, participants can pass objective tests of awareness, such as recognition and generation. There remains the special case of preserved learning capacity in amnesic populations. Studies with amnesic patients perhaps provide the most compelling evidence for the existence of distinct implicit and explicit memory systems. Amnesics, whose explicit memory is poor, have been demonstrated to show intact learning in various implicit memory tasks (see Gabrieli, 1998). However, even amnesic patients pass objective tests of awareness. They are able to discriminate between old and new sequences in a recognition test (Knowlton, Ramus, & Squire, 1992) and can recognize fragments of grammatical sequences at above-chance levels (Knowlton & Squire, 1996). These studies do show functional independence of implicit and explicit learning, in the sense that recognition decisions are impaired in amnesics relative to controls, whereas grammaticality decisions remain relatively intact. Nonetheless, we know of no artificial grammar learning experiment in which either normal participants or amnesics have failed objective tests of awareness. In this sense, artificial grammar learning is explicit. Moreover, Kinder and Shanks (2001) have shown that a single-system computational model can simulate the dissociation between recognition and classification in amnesics, relative to controls.

Do amnesics fail subjective tests of awareness? Recently, Channon et al. (2002) asked a group of amnesics and a group of age-matched controls to rate their confidence in grammaticality decisions on a 6-point scale. Both groups of participants were as confident in incorrect decisions as they were in correct decisions, indicating that their knowledge was implicit. However, as the experiments reported in this article show, estimates of subjective confidence based on scales are not as sensitive as those based on binary responses. The results of our experiments indicate that, at least for normal populations, when a suitably unbiased and sensitive measure is used, artificial grammar learning is found to occur above the subjective threshold of awareness. This finding is also likely to be true for amnesic populations if

the 6-point scale used by Channon et al. is as insensitive as the continuous scale. Future research will have to determine whether participants would be subjectively aware of the accuracy of their decisions in Channon et al.'s procedure if awareness is assessed with binary confidence judgments.

The experiments reported here indicate, on the basis of confidence, that artificial grammar learning occurs above subjective thresholds of awareness. That is to say, participants appear to have some insight into their decision-making processes, so that they can distinguish between correct and incorrect decisions in terms of confidence. The present results may be viewed as a useful supplement to studies employing objective tests, in that they demonstrate a degree of subjective, conscious introspection during classification in a standard implicit learning task.

REFERENCES

- ALLWOOD, C. M., GRANHAG, P. A., & JOHANSSON, H. (2000). Realism in confidence judgments of performance based on implicit learning. *European Journal of Cognitive Psychology*, **12**, 165-188.
- BARANSKI, J. V., & PETRUSIC, W. M. (2001). Testing architectures of the decision-confidence relation. *Canadian Journal of Experimental Psychology*, **55**, 195-206.
- BROOKS, L. R., & VÓKEY, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, **120**, 316-323.
- CHAN, C. (1992). *Implicit cognitive processes: Theoretical issues and applications in computer systems design*. Unpublished doctoral thesis, University of Oxford.
- CHANNON, S., SHANKS, D., JOHNSTONE, T., VAKILI, K., CHIN, J., & SINCLAIR, E. (2002). Is implicit learning spared in amnesia? Rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia*, **40**, 2185-2197.
- CHEESMAN, J., & MERIKLE, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, **36**, 387-395.
- DEHAENE, S., & NACCACHE, L. (2001). Toward a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, **79**, 1-37.
- DIENES, Z., & ALTMANN, G. [T. M.] (1997). Transfer of implicit knowledge across domains: How implicit and how abstract? In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 107-123). Oxford: Oxford University Press.
- DIENES, Z., ALTMANN, G. T. M., KWAN, L., & GOODE, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1322-1338.
- 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.
- ELDRIDGE, L. L., SARFATTI, S., & KNOWLTON, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review*, **9**, 139-145.
- ERDELYI, M. H. (1974). A new look at the new look: Perceptual defense and vigilance. *Psychological Review*, **81**, 1-25.
- ERIKSEN, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review*, **67**, 279-300.
- GABRIELI, J. D. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, **49**, 87-115.
- GOMEZ, R. L., & SCHVANEVELDT, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 396-410.
- HOLENDER, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral & Brain Sciences*, **9**, 1-66.
- JUSLIN, P., WINMAN, A., & OLSSON, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review*, **107**, 384-396.
- KELLEY, C. M., & LINDSAY, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory & Language*, **32**, 1-24.
- KELLY, S. W., BURTON, A. M., KATO, T., & AKAMATSU, S. (2001). Incidental learning of real world regularities in Britain and Japan. *Psychological Science*, **12**, 86-89.
- KINDER, A., & SHANKS, D. R. (2001). Amnesia and the declarative/procedural distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience*, **13**, 648-669.
- KNOWLTON, B. J., RAMUS, S. J., & SQUIRE, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, **3**, 172-179.
- KNOWLTON, B. J., & SQUIRE, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 169-181.
- KORIAT, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness & Cognition*, **9**, 149-171.
- KUNIMOTO, C., MILLER, J., & PASHLER, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness & Cognition*, **10**, 294-340.
- LICHTENSTEIN, S., FISCHOFF, B., & PHILLIPS, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). Cambridge: Cambridge University Press.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MCCLELLAND, A. G. R., & BOLGER, F. (1994). The calibration of subjective probabilities: Theories and models 1980-94. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 453-482). Chichester, U.K.: Wiley.
- MEULEMANS, T., & VAN DER LINDEN, M. (2002). Artificial grammar learning in amnesia. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making* (pp. 144-164). Hove, U.K.: Psychology Press.
- NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.
- NEWELL, B. R., & BRIGHT, J. E. H. (2002). Well past midnight: Calling time on implicit invariant learning? *European Journal of Cognitive Psychology*, **14**, 185-205.
- PEIRCE, C. S., & JASTROW, J. (1884). On small differences of sensation. *Memoirs of the National Academy of Sciences*, **3**, 75-83.
- 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 & Verbal Behavior*, **6**, 855-863.
- REBER, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, **81**, 115-119.
- REDINGTON, M., & CHATER, N. (2002). Knowledge representation and transfer in artificial grammar learning. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making*. (pp. 121-143). Hove, U.K.: Psychology Press.
- SHANKS, D. R., & ST. JOHN, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral & Brain Sciences*, **17**, 367-447.
- SHAW, J. S., III, MCCLOURE, K. A., & WILKENS, C. E. (2001). Recognition instructions and recognition practice can alter the confidence-response time relationship. *Journal of Applied Psychology*, **86**, 93-103.
- TUNNEY, R. J., & ALTMANN, G. T. M. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1322-1333.

- WELLMAN, H. M. (1977). Tip of the tongue and feeling of knowing experiences: A developmental study of memory monitoring. *Child Development*, **48**, 432-444.
- WRIGHT, G., & AYTON, P. (1988). Decision time, subjective probability, and task difficulty. *Memory & Cognition*, **16**, 176-185.
- YATES, J. F. (1990). *Judgment and decision making*. London: Prentice-Hall.

NOTES

1. For practical purposes, the discrimination index d' varies from approximately -4.65 to 4.65 . Null discrimination returns values of d' close to zero.
2. This and all the subsequent correlations reported are one-tailed.

APPENDIX
Sequences Generated by the Artificial Grammar Shown in Figure 1
and Used in Experiments 1-4

Training (Grammatical)	Testing	
	Grammatical	Ungrammatical
MXRVXT	MXRMXT	MXRRXT
VMTRRRR	VMTRRRX	VMTRRRT
MXTRRR	VXTRRR	TXTRRR
VXVRMXT	VXVRVXT	VXVRTXT
VXVRVM	VXVRV	VXVRT
VMRVVV	VMRVVVM	VMRVVVR
MXRTMVR	MXRTMXR	MXRTMTR
VMRMXTR	VMRVXTR	VMRTXTR
MXR	MVR	MTR
VMRVXVR	VMRVXVT	VMRVXX
MVRVM	MXRVM	MTRVM
VMRMVR	VMRMXRV	VMRMTRV
VMRMVXR	VMRMVXT	VMRMVXX
MXRTVXT	MXRTMXT	MXRTRXT
MXRMVXR	MXRMVXT	MXRMVXX
MVXTR	MVXTX	MVXTT
MVXRM	MVXRV	MVXRT
MVXRMVR	MVXRMXR	MVXRMTR
VXVT	VXVR	VXVM
MXRMXRV	MXRMVRV	MXRMTRV
MXRTVMT	MXRTVMR	MXRTVMM
VMT	VMR	VMM
MVXRVMR	MVXRVMT	MVXRVMX
MVXRVVV	MVXRVVM	MVXRVVR
MXRMXRM	MXRMVRM	MXRMTRM
VMRVMT	VMRVMR	VMRVMM
MXRVVVM	MVRVVVM	MTRVVVM
VXVRMXR	VXVRMVR	VXVRMTR
MXTRRRX	VXTRRRX	TXTRRRX
VXVTRRX	VXVTRRR	VXVTRRM
MXTRRX	VXTRRX	TXTRRX
VMRTMXT	VMRTVXT	VMRTTXT

(Manuscript received July 15, 2002;
revision accepted for publication May 27, 2003.)