Cognitive Psychology **42,** 61–112 (2001)
doi:10.1006/cogp.2000.0743, available online at http://www.idealibrary.com on IDEAL®

Abstractionist and Processing Accounts of Implicit Learning

Theresa Johnstone and David R. Shanks

University College London, London, England

Five experiments evaluated the contributions of rule, exemplar, fragment, and episodic knowledge in artificial grammar learning using memorization versus hypothesis-testing training tasks. Strings of letters were generated from a biconditional grammar that allows different sources of responding to be unconfounded. There was no evidence that memorization led to passive abstraction of rules or encoding of whole training exemplars. Memorizers instead used explicit fragment knowledge to identify the grammatical status of test items, although this led to chance performance. Successful hypothesis-testers classified at near-perfect levels by processing training and test stimuli according to their rule structure. The results support the episodic-processing account of implicit and explicit learning.

Key Words: artificial grammar learning, episodic processing, episodic knowledge; explicit learning, hypothesis testing, implicit learning, memorization, rule abstraction.

People have an impressive capacity for storing information about particular events. This ‘‘episodic’’ memory allows us to recall the context of specific experiences, such as what we did on our last holiday, and is now well understood both psychologically (Tulving, 1983) and at the neural level (McClelland, McNaughton, & O'Reilly, 1995; Treves & Rolls, 1994). We also have the ability, however, to acquire knowledge about generalities, that is, properties true of classes of objects or events. We can judge the grammaticality of a novel sentence, read a word in an unfamiliar script, perform arithmetic operations, and so on. These abilities seem to require representations of ab-

Theresa Johnstone and David R. Shanks, Department of Psychology, University College London, England. This research was supported by United Kingdom Medical Research Council Research Studentship G78/4925 and by a grant from the United Kingdom Economic and Social Research Council (ESRC). The work is part of the program of the ESRC Centre for Economic Learning and Social Evolution, University College London. We thank Shelley Channon, Axel Cleeremans, Zoltán Dienes, Koen Lamberts, Mark St. John, Richard Tunney, and Bruce Whittlesea for their helpful comments on this work.

Address correspondence and reprint requests to David Shanks, Department of Psychology, University College London, Gower Street, London WC1E 6BT, England. E-mail: d.shanks@ucl.ac.uk.

0010-0285/00 $35.00

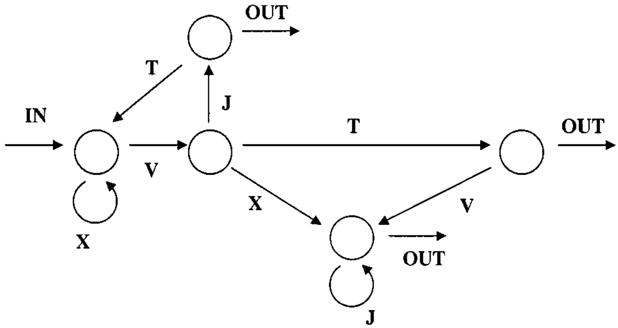
stract, general properties such as the rules of a grammar that transcend and are separate from knowledge of specific objects or events.

Cognitive psychology has traditionally dealt with this distinction in a straightforward way, by assuming separate psychological and neural processes for representing specific and general knowledge. Under various terms (e.g., episodic, explicit, and declarative), knowledge of specific events is usually assumed to be distinct from knowledge about general properties (e.g., semantic, implicit, and procedural). Indeed, learning about specifics and extracting generalities are often thought to be computationally incompatible tasks. A puzzle, however, is to explain how general knowledge can be acquired, since abstract properties themselves are never directly observed (see Whittlesea, 1997a, 1997b). Instead, such properties must be induced from multiple experiences with specific objects or events. Hence, the separate-systems account assumes that there exists a mechanism for creating abstractions across specific experiences. Moreover, since we do not normally deliberately intend to perform such abstraction, it must be largely an incidental and unconscious process.

Undoubtedly, there is a wealth of evidence consistent with the separate systems account with a good deal of that evidence coming from artificial grammar learning (AGL) research. For example, Knowlton, Ramus, and Squire (1992) trained normal participants and amnesic patients to memorize strings of letters generated from the finite-state transition grammar shown in Fig. 1. This grammar specifies certain constraints that exist in the order of string elements, much as exist in natural languages. Grammatical strings are generated by entering the diagram at the leftmost node and moving along legal pathways, as indicated by the arrows, collecting letters until an exit point is reached on the right-hand side. The letter string XXVXJJ is grammatical as it can be generated from the diagram, whereas TXXXVT is ungrammatical, as strings must begin with a V or an X.

Knowlton, Ramus, and Squire (1992) tested specific knowledge by asking participants to recognize which letter strings they had seen during training using a set of test strings half of which had been presented as training strings and half of which were novel. In contrast, general knowledge was tested by informing participants of the existence of a set of rules governing the structure of the training items—though they were not told what those rules are—and then asking participants to classify novel letter strings as grammatical or ungrammatical depending on whether the letter strings appeared to conform to the rules or not. The fact that the amnesic patients were selectively impaired in making judgments about specific items, while their general knowledge of the grammar was intact, seems strongly to support the idea of separate ''implicit,'' general and ''explicit,'' specific learning systems (but see Kinder & Shanks, 2000, for an alternative account).

Theories of implicit learning are based on three major claims. First, there has been much debate about the conditions required for implicit learning.



Grammatical	Ungrammatical
XXVT	TVT
XXVXJJ	TXXXVT
VXJJ	VXXXVJ
VTV	VJTVTX

FIG. 1. This artificial grammar was used by Knowlton, Ramus, and Squire (1992) and was originally created by Abrams and Reber (1989).

Reber (1967, 1989) suggested that implicit learning occurs when participants observe or memorize representative examples of a complex rule-governed concept without being told that the examples conform to a set of rules. These incidental learning conditions create passive ‘‘consumers’’ and sometimes even ‘‘victims’’ of the knowledge acquired (Lewicki & Hill, 1989, p. 240), with structure emerging in a stimulus-driven way (Cleeremans, 1993, p. 19).

Second, there has been a good deal of controversy over the form of knowledge acquired. Implicit learning was initially assumed to create abstract mental representations of complex rules (Reber, 1967, 1989), but more recently other forms of knowledge representation such as abstract patterns of family resemblance (Mathews, Buss, Stanley, Blanchard-Fields, Cho, & Druhan, 1989) or first-order dependencies between adjacent letters (Gomez, 1997) have been proposed. In contrast explicit learning is assumed to depend on mental representations of specific whole or partial training items in a separate episodic memory.

Third, it has been argued that participants lack awareness of the knowledge they use to classify test items. This conclusion is based on an assumption that participants are using rule knowledge to classify at above-chance levels, together with evidence that participants cannot fully state the rules of the grammar and feel as though they are guessing (Dienes, Altmann, Kwan, & Goode, 1995; Reber & Lewis, 1977). In contrast, because participants are aware of observing or memorizing whole or partial training examples, conscious recollection of ‘‘old’’ items and a sense of novelty for ‘‘new’’ items accompany recognition performance.

Despite 30 years of research, there is still debate about the validity of these three key assumptions of implicit learning theory. One major issue is the difficulty of reliably establishing whether participants are using complex rule-based knowledge in classification tests or whether performance is based on knowledge of training examples or fragments of examples (see Shanks & St. John, 1994). The knowledge issue is clearly intertwined with questions of how sensitive tests of awareness are to the knowledge that participants actually use in classification tests (see Shanks & St. John, 1994) and how we assess whether the knowledge used to classify test items is computed during training or at test (see Redington & Chater, 1996). The second major issue is whether participants really are ‘‘passive victims’’ of incidental learning situations or whether, as suggested by Whittlesea and Dorken (1993), they actively engage with training stimuli to meet the demands of the task. Whittlesea and Dorken (1993) suggested that test performance might appear to be implicit because we are only aware of knowledge if the training task draws our attention to it. Memorization instructions disguise the fact that knowledge acquired during training is relevant to the test; and before we can test for particular types of knowledge we need some kind of theory about how knowledge acquired during training is relevant to classification and free recall tests.

FORM OF KNOWLEDGE

Four structural accounts suggest that the information participants use to classify test stimuli is based solely on knowledge available from the training stimuli. In contrast, Whittlesea’s (1997a, 1997b) episodic account focuses on the overlap of processing between the training and test tasks. The structural accounts are based on (1) general rule knowledge (e.g., Reber, 1967, 1989; Reber & Allen, 1978; Reber & Lewis, 1977), (2) specific item knowledge (Brooks, 1978; Brooks & Vokey, 1991; McAndrews & Moscovitch, 1985; Vokey & Brooks, 1992), (3) letter fragments (Dulany, Carlson, & Dewey, 1984; Perruchet & Pacteau, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990), and (4) both rule and fragment knowledge (Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997). We briefly consider these and the episodic account in turn.

Evidence for Rule Knowledge

Convincing evidence that participants classify on the basis of rules depends on having a clear definition of what a rule is and on unconfounding rule knowledge from other explanations of test performance. Unfortunately, researchers have been less than forthcoming on the definitional issue. However, the general idea is that, at least in the case of finite-state grammars, learning the structure of the grammar entails forming some abstract mental representation which describes each of the states of the grammar (i.e., the nodes in Fig. 1), together with the legal letter continuations from that state, and the ensuing state. This mental representation is usually thought of as a symbolic or ‘‘algebraic’’ structure (Marcus, Vijayan, Bandi Rao, & Vishton, 1999) and is assumed to be quite independent of, and distinct from, a specification of the transitional probabilities or distributional statistics of the surface elements instantiating the grammar. Rules go beyond perceptual features and guide classification instead on the basis of deep, conceptual, features (see Redington & Chater, 1996): Thus, no matter how similar a test string is in its surface form to previously seen grammatical strings, it is not grammatical if it breaks a rule of the grammar. Test strings are classified as grammatical if they can be parsed by the grammar. Manza and Reber (1997, p. 75) wrote that

> This position is based on the argument that the complex knowledge acquired during an AG learning task is represented in a general, abstract form. The representation is assumed to contain little, if any, information pertaining to specific stimulus features; the emphasis is on structural relationships among stimuli. The key here is the notion that the mental content consists, not of the representation of specific physical forms, but of abstract representations of those forms.

The strongest evidence for abstract rule knowledge is found in ‘‘transfer’’ tests where participants train on items in one letter set or modality and successfully classify test items presented in a different letter set or modality (e.g., Altmann, Dienes, & Goode, 1995; Brooks & Vokey, 1991; Gomez & Schvaneveldt, 1994; Reber, 1969). The only common factor between training and test items is their underlying abstract structure. For example, Altmann et al. (1995, Experiment 1) trained one group of participants on standard letter strings and a second group on sequences of tones, with both the letter strings and tone sequences conforming to the same rule structure. Thus each letter string had an equivalent tone sequence in which, for instance, the letter M was translated into a tone at the frequency of middle C. In the test phase, participants classified strings presented in the same modality as their training strings (letters/letters or tones/tones) or in the opposite modality (letters/tones or tones/letters). There were two types of control groups who either received no training or who were trained on randomly generated sequences. The results suggested that prior exposure to the grammar led to above-chance classification performance (same modality 56% correct, changed modality 54% correct), whereas control groups performed at chance levels (50%).

Although this experiment appears to provide evidence that changed modality groups used general, abstract, rule knowledge that goes beyond perceptual features, Redington and Chater (1996) demonstrated that participants could have used surface fragments of two or three letters to perform abstraction at test. This is explained in more detail in a later section on evidence for fragment learning. Moreover, Gomez (1997) has presented convincing evidence that transfer is always accompanied by explicit knowledge: Participants who achieved above-chance transfer scores also scored above chance on direct tests in her experiments. Thus there is little evidence at present that transfer is mediated by implicit, abstract knowledge.

Evidence for Exemplar Knowledge

The exemplar account assumes that participants retrieve specific training examples from memory when they classify test items (Brooks, 1978; Brooks & Vokey, 1991; Neal & Hesketh, 1997; Vokey & Brooks, 1992). For example, Vokey and Brooks (1992) trained participants on grammatical strings and tested them on novel strings, where half the test strings were grammatical and half ungrammatical. Orthogonal to grammaticality, half the test items were similar to one training item (differing by only one letter) while half were dissimilar to all training items (differing by two or more letters). Independent effects of grammaticality and similarity were found in both classification and recognition tests.

Vokey and Brooks (1992, p. 328) used instance models (e.g., Hintzman, 1986, 1988; Medin & Schaffer, 1978; Nosofsky, 1986) to argue that independent effects of grammaticality and similarity are consistent with models that rely solely on retrieval of specific items. As new grammatical test items are likely to resemble a large number of grammatical training items, the difference between classification of grammatical versus ungrammatical test items can be explained by ‘‘retrieval time averaging.’’ On the other hand, the difference between similar and dissimilar test items can be explained on the basis of nonlinear generalization gradients where a test item that is highly similar to an item in memory has a disproportionately large effect on test performance. Hence the grammaticality effect could arise because grammatical test items are moderately similar to many training items and the similarity effect could arise because each similar test item is highly similar to one training item. However, Vokey and Brooks (1994) conceded that their results did not allow them to falsify the abstract rule knowledge account.

Evidence for Fragment Knowledge

An opposing theory is that participants learn about the frequency of occurrence of fragments of letters (i.e., two-letter bigrams, three-letter trigrams, etc.) in the training strings and classify novel test strings as grammatical to the extent that a test string contains fragments that were present in the training strings (Dulany, Carlson, & Dewey, 1984; Perruchet, 1994; Perruchet, Gallego, & Pacteau, 1992; Perruchet & Pacteau, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990). Perruchet and Pacteau (1990) compared the performance of participants trained on grammatical letter strings with those trained on the bigrams used to construct the grammatical training strings. The finding that both groups were able to classify novel test strings at above-chance levels suggests that fragment knowledge alone is sufficient to account for accurate classification performance. In fact Perruchet (1994) was able to explain both the grammaticality and similarity effects found by Vokey and Brooks (1994) solely on the basis of trigram knowledge.

However, Gomez and Schvaneveldt (1994) demonstrated that there are two types of bigram violations within ungrammatical strings and participants trained on bigrams were only sensitive to one violation type. Participants trained on grammatical strings could detect both illegal letter pairs and legal letter pairs in illegal positions within a string, while participants who memorized bigrams were only able to detect illegal letter pairs. But Redington and Chater (1996) added a further dimension to this debate by showing that Gomez and Schvaneveldt's results can be predicted by ''toy models'' which call a test string grammatical if all bigrams and trigrams have been seen in training items and call a test string ungrammatical if it contains novel letter fragments. Overall, then, the evidence that grammaticality judgments are to some extent mediated by fragment knowledge is quite strong.

Evidence for Abstract Rule and Fragment Knowledge

Knowlton and Squire (1994, Experiment 2b) challenged the exemplar account by using test stimuli that contained the same orthogonal grammaticality and whole-item similarity manipulations as Vokey and Brooks (1992) had used, but with an added manipulation where fragment similarity was held constant across similar and dissimilar test item types. The results showed that Vokey and Brooks' results were more likely to have been produced by rule and fragment knowledge than by rule and whole-item knowledge. However, these results leave open the debate about whether the grammaticality and fragment effects are derived from dual knowledge sources or ''retrieval time averaging'' from only one knowledge base. Finally, as amnesic patients classified test items with the same degree of accuracy as normal participants, Knowlton and Squire concluded that both rule and fragment knowledge are implicit.

Meulemans and Van der Linden (1997) have provided the most convincing evidence for this dual-mechanism account using test stimuli that balanced rule knowledge orthogonally to fragment knowledge. They found that after training on 32 letter strings (Experiment 2a), participants classified test strings using fragment knowledge, whereas after training on 125 letter strings (Experiment 2b) they classified on the basis of rule knowledge. However, Johnstone and Shanks (1999) demonstrated that in Experiment 2b, information about grammatical rules and familiar training fragments was confounded with knowledge of the positional constraints on letter fragments. This argument is explained in more detail shortly in the section on problems with transitional grammars.

Evidence for the Episodic-Processing Account

The episodic-processing account challenges all of the above accounts, as those accounts focus solely on stimulus-driven acquisition of structural aspects of training items (i.e., rules, exemplars, or letter fragments), whereas the episodic-processing account suggests that (1) processing knowledge is acquired in addition to structural knowledge, (2) training instructions dictate which aspects of the structure of training items are encoded, and (3) participants can apply the same knowledge explicitly or implicitly depending on whether they understand the relationship between processing fluency and the knowledge they acquired by processing training items in particular ways.

Evidence that knowledge of both structure and processing is encoded during training was provided by Whittlesea and Dorken (1993). Participants memorized items such as ENROLID that were generated from a grammar, either by pronouncing or spelling them aloud, and then classified test items by pronouncing half of them and spelling the remainder. Test performance was only reliably above chance when the study and test processes were the same. When items were spelled in training and pronounced at test or pronounced during training and spelled at test, participants classified at chance levels. Thus, the knowledge gained during training included details of processing as well as structural aspects of stimuli, and test performance was successful to the extent that the test instructions cued prior processing episodes.

Evidence that training instructions dictate which aspects of the structure of training items are encoded was presented by Wright and Whittlesea (1998). In the study phase, participants were presented with digit strings such as 1834, all of which conformed to an odd-even-odd-even rule. One group processed each digit by saying it aloud and immediately making a judgement about whether it was a low (less than five) or high number (greater than four). For example, 1834 would be processed as ‘‘1-low-8-high-3-low-4-low.’’ A second group processed each string by pronouncing the two digit pairs. In this case, 1834 would be processed by saying ‘‘eighteen thirty-four.’’ At test, half the strings were created by reversing the order of the two familiar digit pairs in training items (e.g., 1834 became 3418) and half the test items comprised novel digit pairs.

Although all test strings were novel, participants were asked to discriminate between ‘‘old’’ items seen during training and ‘‘new’’ items. The group who said the training items as two-digit pairs were more likely than the group who read strings digit-by-digit to say that test items containing familiar digit pairs were old and test items containing unfamiliar digit pairs were new. Thus the manner in which training items were processed dictated which aspect of the structure of test items was encoded (single digits or digit pairs) and subsequent test performance. These results cast doubt on the idea that there is a ‘‘neutral’’ form of coding, whether it is of whole items, fragments, or rules. Instead, and consistent with the principles of transfer-appropriate processing (Morris, Bransford, & Franks, 1977) what is learned depends on the processing demands of the task.

Whittlesea and Williams (1998) put forward a discrepancy-attribution hypothesis which suggests that participants will apply knowledge explicitly or implicitly at test depending on whether they understand the relationship between processing particular test items fluently and the knowledge they acquired during training. During training, participants pronounced natural words (e.g., TABLE), orthographically regular easily pronounced nonwords (e.g., HENSION), and less regular and hence harder to pronounce nonwords (e.g., LICTPUB). At test, participants were asked to pronounce old and novel versions of these three types of items and to indicate whether each item had been seen during training.

Whittlesea and Williams found in one experiment that novel regular nonwords were 21% more likely to be called old than novel words and 28% more likely to be called old than novel irregular nonwords. Since the words were pronounced more rapidly than the regular nonwords, followed by the irregular nonwords, it is clear that fluency *per se* is not the critical variable. Whittlesea and Williams suggested that participants did not expect nonwords to be processed fluently and as a result unconsciously attributed the surprising fluency of reading the orthographically regular ones to those items having been presented during training. In contrast, there was no discrepancy between the first impression that a letter string such as TABLE is a word and the subsequent fluency of processing. Participants were therefore able to discount fluency and use conscious recollection to make test responses for natural words.

PROBLEMS WITH TRANSITIONAL GRAMMARS

Johnstone and Shanks (1999) questioned the use of artificial grammars that are based on finite-state transition rules to investigate implicit learning, as these grammars do not provide a means of convincingly determining the contributions of rule and fragment knowledge in classification performance. In short, it is very hard to create test items that unconfound grammaticality and fragment composition. The problem with transition-rule grammars is that they use a rule structure that dictates legal consecutive letters tied to particular letter-string locations. For example, all legal strings generated from the grammar used by Meulemans and Van der Linden (created by Brooks & Vokey, 1991 and shown in Fig. 2) start with MV, MX, VM, or VX. But, if participants classify test strings because they know what letters are legal in the first two positions, it is not clear what type of knowledge they are using to make this decision. They could be using rules (i.e., ‘‘all legal strings must begin with M or V,’’ ‘‘an initial M can only be followed by V or X,’’ and ‘‘an initial V can only be followed by M or X’’), but they

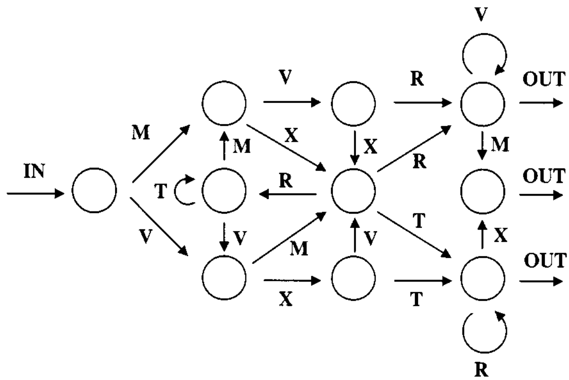


FIG. 2. The artificial grammar used by Brooks and Vokey (1991).

could also be using bigram knowledge (i.e., all training strings began with MV, MX, VM, or VX).

During the past 30 years there has been a trend to control for and quantify fragment statistics at increasing levels of detail. However, grammatical knowledge has remained a vague concept, quantified only in terms of two distinct categories (grammatical versus ungrammatical) that is assumed to exist whenever fragment statistics do not account for all of the variance in test performance. One way of clarifying matters is to quantify grammaticality, though we are not convinced that this will allow unequivocal conclusions (see Johnstone & Shanks, 1999, for a preliminary effort along these lines). A sounder way is to use a different type of finite-state grammar that allows us to unconfound grammaticality, fragment similarity, and whole-item similarity more convincingly.

AN ALTERNATIVE GRAMMAR

Shanks, Johnstone, and Staggs (1997, Experiment 4) constructed letter strings from a biconditional grammar, originally designed by Mathews et al. (1989, Experiment 4). This biconditional grammar generates strings of length eight from a vocabulary of six letters and has three rules governing the relationship between letters in positions 1 and 5, 2 and 6, 3 and 7, and 4 and 8 such that when one position contains a D, the linked letter must be an F, and similarly for G/L and K/X (e.g., DFGK.FDLX is legal, whereas LFGK.KDLX is not). This grammar has three advantages over transition-rule grammars.[1] First, each of the three rules can occur in any of the letter locations. For example a D can be placed in any of the eight positions, as

[1] From a computational linguistics point of view the grammar is still finite-state because it generates a finite number of sentences and the sentences are of finite length.

long as an F occurs in the associated letter location. Second, as the rule-related positions have three intervening letters, it is possible to unconfound rule and fragment knowledge. Finally, it is straightforward to quantify how grammatical test strings are. All grammatical strings contain four valid rules and in our studies all ungrammatical test strings contain three valid rules and one illegal letter pairing. The strings generated from this biconditional grammar allow whole-item and fragment information to be unconfounded from grammaticality more successfully than has been achieved with transitional grammars. The first aim of the present research, therefore, is to reevaluate the key issues (What are the conditions of implicit learning? What is its content? To what extent is it consciously accessible?) that have driven AGL research in the past 30 years but which have yet to be settled.

RULE LEARNING

In addition to studying implicit rule learning, Shanks, Johnstone, and Staggs (1997, Experiment 4) also looked at the performance of participants who consciously tried to learn the rules of a grammar. In most previous studies (Dulany, Carlson, & Dewey, 1984; Perruchet & Pacteau, 1990; Reber, 1976; Reber, Kassin, Lewis, & Cantor, 1980; Turner & Fischler, 1993), instructions aimed at encouraging rule learning were minimal (e.g., participants were simply informed prior to a standard study phase that the strings conformed to a set of rules and that discovering these rules may be helpful). However, Shanks et al. used a task, originally created by Mathews et al. (1989, Experiments 3 and 4), that was designed to encourage rule learning. Participants were shown flawed examples of grammatical strings, asked to indicate which letters they thought created violations of the grammar, and then given feedback about their accuracy. Training strings contained one or two violations of the biconditional rules, and participants adopted a hypothesis-testing strategy to determine the underlying rules used to generate grammatical strings. Like Mathews et al. (Experiment 4) we found a clear dissociation in classification test accuracy, with chance-level performance by some participants and almost perfect performance by others. Shanks et al. found that these latter participants showed a strong effect of grammaticality and no effect of whole-item similarity, suggesting that the mental representations underlying their performance were the abstract principles of the grammar. These results suggest that, as predicted by the episodic-processing approach (Whittlesea, 1997a, 1997b; Whittlesea & Dorken, 1993), rule abstraction does not take place under implicit learning conditions, but depends on active, conscious efforts to identify the rules of the grammar, leading to explicit knowledge. The second aim of the present article is to assess how well the findings of five AGL experiments fit this episodic-processing approach. We begin by asking whether whole-item knowledge contributes to grammaticality decisions under explicit and implicit training conditions.

EXPERIMENT 1

Shanks, Johnstone, and Staggs (1997, Experiment 4) used the biconditional grammar and match and edit tasks created by Mathews et al. (1989, Experiments 3 and 4), along with new training and test strings that manipulated rule knowledge (i.e., grammaticality) orthogonal to whole-item similarity while ensuring that these two factors had minimal overlap with fragment similarity. As fragment similarity is generally referred to as associative chunk strength (ACS) in the AGL literature, this term is used here. ACS provides a measure of the frequency with which fragments of two letters (bigrams) and three letters (trigrams) in test items overlap with training stimuli. At the level of whole items, test strings that differ from one training item by only two letters are defined as similar, whereas test items that differ by three or more letters from all training items are defined as dissimilar.

One group of participants was induced to process the perceptual characteristics of the training stimuli by asking them to memorize letter strings without telling them that these strings were constructed according to the rules of a grammar (match group). A second group was induced to process the abstract properties of the letter strings by asking them to test hypotheses in order to discover the rules of the grammar (edit group). The results showed a clear dissociation in classification accuracy, with edit participants who learned the rules performing at near-perfect levels, while the match group performed at chance. Neither group showed an effect of whole-item similarity.

Experiment 1 sought to extend the findings of Shanks et al. (1997, Experiment 4) with three major modifications. We included a control group, counterbalanced the rule letter pairs across participants, and assessed participants' awareness of the rules of the grammar. While the match and edit groups trained on the same grammatical training items, a control group was asked to memorize letter strings that contained neither rules, whole-item similarity, nor ACS relationships with the test strings. All groups classified the same set of novel test strings. Participants in Shanks et al.'s experiment trained on strings based on the three rule pairings of D with F, G with L, and K with X. In the present experiment each participant within each group saw a different version of the 15 possible sets of letter pairs that can be created from the letters D, F, G, K, L, and X. Finally, a questionnaire was used fairly exhaustively to assess participants' knowledge of the rules of the grammar. It is true that posttask questionnaires have often been criticized as instruments for assessing awareness on the grounds that they may be somewhat insensitive (e.g., Shanks & St. John, 1994). However, this is only a concern if participants demonstrate behavioral sensitivity to some feature of a task (e.g., a rule structure) which they cannot report; in the present experiments the rule structure is so simple that we were confident participants would have no difficulty reporting it.

There were three hypotheses. The first was that as the match group had not been asked to process the rule structure of the training strings, they would show no effect of grammaticality in their classification performance. In fact we conjectured that the match group would perform at the chance level anticipated in the control group. Second, as the edit participants had processed the rule structure, it was predicted that they would show an effect of grammaticality. Third, based on the results of Shanks et al. (1994, Experiment 4), it was predicted that none of the groups would show an effect of whole-item similarity.

Method

Participants. Twenty-four psychology undergraduates from University College London (UCL) were paid £5 to take part in the experiment and were randomly assigned to a match, edit, or control group. The control and match groups were initially told that they were taking part in a short-term memory experiment, while the edit group was told that they would be taking part in a rule-discovery experiment. All three groups carried out the same classification test.

Match task. The control and match groups were told that they were being tested on how good their short-term memory was for strings of letters like DFGX.FDLK. On each of 72 trials a string appeared on the screen and the participant was asked to mentally rehearse it. The string stayed on the screen for 7 s and then the screen went blank for 2 s. Then a list of three strings was displayed and the participant was asked to type the number (1–3) of the string that matched the one they were rehearsing. The two foils were illegal versions of the correct string. The order of strings was randomized across blocks and participants.

Edit task. The edit group was told that they would be shown strings of letters such as DFGX.FDLK that were constructed from the six letters D, F, G, K, L, and X and that the computer was programmed with a set of rules for putting letters into acceptable orders. Participants were told that their task was to work out what these rules were. They would see one string at a time for each of 72 trials. Each string would have between two and four letters that violated the rules, in terms of the relationships between the letters. Participants were asked to indicate whether they felt that each letter conformed to or violated the rules by putting a Y below letters that they believed conformed to the rules and an N below letters that they believed did not. It was explained that at the beginning of the experiment the participant would not know the rules and therefore they would have to start by guessing. But on each trial they would be given feedback in the form of the correct string of Ys and Ns, as well as the corrected string itself, and they should try to learn from this feedback in order to induce the rules.

Classification task. Immediately before the classification task began, participants in the control and match groups were informed that the letter strings they had been asked to memorize in the first part of the experiment were generated from a complex set of rules. They were told not to worry if they did not notice any rules, as the task that they had performed made it very unlikely that they would know them. In fact only the match group had seen rule-governed strings, whereas the control group had not. Participants in the edit group were reminded that in the first part of the experiment they had used a hypothesis-testing strategy to try to learn the rules of the grammar. They were also told not to worry if they did not feel completely confident in their understanding of the rules, as the task was very difficult.

The 144 strings presented for classification comprised two blocks of the same 72 strings presented in different random orders across blocks and participants. Each string was presented in turn, and participants were asked to rate how well it conformed to the rules on a scale from 1 to 6. The points on the scale indicated the following: (1) *certain,* (2) *fairly certain,* (3) *guess that the string obeys the rules,* (4) *guess,* (5) *fairly certain,* and (6) *certain that the string does not obey the rules.*

Questionnaire. After participants had finished the classification test, they were asked a series of questions in order to explore how much they had learned about the letter pair rules. Participants were asked if they had adopted any particular strategy in the test phase to determine if the strings conformed to the rules. If this failed to elicit the rules of the grammar they were then asked if they had noticed any rules in the construction of the training strings. If this failed to elicit the rules, they were then asked if they knew the rules linking letters in the first half of the string to corresponding letters in the second half of the string. If this third question failed to elicit the rules of the grammar, participants were told that there were three rules that dictated which letters could appear in location 5 depending on what letter was in location 1 and they were then asked if they could say what those rules were. This question was repeated for each pair of rule-related letter locations.

Materials. Three separate sets of letter strings were created to train the control group, to train the match and edit groups, and to provide a classification test for all three groups (see Appendix A). In addition, allocation of two sets of training strings (Training Lists 1 and 2) was a between-subjects manipulation in both the match and edit groups. Though each participant within each group saw a different example of the 15 possible sets of three letter pairs that can be created from D, F, G, K, L, and X, the examples in this article were all generated from the rule-set D↔F, G↔L, and K↔X.

The training strings used by all three groups were designed so that each letter was evenly distributed across each of the eight locations and so that ACS was equivalent across test items. The control group training strings did not contain biconditional rules, whereas the match and edit group training strings did. For each training string used for the match and edit groups, two ungrammatical versions were created with two or four letter violations. For each training string used for the control group, two versions were created that differed from the original string by one or two letters. Participants in the match and edit groups saw each string (from Training List 1 or 2, see Table A1) four times for a total of 72 training trials. Participants in the control group saw each string (Table A2) twice again yielding 72 training trials.

In relation to the training strings seen by the match and edit groups, half of the test strings were grammatical and the other half ungrammatical and within each of these two categories, half of the strings were similar to training strings and the other half were dissimilar. Similar test items only differed from a specific training item by two letters, whereas dissimilar test items differed from all training items by more than two letters. This created four types of test items: grammatical and similar (GS), grammatical and dissimilar (GD), ungrammatical and similar (US), and ungrammatical and dissimilar (UD). Participants were presented with all the strings shown in Table A3. For those trained on Training List 1, the Test List 1 strings are similar and the Test List 2 strings dissimilar. The converse is true for participants trained on Training List 2. There were no relationships of grammaticality or similarity between the test strings and the training strings that the control group saw.

Calculation of Associative Chunk Strength (ACS). ACS was calculated on the basis of the theoretical perspective on chunking presented by Servan-Schreiber and Anderson (1990) and as applied by Knowlton and Squire (1994, p. 85). The actual ACS statistics for each experiment are shown in the appendices. ACS is a measure of the frequency with which fragments of two letters (bigrams) and three letters (trigrams) within test items appeared in training strings. Two measures of ACS were calculated for the initial and terminal fragments within each test string (anchor ACS) and for all fragments in a test string (global ACS). For example, the anchor ACS for the grammatical test string LFGK.GDLX in relation to List 1 training items is (LF (1) + LX (1) + LFG (0) + DLX (1)) / 4 = 0.75 and in relation to List 2 training items is (LF (0) + LX (1) + LFG (0) + DLX (0)) / 4 = 0.25.

Global ACS was calculated by breaking each test string down into its constituent bigrams and trigrams and then calculating how many times each fragment had occurred in any location within Training Lists 1 and 2 and dividing the totals by the number of fragments (7 bigrams and 6 trigrams). For example, LFGK.GDLX can be broken down into LF, FG, GK, KG, GD, DL, LX, LFG, FGK, GKG, KGD, GDL, and DLX, which when compared to the training strings contributes (((4 + 3 + 5 + 3 + 4 + 2 + 5) / 7) + ((0 + 1 + 0 + 0 + 0 + 1) /

TABLE 1
Mean Percentage of Correct Responses across Blocks in the Training Phase

Experiment	Group	Block 1	Block 2	Block 3	Block 4	Overall accuracy
1	Control	83	83	86	92	86
	Match	82	88	87	92	87
	Edit	60	72	75	77	71
2	Match	93	92	97	92	94
	Edit	64	77	76	79	74
3	Match	86	89	91	92	90
	Control	84	84	89	90	87
4	Match	91	94	91	94	93
5	Match	90	90	87	93	90

6)) / 2 = 2.02 to the Training List 1 similar global ACS score and (((2 + 5 + 5 + 2 + 1 + 4 + 5) / 7) + ((0 + 0 + 1 + 0 + 0 + 0) / 6)) / 2 = 1.80 to the Training List 2 dissimilar global ACS score. Appendix A shows that grammatical versus ungrammatical and similar versus dissimilar test strings did not differ in ACS.

Results

A significance level of 0.05, two-tailed, is assumed for all statistical tests in this article, unless the level and direction of the test are specifically stated. Table 1 shows data from four blocks of 18 training trials. Responses in both the control and match groups were scored as correct if the same string as that initially presented for rehearsal was selected from the list. A one-way ANOVA for the control group, with block as a within-subjects variable, indicated that there was a significant effect of block, $F(3, 21) = 3.53$, $MS_e = 44.46$, demonstrating that participants' ability to memorize the training strings improved as training progressed. In contrast, a two-way ANOVA for the match group, with block as a within-subjects variable and training list as a between subjects variable, found no overall effect of block, $F(3, 18) = 2.46$, $MS_e = 52.73$, or list, $F < 1$, and no Block $\times$ List interaction, $F < 1$. The control and match groups' performance was close to ceiling and participants were performing the memorization task accurately across training blocks.

The edit group was asked to indicate whether each letter in a training string was grammatical or ungrammatical by placing Y or N beneath it. The accuracy of these responses was scored at the level of individual letters (see Table 1). A two-way ANOVA with block as a within-subjects variable and training list as a between-subjects variable yielded an effect of block, $F(3, 18) = 6.91$, $MS_e = 66.37$, but no effect of list, $F < 1$, and no Block $\times$ List interaction, $F < 1$. These results suggest, as predicted, that the edit group acquired new knowledge as training progressed by successfully identifying the rules of the grammar as a result of hypothesis testing and feedback.

Table 2 shows the mean percentage of items classified as grammatical for each group. Within each group the classification responses are shown for the four test item types. The mean percentage of correct responses for the control group was 51%. This provides a measure of chance performance that can be compared with the classification results of the match and edit groups. The mean percentage of correct responses for the match group was 55%, and the 95% confidence intervals (CI) shown in Table 2 indicate that this is not significantly different from the percentage correct in the control group, $t(14) < 1$, $SE = 4.69$. The mean percentage correct for the edit group was 75%, with a confidence interval of 58 to 92% indicating above-chance performance.

A three-way ANOVA comparing the percentage of items classified as grammatical (ratings ≤ 3), with group (control, match, or edit) as a between-subjects variable and both grammaticality and similarity as within-subjects variables, found a significant effect of grammaticality, $F(1, 21) = 10.43$, $MS_e = 1015.27$, and a Group $\times$ Grammaticality interaction, $F(2, 21) = 5.16$, $MS_e = 1015.27$. The main effects of group and similarity and the Group $\times$ Similarity, Grammaticality $\times$ Similarity, and Group $\times$ Grammaticality $\times$ Similarity interactions were not significant, with $F < 1$ in each case.

Next, separate two-way ANOVAs comparing the percentage of items classified as grammatical were conducted on the data for each of the three groups, with both grammaticality and similarity as within-subjects variables. In the control group there was no effect of grammaticality, $F < 1$, or similarity, $F < 1$, nor a Grammaticality $\times$ Similarity interaction, $F(1, 7) = 2.03$, $MS_e = 26.70$. In the match group there was no effect of grammaticality, $F(1, 7) = 1.39$, $MS_e = 647.01$, or similarity, $F < 1$, nor a Grammaticality $\times$ Similarity interaction, $F < 1$. The edit group showed an effect of grammaticality, $F(1, 7) = 8.60$, $MS_e = 2342.61$, whereas the effect of similarity, $F < 1$, and the Grammaticality x Similarity interaction, $F < 1$, were not significant. This suggests that only edit group participants learned the rules of the grammar and that they classified test items based on rules alone.

A grammatical sensitivity measure (d'_g) was calculated by comparing the percentage of grammatical test strings that were correctly classified as grammatical (hits) with the percentage of ungrammatical test strings incorrectly classified as grammatical (false alarms). Figure 3 and Table 2 show that only edit group participants discriminated grammatical from ungrammatical items at above-chance levels ($d'_g = 2.18$), as $d'_g = 0$ fell inside the 95% confidence interval of the d'_g scores of both the control and match groups. Participants in all three groups showed a slight bias toward calling strings ungrammatical.

Table 2 also shows the mean percentage of ‘‘correct’’ responses when performance is based on the similarity, rather than the grammaticality, of test strings. In this case, a test response is ‘‘correct’’ when a similar item is classified as grammatical or a dissimilar item is classified as ungrammatical. Hits were grammatical responses to high ACS (grammatical and ungrammat-

TABLE 2
Mean Percentages of Items Classified as Grammatical in Experiments 1–3

Group	*N*	Mean percentage of ''grammatical'' responses				Grammatical measures			Similarity measures		
		GS	GD	US	UD	Correct and CI	Sensitivity (d'_g) and CI	Bias c	Correct and CI	Sensitivity (d'_s) and CI	Bias c
Experiment 1											
Control	8	49	51	49	47	51 ± 2.6	0.06 ± 0.14	0.03	50 ± 3.4	−0.01 ± 0.18	0.03
Match	8	52	55	43	43	55 ± 8.8	0.32 ± 0.55	0.03	50 ± 2.6	−0.02 ± 0.14	0.04
Edit	8	74	71	22	23	75 ± 16.8*	2.18 ± 1.43*	0.13	50 ± 1.9	−0.02 ± 0.16	0.11
Nonlearners	11	48	48	46	44	52 ± 1.8	0.16 ± 0.24	0.14	50 ± 2.1	−0.01 ± 0.14	0.12
Learners	5	97	96	4	8	95 ± 4.9*	3.67 ± 0.90*	−0.07	49 ± 2.3	−0.04 ± 0.12	−0.03
Experiment 2											
Match	8	73	33	61	31	54 ± 5.1	0.18 ± 0.27	0.01	68 ± 8.1*	1.04 ± 0.63*	0.05
Edit	8	77	63	29	13	75 ± 17.5*	2.14 ± 1.55*	0.09	58 ± 8.3	0.44 ± 0.49	0.14
Nonlearners	12	67	31	59	28	53 ± 3.7	0.14 ± 0.20	0.10	67 ± 6.6*	0.99 ± 0.47*	0.13
Learners	4	99	99	2	3	98 ± 1.7*	4.22 ± 0.39*	−0.08	50 ± 1.3	−0.01 ± 0.06	−0.02
Experiment 3											
Control	29	58	61	61	62	49 ± 2.3	−0.05 ± 0.12	−0.29	49 ± 2.6	−0.04 ± 0.14	−0.29
Unaware Match	54	59	36	58	38	50 ± 1.2	0.00 ± 0.06	0.08	61 ± 2.9*	0.67 ± 0.20*	0.11
Aware Match	16	59	31	45	34	53 ± 2.6*	0.15 ± 0.14*	0.21	60 ± 6.2*	0.68 ± 0.45*	0.26

Note. GS = grammatical and similar; GD = grammatical and dissimilar; US = ungrammatical and similar; UD = ungrammatical and dissimilar; CI = 95% confidence interval.
*indicates $p < .05$.

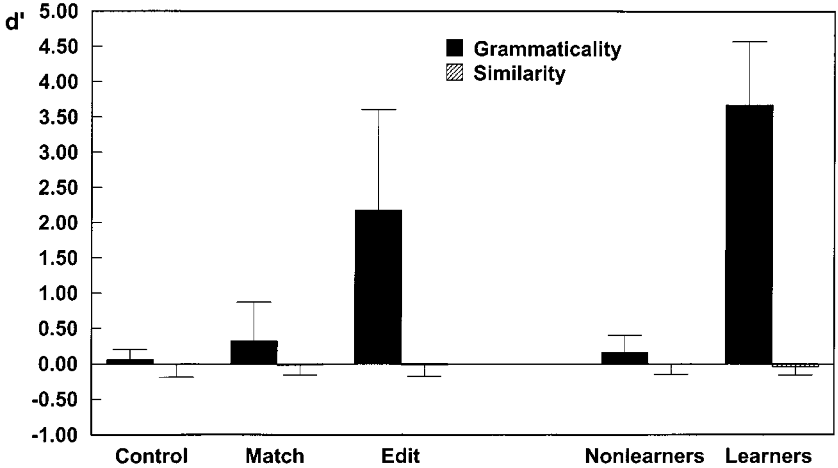


FIG. 3. Mean d' for grammaticality-based (d'_g) and similarity-based (d'_s) classification in the control, match, edit, nonlearner, and learner groups of Experiment 1. Error bars represent 95% confidence intervals.

ical) strings, whereas false alarms were grammatical responses to low ACS (grammatical and ungrammatical) strings. On the basis of similarity the control, match, and edit groups each classified 50% of strings as grammatical according to their similarity to training strings. Figure 3 shows that the 95% confidence intervals around the corresponding similarity sensitivity scores (d'_s) for all three groups encompass chance-level sensitivity.

Inspection of participants' verbal reports indicated that every participant in the control group, seven members of the match group, and four members of the edit group had no knowledge of the rules of the grammar. Based on the verbal report data, participants in the match and edit groups were partitioned into those who successfully identified the rules (Learners, $N = 5$) and those who did not (Nonlearners, $N = 11$), and a second set of analyses were conducted for these two subgroups (see Table 2). One participant in the match group was not aware of the rules at the end of the match task, but worked out what they were during the classification test. Three of the edit group learners reached ceiling in their first block of training and one participant reached ceiling in the fourth block.

The mean percentages correct were 52% for the nonlearners and 95% for the learners. Figure 3 shows that the nonlearners performed at chance levels with a d'_g score of 0.16 and confidence interval of -0.08 to 0.40, while the learners had a d'_g score of 3.67 and a confidence interval of 2.77 to 4.57. The nonlearners had a bias toward calling test strings ungrammatical, whereas the learners had a slight bias toward calling strings grammatical. When the classification scores were examined to see if participants were sensitive to the similarity of test items to whole training items (see Table 2), the nonlearners classified 50% of test strings accurately and the learners classified 49% of the test strings accurately. The 95% confidence levels around the d_s' scores for both groups confirmed chance levels of performance.

Separate two-way ANOVAs were carried out for these two subgroups on the percentages of items classified as grammatical, with both grammaticality and similarity as within-subjects variables. For the nonlearners there was no effect of grammaticality, $F(1, 10) = 2.79$, $MS_e = 36.20$, or similarity, $F < 1$, nor a Grammaticality × Similarity interaction, $F < 1$. For the learners, on the other hand, there was a significant grammaticality effect, $F(1, 4) = 328.52$, $MS_e = 124.81$, with no effect of similarity, $F < 1$, and no Grammaticality × Similarity interaction, $F(1, 4) = 1.54$, $MS_e = 16.01$.

The sums of squares calculated for the two within-subjects ANOVAs indicated that rule knowledge accounted for 1% of the variance in the performance of the nonlearners while it accounted for 98% of the variance in performance of the learners. Whole-item similarity accounted for 0% of the variance in performance of both groups.

Discussion

This experiment replicates and extends the findings of Shanks, Johnstone, and Staggs (1997, Experiment 4). Neither the match nor edit groups showed any knowledge of whole items in their classification performance. The match and edit nonlearners showed no knowledge of the rules of the grammar in their classification performance and did not differ from the control group. Only the edit learners who successfully identified the rules of the grammar succeeded in the classification test and these participants were fully aware of and able to say what the rules of the grammar are. Thus the rules of this biconditional grammar cannot be learned under standard implicit learning conditions, despite the fact that the performance of the participants in the edit group shows that the rules are learnable.

EXPERIMENT 2

The aim of the second experiment was to examine whether match and edit participants could learn about the two- and three-letter (bigram and trigram) fragments used to construct their training strings. While the same grammar, tasks, and questionnaire were used as in Experiment 1, a new set of training and test strings were constructed from a subset of the bigrams and trigrams that can be created from the letters D, F, G, K, L, and X. Again, half the classification test strings were grammatical and half ungrammatical, but this time within each of these categories half of the test items were constructed from the bigrams and trigrams used to construct the training items, while the other half of the test items were constructed from novel bigrams and trigrams not seen during training. This created a test manipulation of fragment similarity (i.e., ACS) orthogonal to grammaticality. Whereas in Experiment 1 similarity referred to the overlap of whole test items with training items, in Experiments 2–4 similarity refers to the overlap of letter fragments between test and training items. Test items with high ACS are similar to training items, while test items with low ACS are dissimilar to training items.

Whittlesea and Dorken's (1993) episodic-processing account suggests that variations in the processing demands of different training tasks will lead to variations in the knowledge acquired during training. In addition, a participant's ability to retrieve knowledge acquired during training depends on the extent to which the test reinstates the original training context in terms of processing and the structure of the stimuli. It was predicted that edit group participants who successfully hypothesis tested would classify solely on the basis of the rules of the grammar, with no effect of ACS, and that they would also be able to state the rules of the grammar. The episodic-processing account predicts these results because the edit learners would have explicitly processed their training strings in the same way required to carry out the classification test successfully. That is, they would scan from one side of the string to the other, checking that positions 1 and 5, 2 and 6, 3 and 7, and 4 and 8 contain valid rule letter pairs.

There were two predictions for the match group and edit nonlearners. First, they would classify on the basis of ACS, with no effect of grammaticality. Second, they would not be able to say what the rules of the grammar are. The episodic-processing account predicts these results because the match group was instructed to mentally rehearse training strings and this should have caused them to process letter strings in the left-to-right order necessary to create knowledge of letter chunks. During the classification test, strings that contained chunks processed during the training task would be processed more fluently than strings comprising largely novel chunks. This chunk fluency effect would be the only knowledge that match and edit nonlearners could use to classify test strings. Since processing in the training stage did not include explicit analysis of the rule structure, these participants would not be able to say what the rules of the grammar are and would report that they were guessing in the classification test.

Method

Participants. A further 16 UCL psychology undergraduates were each paid £5 to participate in the experiment and were divided equally between a match and an edit group.

Materials. A set of 36 grammatical training strings was created from a subset of 18 of the possible 36 bigrams and 216 trigrams that can be created from D, F, G, L, K, and X (see Appendix B). Again, two ungrammatical training strings were created for each grammatical training string by violating the rules for one or two letter pairs. The violations were made so that the ungrammatical training strings comprised the same subset of bigrams and trigrams as the grammatical training strings.

A set of 48 test strings was created using the full set of possible bigrams and trigrams in order to manipulate ACS independently of grammaticality (see Appendix B). Half of the test strings were grammatical and half were ungrammatical. Orthogonal to this, half of the test strings had high ACS and half had low ACS. All participants classified the 48 test strings twice. The training and test strings are shown in Appendix B, along with string statistics that show that while similar and dissimilar test strings differ in ACS, there is no ACS difference between grammatical and ungrammatical test strings.

Results

Table 1 shows the mean percentage of training items on which subjects in the match and edit groups made correct responses. A one-way ANOVA for the match group, with block as a within-subjects variable, indicated that there was no significant effect of block, $F(3, 21) = 2.13$, $MS_e = 23.38$. Performance was close to ceiling and the results show that participants were performing the memorization task accurately across training blocks. A one-way ANOVA for the edit group yielded a significant effect of block, $F(3, 21) = 3.51$, $MS_e = 103.70$. This suggests, as predicted, that the edit group acquired new knowledge as training progressed with successful hypothesis testing.

Table 2 presents the mean percentage of items classified as grammatical for each group, with the overall mean percentage correct. The mean percentage of correct responses for the match group was 54% (CI 48–59%), which suggests that these results could have occurred by chance. The mean percentage correct for the edit group was 75% (CI 58–93%), indicating above-chance performance. A three-way ANOVA comparing the percentage of items classified as grammatical (ratings ≤ 3), with group (match or edit) as a between-subjects variable and both grammaticality and ACS as within-subjects variables, found significant effects of grammaticality, $F(1, 14) = 9.22$, $MS_e = 1384.97$, and ACS, $F(1, 14) = 17.90$, $MS_e = 564.35$, and a significant Group $\times$ Grammaticality interaction, $F(1, 14) = 5.20$, $MS_e = 1384.97$. The effects of group, $F(1, 14) = 1.42$, $MS_e = 208.29$, and the Group $\times$ ACS, $F(1, 14) = 2.85$, $MS_e = 564.35$, Grammaticality $\times$ ACS, $F < 1$, and Group $\times$ Grammaticality x ACS, $F(1, 14) = 1.05$, $MS_e = 161.79$, interactions were not significant.

Separate two-way ANOVAs comparing the percentage of items classified as grammatical were conducted for each of the groups, with both grammaticality and ACS as within-subjects variables. In the match group there was a significant effect of ACS, $F(1, 7) = 17.94$, $MS_e = 551.14$, but no effect of grammaticality, $F(1, 7) = 1.85$, $MS_e = 213.22$, and no Grammaticality $\times$ ACS interaction, $F(1, 7) = 1.24$, $MS_e = 193.37$. The edit group showed an effect of grammaticality, $F(1, 7) = 7.66$, $MS_e = 2556.73$, whereas the effect of ACS, $F(1, 7) = 3.16$, $MS_e = 577.57$ and the Grammaticality $\times$ ACS interaction, $F < 1$, were not significant.

The sensitivity measure d'_g (see Fig. 4) shows that participants in the edit group (CI 0.59–3.69) were better at discriminating grammatical from ungrammatical items than those in the match group (CI -0.09 to 0.45), since there is no overlap in the confidence intervals. Indeed, the level of chance

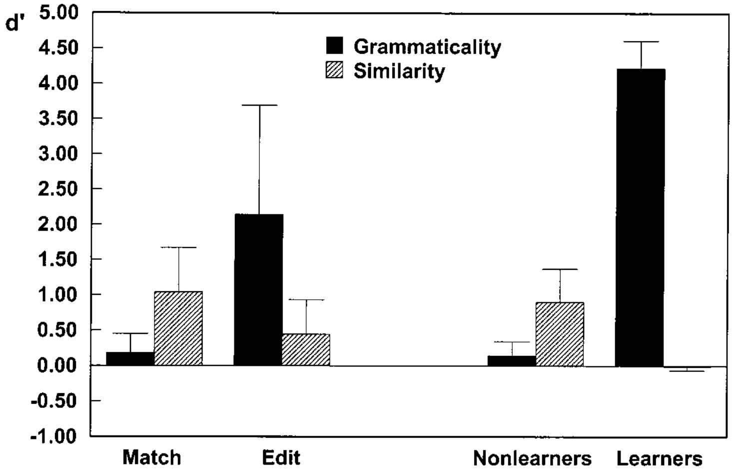


FIG. 4. Mean d' for grammaticality-based (d'_g) and similarity-based (d'_s) classification in the match, edit, nonlearner, and learner groups of Experiment 2. Error bars represent 95% confidence intervals.

responding ($d'_g = 0$) fell inside the 95% confidence interval of the d'_g scores of the match group. In contrast, discrimination in the edit group was well above chance.

Percentage correct and signal detection measures were also computed to assess whether performance had been influenced by the degree of ACS overlap between training and test strings. This time test responses were ‘‘correct’’ if high-ACS items were classified as grammatical and low-ACS items were classified as ungrammatical. Hits were grammatical responses to high-ACS (grammatical and ungrammatical) strings, whereas false alarms were grammatical responses to low-ACS (grammatical and ungrammatical) strings. Table 2 shows that based on ACS, the match group classified 68% of test strings correctly, while the edit group classified 58% of test strings accurately. Figure 4 shows that the 95% confidence intervals around the sensitivity scores (d'_s) indicate above-chance levels of sensitivity to the similarity of test strings to training strings in the match group (CI 0.41–1.67), while the edit group performed at chance (CI −0.05 to 0.93). The groups differed significantly, $t(14) = 2.45$, $SE = 0.80$.

Verbal reports showed that all the participants in the match group, and four participants in the edit group, had no knowledge of the rules of the grammar. The results for the subgroup who successfully identified the rules (learners) and those who did not (nonlearners) are shown in Table 2. The mean percentage correctly classified as grammatical or ungrammatical for the nonlearners was 53% while the learners had a mean percentage correct of 98%.

Separate two-way ANOVAs were carried out for these two subgroups on the percentages of items classified as grammatical, with both grammaticality and ACS as within-subjects variables. For the nonlearners there was an effect of ACS, $F(1, 11) = 25.04$, $MS_e = 543.59$, but no effect of grammaticality, $F(1, 11) = 2.23$, $MS_e = 165.98$, nor a Grammaticality $\times$ ACS interaction, $F < 1$. For the learners, by contrast, there was a significant grammaticality effect, $F(1, 3) = 2933.57$, $MS_e = 12.66$, with no effect of ACS, $F < 1$, and no Grammaticality $\times$ ACS interaction, $F(1, 3) = 1$, $MS_e = 1.09$.

Finally, signal-detection measures were calculated for sensitivities to the grammaticality and ACS of test strings in these two subgroups. Figure 4 indicates that the nonlearners showed chance sensitivity to the rules of the grammar with a mean d'_g score of 0.14, while the learners showed near-perfect sensitivity to the grammaticality of test strings with a mean d'_g score of 4.22. In contrast, the nonlearners were sensitive to the ACS of test strings ($d'_s = 0.99$), while the learners showed chance level performance ($d'_s = -0.01$).

The sums of squares calculated for the separate two-way ANOVAs were analyzed to identify how much of the performance of the nonlearners and learners could be accounted for by knowledge of the rules of the grammar versus ACS. Only 1% of the variance in the performance of the nonlearners is attributable to knowledge of the rules of the grammar, whereas ACS explains 50% of the variance in their performance. In contrast, 100% of the variance in the performance of the learners could be explained by knowledge of the rules of the grammar, with ACS accounting for 0% of their performance.

Discussion

The results supported our predictions that edit learners would classify on the basis of rule knowledge with no effect of ACS, while the match group and edit nonlearners would show the opposite pattern, classifying on the basis of ACS with no knowledge of the rules of the grammar. The edit learners had explicit knowledge of the rules that they used to classify test items. The match group and edit nonlearners did not abstract the rules of the grammar during the training phase, they could not say what the rules of the grammar were, and did not classify test strings on the basis of the rules of the grammar. Note, moreover, that the string endorsements of these participants reveal a ‘‘reverse’’ grammaticality effect for US versus GD items whereby more ungrammatical than grammatical strings were classified as grammatical when the former contained a greater proportion of familiar chunks.

EXPERIMENT 3

Experiments 1 and 2 have failed to find any convincing evidence of rule abstraction under implicit learning (match) conditions. That is, participants unaware of the structure of the domain do not show behavioral sensitivity to that structure. Although this result is, we believe, consistent with results from experiments using transitional grammars (see Johnstone & Shanks, 1999), it runs counter to the claims of a large number of researchers (e.g., Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997; Reber, 1967, 1989; Reber & Allen, 1978; Reber & Lewis, 1977). Given the crucial significance of this issue for the theoretical understanding of the representation of general and specific knowledge, the aim of Experiment 3 was to replicate the results of the match group in Experiment 2, using a larger sample of participants, a new control group, and a different questionnaire that allowed us to quantify both participants' explicit knowledge of the rules of the grammar and their subjective confidence in the accuracy of their explicit knowledge. A match group was trained on modified strings from Experiment 2, while a control group was trained on a new set of strings that did not overlap with the test strings at the level of rules, whole items, or ACS. Both groups classified the test strings used in Experiment 2. The questionnaire was used to divide match participants into those who were aware versus those who were unaware of the rules of the grammar.

It was predicted that the unaware match group would replicate the chance performance of the match group in Experiment 2 and that their performance would not differ from that of the control group, who trained on strings that did not contain the rule manipulation. In contrast, the aware match group was expected to show above-chance classification performance. The unaware and aware match groups were expected to show equivalent, significant levels of ACS knowledge as they had processed the training strings in the same left-to-right manner required accurately to mentally rehearse the strings. This prediction differs from that of Experiment 2 where the edit learners were expected to show no effect of ACS. We assume that the edit learners in Experiment 2 did not show sensitivity to ACS because they had processed the training strings by scanning backward and forward across the central dot to check whether the letters were in accordance with the rules of the grammar. In the present experiment, the control group was expected to show chance levels of ACS knowledge as their training strings had no ACS relationship to the test strings. It was predicted that the aware match group would reveal significantly more explicit rule knowledge than either the control or unaware match groups, who were both expected to show equivalent chance levels of explicit knowledge. Finally, it was anticipated that the aware match group would show positive correlations between both their explicit knowledge and subjective confidence and the accuracy of their classification performance.

Method

Participants. Ninety-nine students at University College London performed the experiment as part of their 1st-year research methods class. Although participants were not paid for taking part in the experiment, a £20 book token was offered to the student who gave the most accurate answers to the questionnaire. Participants were randomly assigned to a control group (N = 29) or a match group (N = 70).

Procedure. Both groups were trained and tested with the match and classification tasks used in Experiments 1 and 2. All participants then completed a questionnaire that examined how much explicit knowledge they had of the rules of the grammar and how confident they were that their rule knowledge was accurate. The only difference between the two groups was that they processed different training strings.

Questionnaire. Participants were told that the strings that they saw during training were constructed from the six letters D, F, G, L, K, and X and asked to answer the following six questions and to guess if they did not know the answer. If there was a D in position 1, what letter appeared in position 5? If there was an F in position 2, what letter appeared in position 6? If there was a G in position 3, what letter appeared in position 7? If there was a K in position 5, what letter appeared in position 1? If there was an L in position 7, what letter appeared in position 3? If there was an X in position 8, what letter appeared in position 4? (Two of the eight possible questions were chosen at random and omitted in order to reduce the total number of questions participants had to respond to). Participants were then asked to say how accurate they thought they had been in specifying the rules by placing a mark on a horizontal line. The line was 16-cm long, with a zero at the left-hand end, indicating ‘‘I do not know any rules,’’ and 100 at the right-hand end, indicating ‘‘I am certain that all my answers are correct.’’

Match group training strings. The letter strings used in Experiment 2 were modified slightly,[2] while retaining the use of a subset of 18 of the possible 36 bigrams that can be created from D, F, G, L, K, and X (see Appendix C). Again, two ungrammatical training strings were created for each grammatical string, violating the rules for one- and two-letter pairs. The violations were made so that the ungrammatical training strings comprised predominantly the same subset of bigrams and trigrams as the grammatical training strings.

Control group training strings. The control group was trained on 36 new letter strings that had no relationship to the test strings in terms of the biconditional rules, whole items, or ACS. The rule-related letter positions (1–5, 2–6, 3–7, and 4–8) contained all possible pairings of the six letters (D, F, G, K, L, and X). All 30 of the bigrams that can be created from the six letters D, F, G, K, L, and X, without using double letters (e.g., DD), were used to construct the strings (see Appendix C)

Classification test strings. The test strings were the same as in Experiment 2. Appendix C specifies how similar versus dissimilar test strings differed in ACS in relation to the match but not the control group's training strings. There was no difference between ACS for grammatical versus ungrammatical test strings for either group.

Results

The mean percentage of training items on which participants in the control and match groups made correct responses are shown in Table 1. A two-way ANOVA comparing accuracy in the match task, with group (control or match) as a between-subjects variable and training block (1 to 4) as a within-subjects variable, indicated that there was a significant effect of block, $F(3, 291) = 11.61$, $MS_e = 56.50$, but no effect of group, $F(1, 97) = 2.73$, $MS_e = 257.77$, and no Group × Block interaction, $F < 1$. The performance of both groups improved across training blocks.

[2] In Experiment 2 we did not constrain the grammatical and ungrammatical strings to be balanced for whole-item similarity, nor did we constrain the high versus low ACS strings to be balanced for whole-item similarity. In Experiment 3 the strings were constrained in this respect.

Match participants' explicit knowledge of the rules of the grammar was assessed by marking their six rule-based questions in relation to the specific set of rules (of 15 different sets) they experienced. All control participants' answers were assessed against the rule set of D↔F, G↔L, and K↔X. Match participants were allocated to an unaware group ($N = 54$) if they answered less than four questions correctly and to an aware group ($N = 16$) if they answered four or more questions correctly. This cutoff point was selected as it created mean correct questionnaire scores for the unaware match group ($M = 1.09$, $SE = 0.13$) that did not differ from the control group ($M = 0.97$, $SE = 0.16$), $t < 1$, while the aware match group ($M = 5.06$, $SE = 0.23$) was reliably more accurate than both the unaware match group, $t(68) = 14.72$, and the control group, $t(43) = 14.76$.

Table 2 presents the mean percentage of items classified as grammatical for each test item type and the overall mean percentage of correct classification responses for the control, unaware match, and aware match groups. The mean percentages of correct responses were 49% for the control group, 50% for the unaware match group, and 53% for the aware match group. The confidence intervals indicate that the control and unaware match groups were classifying at chance levels and that the aware group was just performing at better than chance.

A three-way ANOVA comparing the percentage of items classified as grammatical (ratings ≤ 3), with group (control, unaware match, and aware match) as a between-subjects variable and both grammaticality and ACS as within-subjects variables, found significant effects of group, $F(2, 96) = 10.74$, $MS_e = 829.20$, and ACS, $F(1, 96) = 32.57$, $MS_e = 415.43$, and significant Group $\times$ ACS, $F(2, 96) = 13.22$, $MS_e = 415.43$, Grammaticality $\times$ ACS, $F(1, 96) = 5.94$, $MS_e = 90.63$, and Group $\times$ Grammaticality $\times$ ACS, $F(2, 96) = 4.95$, $MS_e = 90.63$, interactions. There was no effect of grammaticality, $F < 1$, and no Group $\times$ Grammaticality interaction, $F(2, 96) = 2.59$, $MS_e = 111.61$. A two-way ANOVA for the control group, with both grammaticality and ACS as within-subjects variables, indicated that there was no effect of grammaticality or ACS and no Grammaticality $\times$ ACS interaction, with $F < 1$ in all three cases. A comparable ANOVA for the unaware match group indicated that there was a significant effect of ACS, $F(1, 53) = 55.19$, $MS_e = 464.94$, but no effect of grammaticality, $F < 1$, and no Grammaticality $\times$ ACS interaction, $F(1, 53) = 1.50$, $MS_e = 89.91$. Finally, for the aware match group there was a significant effect of ACS, $F(1, 15) = 9.52$, $MS_e = 641.13$, a marginal effect of grammaticality, $F(1, 15) = 4.14$, $MS_e = 115.67$, $p = .06$, and a Grammaticality $\times$ ACS interaction, $F(1, 15) = 9.45$, $MS_e = 103.30$. The latter interaction derives from the fact that the aware match participants were less likely to call ungrammatical/similar test strings grammatical ($M = 45\%$, $SE = 4.84$) than grammatical/similar strings ($M = 59\%$, $SE = 4.07$), $t(15) = 3.58$.

Signal-detection measures were calculated to assess sensitivity in judging

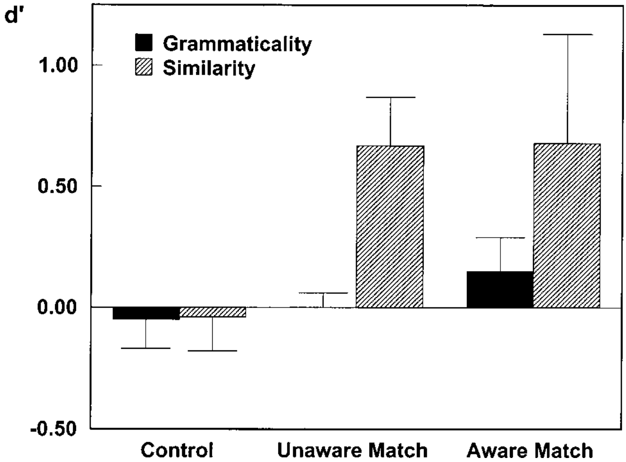


FIG. 5. Mean d' for grammaticality-based (d'_g) and similarity-based (d'_s) classification in the control, unaware match, and aware match groups of Experiment 3. Error bars represent 95% confidence intervals.

the grammaticality of test strings (d'_g). Both the control and unaware match groups showed chance levels of sensitivity to the rules of the grammar, in contrast to the aware match participants who were sensitive to the rules. The control group showed a bias toward calling all test strings grammatical while the match participants showed a bias toward calling strings ungrammatical.

The accuracy of classification performance on the basis of sensitivity to ACS is shown in Table 2 and was 49% for the control group, 61% for the unaware match group, and 60% for the aware match group. Figure 5 shows that compared to the control group, both the unaware and aware match groups were significantly more sensitive to ACS. This is confirmed by the d'_s scores.

The questionnaire answers were analyzed to assess whether the aware match group used implicit or explicit knowledge of the rules of the grammar in the classification test. There were positive one-tailed correlations between the percentage of items classified correctly and both the number of rule-related questions answered correctly ($r = .46$, $p = .04$), and the subjective confidence in the accuracy of the explicit knowledge ($r = .44$, $p = .04$).

The Grammaticality $\times$ ACS interaction in the aware match group ANOVA suggested that participants in this group were able to use their explicit rule knowledge to suppress calling ungrammatical/high ACS test strings grammatical. The accuracy of this group's performance was examined to see whether they had abstracted the rules of the grammar during the training phase or whether they had consciously looked for them during the test, after being told that the training strings had been constructed according to a set of rules. A comparison of the aware group's performance between the two blocks of 12 ungrammatical/similar test trials indicated that there was no difference in classification accuracy between test block 1 (M = 46% correct trials, SE = 4.87) and block 2 (M = 45%, SE = 5.84), $t(15)$ = .22. This suggests that the aware match group acquired their explicit rule knowledge during training.

Discussion

The classification test, subjective confidence ratings, and questionnaire answers indicated that the aware match group had explicit knowledge of the rules of the grammar, while the unaware match group had a chance level of rule knowledge. Both match groups classified on the basis of ACS. There were reliable positive relationships between the aware match group's test accuracy and both their explicit rule knowledge and their subjective confidence in the accuracy of their rule knowledge.

There are three important points to be made about these results. First, there is absolutely no evidence for implicit abstraction of the rules of the grammar. Participants who used rules in classification performance could explicitly specify at least four of those rules and gave subjective confidence ratings that correlated with the accuracy of their rule knowledge. We therefore have evidence that their rule knowledge is explicit by both verbal report and subjective confidence ratings. At the same time, participants who were unaware of the rules did not discriminate grammatical from ungrammatical strings.

Second, the design of the classification test pitted rule knowledge against the perceptual fluency of high ACS ungrammatical test strings. The unaware match participants were significantly more likely to be swayed by the familiarity of the chunks in ungrammatical/similar strings and to call these strings grammatical, whereas the aware match participants were more able to use their explicit rule knowledge to override chunk familiarity.

Third, these results raise the issue of why rule learners in this experiment showed a strong effect of ACS, whereas rule learners in Experiment 2 did not. Whittlesea and Dorken (1993) suggested that every act of learning carries with it a change in the potential to perform an infinite number of possible future activities. When participants memorize grammatical letter strings they encode information about the training stimuli that indirectly (or incidentally) gives them an ability to process related stimuli in an unanticipated classification test, but the participants' goal in the training task is not the direct acquisition of a classification skill. This account suggests that the edit/learners in Experiment 2 showed no effect of ACS in their classification performance, while the aware match group in Experiment 3 did because of differences in the way they processed training strings. The edit/learners in Experiment 2 presumably processed training strings by glancing from letter positions 1 to 5, 2 to 6, 3 to 7, and 4 to 8 to check that the letters conformed to the three biconditional rules of the grammar. This means that they will not have processed the letter strings in the sequential left-to-right manner necessary for them to become familiar with the two- and three-letter contiguous fragments embodied in the ACS measure. In contrast, participants in the match group in Experiment 3 will have processed the training strings in the sequential left-to-right manner necessary to carry out both the demands of the memorization training task and to accumulate knowledge of the distributional statistics of the strings (i.e., ACS).

EXPERIMENT 4

The results of Experiments 2 and 3 clearly demonstrated that when participants memorized grammatical letter strings they acquired knowledge of letter fragments, but these experiments do not tell us whether that letter fragment knowledge was implicit or explicit. This question is explored in the present experiment.

Perruchet and Pacteau (1990, Experiment 3) trained participants on grammatical letter strings generated from a finite-state grammar that used five letters. Participants' explicit fragment knowledge was then tested by asking them to recognize which of 25 two-letter fragments (bigrams) they had seen during training and which fragments were new. In fact, 14 of the 25 bigrams were old and 11 were new. The frequency of occurrence of the old bigrams in the training strings varied from 1 to 6. Participants were able to differentiate between old and new bigrams and their accuracy was correlated with the frequency of occurrence of the old bigrams in training items. These results suggest that bigram knowledge is explicit and that the strength of that explicit knowledge is related to ACS.

The aim of the present experiment was to address this question using a biconditional grammar. A match group trained on the letter strings used in Experiment 2 and then carried out 60 trials of a bigram recognition task. The 60 trials comprised two blocks of 30 randomly presented bigrams. Eighteen of the 30 bigrams were old, as they had been seen during training, and 12 bigrams were new, as they were seen for the first time during the recognition test. On each trial participants indicated whether they believed each bigram was old or new. As the classification results in Experiments 2 and 3 suggested that match nonlearners met the demands of their test instructions by classifying familiar fragments as grammatical and unfamiliar test fragments as ungrammatical, it was predicted that in this experiment match nonlearners would recognize old bigrams on the basis of the episodic knowledge (Whittlesea, 1997a, 1997b) gained by fragment rehearsal processing during training.

Method

Participants. Twelve students at University College London performed the experiment and were paid £5 for taking part.

Materials. The same training letter strings were used as in Experiment 2. Thirty bigrams were used for the recognition task. Eighteen of these test bigrams had been used to construct training strings and were therefore ''old'' (e.g., DF and KG), while 12 test bigrams had not been seen during training and were therefore ''new'' (e.g., DG and KF), though they were constructed from the same individual letters. Each participant saw the bigram set that matched the specific string set that they saw in training (of 15 different sets of rules).

Procedure. All participants were trained using the standard match task and they then carried out a new recognition test.

Recognition test. Participants were told that they would be presented with 60 letter pairs and asked to indicate whether they had seen each letter pair in their training phase or not. They were told that they should not worry if they found this task difficult and should try to base their judgment on how familiar the letter pair felt to them. The test comprised two blocks of 30 bigrams presented in a different random order across blocks and between participants.

Each letter pair was presented in turn and participants were asked to rate how confident they were that they had seen it in the first part of the experiment, using the following scale: (1) *certain,* (2) *fairly certain,* (3) *guess that I have seen this letter pair before,* (4) *guess,* (5) *fairly certain,* and (6) *certain that I have not seen this letter pair before.* In this part of the experiment they were not told whether their responses were correct.

Results

Training responses were scored as correct if the identical string as that initially presented for rehearsal was selected from the list. Table 1 shows the mean percentage of items correctly selected across four blocks of 18 trials. A one-way ANOVA found no overall effect of block, $F(3, 33) = 1.05$, $MS_e = 31.18$.

The mean percentage of items correctly recognized as old or new was 63%, $SE = 3.12$, with a d' sensitivity score of 0.67 (CI 0.16–1.18), indicating above-chance performance. The criterion score c ($M = -0.97$, $SE = 0.17$) indicates that participants were biased toward calling test items ''old.'' A related t test comparing the percentage of old bigrams correctly recognized as old (hits, $M = 85\%$, $SE = 4.72$) with the percentage of new bigrams incorrectly recognized as old (false alarms, $M = 69\%$, $SE = 6.25$) indicated that participants could reliably discriminate between old and new bigrams, $t(11) = 2.41$. These results support the prediction that match participants would acquire explicit bigram knowledge, as they were able to discriminate between bigrams they had seen during training and bigrams that only appeared in the recognition test.

One possible objection to this conclusion is that recognition need not always be based on a conscious recollection process but may instead be based on an automatic familiarity process (Jacoby, 1991; Hintzman & Curran, 1994). Participants may simply find old chunks more familiar and on that basis call them old without necessarily consciously recollecting them. However, this view is unlikely to be appropriate because Knowlton and Squire (1996) have shown that amnesics are significantly worse than controls at chunk recognition. On the assumption that conscious recollection is far more impaired than familiarity in amnesia (Aggleton & Shaw, 1996; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1988), Knowlton and Squire's data suggest that chunk recognition is an explicit memory task and is probably little contaminated by familiarity. Moreover, Yonelinas (1997) has shown that in contrast to item recognition, associative recognition in normal participants is based purely on recollection and is not mediated at all by familiarity. The pair recognition test used in the present experiment is an associative recognition task because participants have to recognize whether a *pair* of letters (e.g., LF) occurred together in training; the individual letters themselves (L, F, etc.) all occurred in training.

EXPERIMENT 5

This experiment returns to the question of whether participants who mentally rehearse training strings memorize whole training exemplars. In Experiment 1 classification test items were designed to manipulate the similarity of whole test items to whole training items, orthogonal to the grammaticality of test strings, while balancing ACS across all test item types. Similar test items differed from *one* training string by only two letters, while dissimilar test items differed from *all* training items by three or more letters. The match group showed no effect of whole-item similarity in their classification test performance.

The aim of this experiment was to provide a stronger test of the whole-item account by investigating whether memorizing letter strings would result in a whole-item similarity effect when there were six similar training strings, each differing by only one letter from each similar test item. For example the test string DDKL.GFFL was similar to the six training strings KDKL.GFFL, DDGL.GFFL, DDKD.GFFL, DDKL.GLFL, DDKL.GFGL, and DDKL.GFFF. In addition test strings were constructed to manipulate ACS orthogonally to whole training-item similarity. Unlike in the previous experiments, training strings were not constructed according to the biconditional rules.

Participants were asked to memorize the 108 letter strings shown in Appendix D. These letter strings were constructed from a subset of 18 of the 36 bigrams that can be created from the six letters D, F, G, K, L, and X. After training participants were given the standard classification instructions used in Experiments 1–3 and asked to classify test strings as grammatical or ungrammatical. The classification test stimuli comprised 18 high-whole-item similarity/high-ACS (HIHA) strings that each overlapped by seven letters with six training strings and also shared a high number of letter fragments with the training strings, 18 low-whole-item/high-ACS (LIHA) test strings that were dissimilar to all training strings (i.e., they differed from all training strings by at least three letters), but shared a high number of letter fragments with training items, and a third set of low-whole-item/low-ACS (LILA) strings that were dissimilar to all training items and had minimal overlap of two- and three-letter fragments with training items.

On the basis of the results of Experiments 2–4 we predicted that participants would be more likely to classify high-ACS (LIHA) letter strings as grammatical than low ACS (LILA) strings. In contrast, we had no firm prediction about whether participants would show a whole-item similarity effect or not in the HIHA/LIHA comparison. We had created three sets of test strings that would enable us to unconfound the contributions of whole-item similarity and ACS in classification performance. If classification performance was also partly mediated by whole-item similarity then participants would classify more HIHA than LIHA items as grammatical. If participants classified solely on the basis of ACS, however, then there would be no difference in performance on HIHA and LIHA items.

Method

Participants. Twelve students at University College London performed the experiment and were each paid £5 for taking part.

Procedure. The same match and classification tasks were used as in Experiments 1–3.

Materials. Three hundred twenty-four training strings were created using a subset of 18 of the 36 bigrams that can be constructed from the six letters D, F, G, K, L, and X (see Appendix D). Unlike Experiments 1–4, double letters (e.g., DD) were used in this experiment. One hundred eight of the training strings were each presented for mental rehearsal in the match task. The remaining 216 strings were the distractors in the list of three strings presented at the end of each match trial. Half of the distractor strings differed from the initial training string by one letter and the other half differed from the original training string by two letters.

Three sets of test strings were created (see Appendix D). Eighteen high-item/high-ACS (HIHA) strings each overlapped with six training strings on seven letters and also overlapped with all training strings in terms of ACS. Eighteen low-item/high-ACS (LIHA) strings were dissimilar to all training strings, but overlapped with all training items in terms of ACS. A third set of low-item/low-ACS (LILA) strings was dissimilar to all training items and did not overlap with the training items in terms of ACS. Appendix D shows that HIHA and LIHA test items differ in whole-item similarity but not ACS, while LIHA and LILA test items have equivalent low levels of whole-item similarity but differ in ACS.

Results

Table 1 shows the mean percentage of training items on which participants made correct responses. A one-way ANOVA, with block (27 training trials) as a within-subjects variable, found no overall effect of block, $F(3, 33) = 2.38$, $MS_e = 32.63$.

The mean percentage of items classified as grammatical was 73% ($SE = 4.80$) for HIHA, 73% ($SE = 3.87$) for LIHA, and 31% ($SE = 5.12$) for LILA test items. If participants classified on the basis of knowledge of whole training items then they would have called HIHA strings grammatical and LIHA strings ungrammatical; we therefore reanalyzed the results for these items taking ‘‘grammatical’’ responses to HIHA and ‘‘ungrammatical’’ responses to LIHA items as ‘‘correct.’’ The mean percentage correct calculated in this way was exactly 50%. This suggests that participants did not memorize whole training items. On the other hand, if participants were classifying on the basis of ACS then they would classify LIHA strings as grammatical and LILA strings as ungrammatical and this in fact was the case as the mean percentage correct calculated in this way on these items was 71%.

Related t tests indicated that there was no difference in the percentages of HIHA and LIHA items classified as grammatical, $t(11) < 1$, but that there was a reliable difference in the percentages of LIHA and LILA items classified as grammatical, $t(11) = 5.13$. The mean d' scores indicated chance sensitivity to whole-item similarity, $d'_{item} = 0.02$ (CI -0.16–0.20), and above chance sensitivity to ACS, $d'_{acs} = 1.30$ (CI 0.67–1.93). Despite the fact that there were six similar training items for each similar test items, participants classified solely on the basis of fragment knowledge.

GENERAL DISCUSSION

We set out to reevaluate the key issues that have driven AGL research over the past 30 years using a finite-state grammar which does not constrain transitions between adjacent letters. Our first aim was to investigate claims that in incidental learning conditions participants passively abstract the underlying structure of rule-based concepts and in a later classification test apply this knowledge unconsciously (e.g., Cleeremans, 1993; Lewicki & Hill, 1989; Reber, 1967, 1989). These claims were compared to exemplar accounts which predict that participants encode a collection of training examples (Brooks, 1978; Brooks & Vokey, 1991; Neal & Hesketh, 1997; Vokey & Brooks, 1992) and to fragment explanations which predict that participants compile frequency counts of two- and three-letter chunks from training examples (e.g., Dulany, Carlson, & Dewey, 1984; Perruchet & Pacteau, 1990; Redington & Chater, 1996). We introduced a new methodology based on a biconditional grammar that allowed us to unconfound the various forms of knowledge available in training examples (see Johnstone & Shanks, 1999).

Experiments 1–3 manipulated the grammaticality of test items orthogonal to similarity to whole training exemplars (Experiment 1) and similarity to fragments of two or three letters (Experiments 2–3). There was no evidence that participants passively abstracted implicit knowledge of the rules of the grammar. Out of a total of 86 match participants in these experiments, 17 showed an effect of grammaticality in their classification performance and in all 17 cases knowledge of the grammar was explicit. These results challenge previous suggestions that participants abstract the rules of the grammar, based on tests with the same letter set as in training (e.g., Gomez & Schvaneveldt, 1994; Knowlton & Squire, 1994; Mathews et al., 1989; Reber, 1967).

Many people have suggested that transfer studies in which the letter set is changed at test provide evidence of acquisition of abstract knowledge (e.g., Brooks & Vokey, 1991; Gomez & Schvaneveldt, 1994; Knowlton & Squire, 1996; Manza & Reber, 1997; Marcus et al., 1999; Mathews et al., 1989; Reber, 1969; Reber & Lewis, 1977). But, how can this be the case when experiments with the same letter set at test fail to yield evidence of rule abstraction? In our view the evidence from the transfer studies can be adequately explained without requiring rule abstraction at study. Brooks and Vokey (1991); Gomez, Gerken, and Schvaneveldt (2000); and Tunney and Altmann (1999) have shown that much of the transfer to ''changed letter-set'' strings is due to abstract similarity between test and training strings. For example the abstract structure of MXVVVM is similar to that of BDCCCB. Whittlesea and Wright (1997, Experiment 2) manipulated repetition patterns orthogonal to rules and found that classification performance was influenced by repetition. They also pointed out that standard finite-state grammars, such as that created by Reber and Allen (1978), produce massive repetition of letter patterns (e.g., MTTVT, MTVRXM, MTVRXRM, MTTVRXRM, and MTV) that are likely to capture a participant's attention. Furthermore, Gomez (1997) has shown that above-chance transfer is invariably associated with above-chance performance on tests of explicit knowledge (e.g., recognition tests). Hence transfer studies do not challenge the claim that *implicit* abstraction is impossible.

Our studies found no evidence of examplar knowledge as participants failed to show a similarity effect in either Experiment 1, where half of the test items only differed from *one* training string by one letter, or in Experiment 5, where half of the test items only differed from *six* training items, again by one letter. These results challenge exemplar models of classification performance (e.g., Brooks, 1978; Brooks & Vokey, 1991; McAndrews & Moscovitch, 1985; Neal & Hesketh, 1997; Vokey & Brooks, 1992) and support findings of no exemplar effects when grammatical and ungrammatical test strings are equated for fragment knowledge (Knowlton & Squire, 1994; Shanks, Johnstone, & Staggs, 1997).

In contrast to the failure to find evidence for implicit rule or exemplar knowledge, there was strong support for the notion that participants acquire knowledge of two- and three-letter fragments (Experiments 2–5) and confirmation that knowledge of two-letter fragments is explicit (Experiment 4). These results support accounts of classification performance based solely on fragment knowledge (e.g., Dulany, Carlson, & Dewey, 1984; Johnstone & Shanks, 1999; Perruchet, 1994; Perruchet & Pacteau, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990) and undermine claims that fragment knowledge alone is not sufficient to account for classification performance (e.g., Gomez & Schvaneveldt, 1994; Knowlton & Squire, 1994, 1996; Mathews et al., 1989; Meulemans & Van der Linden, 1997; Reber & Allen, 1978). These findings also challenge suggestions that knowledge of simple associations (Gomez, 1997) or more complex fragments (e.g., Servan-Schreiber & Anderson, 1990) is implicit.

It could, of course, be argued that our conclusion that abstract rules are not learned in implicit AGL tasks applies to the biconditional grammar but not to more standard transition-rule grammars such as those shown in Fig.

1 and 2. Perhaps the biconditional grammar is sufficiently different from transition rule grammars that a rather different set of mental processes is engaged, and hence our results do not directly undermine the claim that rules are learned when strings are generated from transitional grammars. We argue, in contrast, that previous research has cast considerable doubt on this claim and that the present results strongly reinforce that doubt. While we acknowledge that the processes engaged by the 2 sorts of grammar may be subtly different, it is crucial to bear in mind that the biconditional grammar can be learned (under edit conditions) and that this can only be achieved by learning an abstract rule of exactly the sort proposed by Manza and Reber (1997), Marcus et al. (1999), and others (see ''Evidence for Rule Knowledge'').

Moreover, in two studies which have directly compared biconditional and transitional grammars under otherwise identical conditions (Experiments 3 versus 4 of Mathews et al., 1989; Experiments 3 versus 4 of Shanks et al., 1997) there is no evidence of qualitative differences in behavior. Instead, these studies simply suggest that the biconditional grammar is much ''sharper'' at demonstrating the influence of different sources of control such as rules and fragment properties.

The Episodic-Processing Account

Our second aim was to assess the episodic-processing account (Whittlesea, 1997a, 1997b; Whittlesea & Dorken, 1993) which predicts that structural aspects of training stimuli will only be encoded to the extent that such information enables participants to meet the demands imposed by the training instructions. Successful test performance is assumed to depend on test instructions reinstating the processing context experienced during training. We evaluated these assumptions by comparing the effects of passively memorizing letter strings with active hypothesis testing to discover the rules of a grammar (Experiments 1–2). The key question was whether edit learners would classify on the basis of the same structural knowledge as match nonlearners or whether there would be differences in the form of knowledge acquired that could be explained by differences in the demands of the training instructions. The results indicated that training instructions determined the form of knowledge acquired, as match nonlearners classified on the basis of fragment knowledge but not rules while edit learners used rule but not fragment knowledge. This double dissociation is entirely consistent with the episodic processing account.

Match participants were told that they were taking part in a short-term memory experiment and were simply instructed to memorize training strings. At the end of the training phase they were told that the stimuli they had seen were constructed according to a set of rules, but they were not told what those rules were and they were not given any specific instructions about how to perform the test. The classification results in Experiments 2 and 3 suggest that match nonlearners met the demands of the test instructions by classifying letter strings with familiar training fragments as grammatical and unfamiliar fragments as ungrammatical. Thus, match nonlearners classified on the basis of episodic knowledge of letter fragments (and rehearsal processes).

In contrast, the edit particpants were asked to hypothesis test in order to identify the rules of the grammar. Those who succeeded in hypothesis testing subsequently carried out a number of training trials where they successfully applied the rules to flawed training strings by glancing from locations 1 to 5, 2 to 6, 3 to 7, and 4 to 8 to check whether these paired locations contained legal letter pairs. When the rule learners were later asked to classify novel test items as grammatical or ungrammatical, they could apply the same processing to the same structural aspects of stimuli as was demanded by the training task.

Effective Concept Learning Strategies

We are grateful to Mathews et al. (1989, Experiments 3–4) for creating the match and edit tasks and biconditional grammar that we have used extensively in our own research (Shanks, Johnstone, & Staggs, 1997). In addition we believe that an analysis of our combined results can provide valuable insights into factors that determine successful training strategies in real-world settings as the match and edit tasks are useful analogs of incidental learning from experience versus active scientific investigation.

First, previous research has been taken to support the notion that a complex rule structure can be learned by memorizing representative examples (e.g., Reber, 1967, 1969). This suggests that incidental learning from experience will similarly lead to accurate knowledge of the underlying structure of a complex concept. However we would caution against this conclusion as our results demonstrate that memorizing examples leads to knowledge of the distributional statistics of surface features, but not to rule knowledge. In our experiments participants who memorized training strings used knowledge of the two- and three-letter fragments in training items to make decisions about test items.

We believe that prior studies that appeared to provide evidence of incidental rule learning were based on designs that confounded rule and fragment knowledge (e.g., Meulemans & Van der Linden, 1997). In practice this means that people may appear to have learned the rules of a concept when actually they have only learned about patterns of family resemblance that correlate with the rules of the concept (e.g., Mathews et al., 1989, Experiment 3). This has implications for many real-world situations such as young children learning their native language and medical practitioners learning disease categories.

Second, the optimal selection of a training strategy is partially dependent on the complexity of the rules governing the domain. For example, the rule structures of typical finite-state transition grammars are too complex to be learned by hypothesis testing (e.g., Brooks, 1978; Mathews et al., 1989, Experiment 3; Reber, 1976; Reber, Kassin, Lewis, & Cantor, 1980; Shanks et al., 1997, Experiment 3), whereas the rules of a biconditional grammar can be learned by hypothesis testing (e.g., Mathews et al., 1989, Experiment 4; Shanks et al. 1997, Experiment 4). Further research is required to clarify why hypothesis testing fails with transition rule grammars. These two types of grammar differ in the number of rules required to specify grammatical structure. The biconditional grammar used in this article can be specified with only three letter-pair rules, whereas a typical transition grammar is based on many more rules. For example, there are 23 rules in the grammar created by Brooks and Vokey (1991) (see Johnstone & Shanks, 1999, for a fuller explanation).

Third, there are large individual differences in hypothesis-testing ability. Only 50% of edit participants succeeded in hypothesis testing in our studies. Those who failed appeared to be trying to impose arbitrary hypotheses on the training stimuli rather than acknowledging that they must start by guessing and then try to identify what hypotheses might explain the feedback. Some people may need preliminary training in hypothesis testing before they can benefit from rule-learning instructions. Alternatively an ‘‘apply rule’’ strategy can be adopted. We have run experiments where participants are told the rules of the grammar at the beginning of the experiment and then asked to apply these rules by correcting flawed training strings. This method leads to perfect classification performance by all participants.

Fourth, accuracy of test performance is determined by the overlap between training and test processing. In our experiments the match and edit groups processed exactly the same letter strings during training yet the match groups encoded fragment knowledge and edit/learners encoded the rules of the grammar. At test match participants used the same rehearsal processes as they had used to process training strings which led to chance performance. In contrast, edit/learners used the rule-checking processes they had practiced during hypothesis testing and this led to near-perfect levels of accuracy.

Finally, previous applied research suggests that a practical strategy for teaching skills that are based on a large number of complex rules (i.e., rule structures that are similar to transition grammars) is to identify and train people on a small number of rules that lead to performance that is ‘‘good enough,’’ though not perfect. For example, Biederman and Shiffrar (1987) demonstrated that naive subjects could determine the sex of day-old chickens with 90% accuracy based on one simple rule, while professional sexers took 2.4 months to reach 95% accuracy based on feedback on examples.

CONCLUSIONS

The results of five experiments challenge previous explanations of AGL based on implicit, general, rule knowledge and on knowledge of a collection of whole exemplars. Instead, our findings support Whittlesea's (1997a, 1997b; Whittlesea & Dorken, 1993) episodic-processing account that predicts that participants will process stimuli in ways that meet the demands of the training instructions and subsequently succeed or fail in test performance to the extent that test instructions reinstate the strategies participants used to meet those demands.

APPENDIX A

Experiment 1 Training and Test Strings

String Statistics

The mean overlaps for the match and edit groups' grammatical versus ungrammatical test strings with their training strings were anchor ACS, $M = .36$, $SE = .03$ versus $M = .38$, $SE = .02$; global ACS, $M = 2.28$, $SE = .05$ versus $M = 2.32$, $SE = .04$. For similar versus dissimilar test strings, the overlaps were anchor ACS, $M = .37$, $SE = .02$ versus $M = .37$, $SE = .02$; global ACS, $M = 2.34$, $SE = .05$ versus $M = 2.26$, $SE = .05$. A two-way ANOVA on anchor ACS, with both grammaticality and whole-item similarity as between-test-item measures, indicated that there was no effect of grammaticality, or whole-item similarity, and no significant Grammaticality x Whole-item Similarity interaction, with $F < 1$ in all three cases. A similar ANOVA on global ACS, indicated that there was no effect of grammaticality, $F < 1$, or whole-item similarity, $F(1, 140) = 1.69$, $MS_e = .161$, and no significant Grammaticality $\times$ Whole-item Similarity interaction, $F < 1$.

In the control group, the overlaps for grammatical versus ungrammatical test strings with their training strings were: anchor ACS, $M = .69$, $SE = .05$ versus $M = .64$, $SE = .04$; global ACS, $M = 4.83$, $SE = .09$ versus $M = 4.80$, $SE = .07$. For similar versus dissimilar test strings the overlaps were anchor ACS, $M = .61$, $SE = .04$ versus $M = .72$, $SE = .05$; global ACS, $M = 4.89$, $SE = .08$ versus $M = 4.73$, $SE = .07$. A two-way ANOVA for anchor ACS indicated no effect of grammaticality, $F < 1$, or whole-item similarity, $F(1, 68) = 2.64$, $MS_e = .07$, and no Grammaticality $\times$ Whole-item Similarity interaction, $F < 1$. A similar ANOVA for global ACS indicated that there was no effect of grammaticality or whole-item similarity and no significant Grammaticality $\times$ Whole-item Similarity interaction, with $F < 1$ in all three cases.

TABLE A1
Experiment 1: Match and Edit Group Training Strings

	Match rehearsal/ edit correct string	Distractor 1/ hypothesis test 1	Distractor 2/ hypothesis test 2
List 1	DFGK.FDLX	LFGK.FDLX	DFGX.FGLX
	DGKX.FLXK	DFKX.FLXK	LGKX.FLDK
	DKFL.FXDG	DKXL.FXDG	DKFG.KXDG
	FDXG.DFKL	FDXK.DFKL	FDLG.DGKL
	FLDK.DGFX	FLDK.LGFX	FLDX.DGKX
	FXLD.DKGF	FXLD.DXGF	FXLG.DKLF
	GKDF.LXFD	GKDF.LXGD	XKDF.LKFD
	GLFX.LGDK	GLFX.LGDF	DLFX.LGFK
	GXKL.LKXG	DXKL.LKXG	GXDL.LFXG
	KLXD.XGKF	KGXD.XGKF	KLGD.XGKL
	KXGL.XKLG	KXDL.XKLG	KXGD.XFLG
	KDLF.XFGD	KDLX.XFGD	KXLF.LFGD
	LFDG.GDFL	LFDG.KDFL	LFDX.GDKL
	LGXF.GLKD	LGXF.GFKD	KGXF.GLFD
	LKGX.GXLK	LKGX.GXDK	FKGX.GDLK
	XDKG.KFXL	XDKG.KFXD	XDFG.KDXL
	XFLK.KDGX	GFLK.KDGX	XFLG.KLGX
	XGFD.KLDF	XFFD.KLDF	LGFD.KLDX
List 2	KXFG.XKDL	KXLG.XKDL	KLFG.GKDL
	XDGK.KFLX	XDGF.KFLX	FDGK.KFDX
	LDKF.GFXD	LDKF.KFXD	LDXF.GKXD
	GFKX.LDXK	GFKX.LFXK	DFKX.LDXF
	KFLD.XDGF	KFLD.XDLF	KFLG.XLGF
	DFXL.FDKG	DFXL.FDKX	KFXL.FXKG
	LGKD.GLXF	XGKD.GLXF	LGKF.GLDF
	XGLF.KLGD	XDLF.KLGD	KGLF.KLXD
	FGXD.DLKF	FGLD.DLKF	FGXL.GLKF
	DKLX.FXGK	DKLG.FXGK	DFLX.FXDK
	LKFG.GXDL	LKFG.KXDL	LKFX.GKDL
	FKDL.DXFG	FKDL.DXLG	FGDL.DXFL
	GLXK.LGKX	GLXK.LGKF	LGXK.LGKX
	FLGX.DGLK	KLGX.DGLK	FLKX.DGLX
	XLDG.KGFL	XKDG.KGFL	XLDK.KGXL
	GXDK.LKFX	GXDK.LGFX	LXDK.LKFG
	KDFL.XFDG	KDXL.XFDG	KDFG.XLDG
	DXGF.FKLD	DXGF.FXLD	KXGF.FKXD

Note. All of the training and test strings in this article are reported using the three rules that D is paired with F, G with L, and K with X. In practice, in all five experiments, 15 different sets of these three rules were used. Rule sets were matched across participants in the Match and Edit groups.

TABLE A2
Experiment 1: Control Group Training Strings

Rehearsal string	Distractor 1	Distractor 2
DLGK.DGKL	GLGK.DGKL	DLGK.DGFX
DGXD.GXLD	DKXD.GXLD	DGXD.FGLD
DKFL.KGLX	DKGL.KGLX	DKXD.KGLX
DFLG.FXGD	DFLX.FXGD	XKLG.FXGD
DLXK.LFXL	DLXK.DFXL	FLDK.LFXL
DXKD.XLFG	DXKD.XKFG	DGKF.XLFG
GXLF.DFKX	GXLF.DFGX	GXKF.LFKX
GDXL.GLDK	GDXL.GLDF	GDXF.GXDK
GDLX.KDFL	KDLX.KDFL	GDLX.GDKL
GKDL.FLDX	GXDL.FLDX	GKDL.FGDF
GFDK.LGXD	GFLK.LGXD	LFDK.LGKD
GLFK.XLKF	GLFD.XLKF	GXFK.XLKD
KLFG.DXFK	KLFG.LXFK	GLFG.DXFD
KXGL.GKFL	KXGL.GXFL	FXGF.GKFL
KGDX.KFLK	KGDX.KFGK	KFDG.KFLK
KDGL.FGLX	KDGL.FGLK	KDFL.FDLX
KGFX.LKDG	LGFX.LKDG	KGFD.LKFG
KFXD.XFDK	KLXD.XFDK	KFXD.LFDL
FKXF.DKFD	FKLF.DKFD	DKXF.DGFD
FLDF.GKLF	FLDL.GKLF	FGDF.GKXF
FXDL.KFGD	FXDL.DFGD	FXGL.KFGK
FKLX.FDLF	FKLX.FKLF	DKLX.KDLF
FDKF.LXDL	FDKF.LXFL	FXKF.LFDL
FGLK.XGDK	FGLK.XGDF	FGDK.XGLK
LGXK.DLKG	KGXK.DLKG	LGXL.DLKX
LKGX.GFDX	LFGX.GFDX	LFGX.KFDX
LFKG.KLXF	LFXG.KLXF	LFDG.KGXF
LXFD.FXKL	LXFG.FXKL	LGFD.FXDL
LXKF.LGFX	LXKF.XGFX	FXKF.LDFX
LDFK.XKGD	LDFK.XLGD	LGFK.XKLD
XDGL.DFXG	XDGL.DFKG	XDKL.DFXD
XGKL.GDXK	XGKL.GDXL	XGKD.GFXK
XKLD.KXDF	FKLD.KXDF	XKLX.LXDF
XFGK.FKXF	XDGK.FKXF	XFLK.FDXF
XLKG.LDGL	XLKG.LDFL	DLKG.LDXL
XFDG.XDKG	XFDG.XDKL	XLDG.XDKX

TABLE A3
Experiment 1: Classification Test Items

	Grammatical test items	Ungrammatical test items
List 1	LFGK.GDLX	LFGK.KDLX
	DLKX.FGXK	DFKX.FGXK
	DKGL.FXLG	DKGL.FXKG
	FDXL.DFKG	FDXK.DFKG
	FGDK.DLFX	FGDK.DKFX
	FKLD.DXGF	FGLD.DXGF
	XKDF.KXFD	XKDF.GXFD
	GLDX.LGFK	GLKX.LGFK
	GXKF.LKXD	GXKD.LKXD
	KLGD.XGLF	KLFD.XGLF
	FXGL.DKLG	FXGL.FKLG
	KDLX.XFGK	KDLG.XFGK
	LFXG.GDKL	LFXG.GDXL
	LGDF.GLFD	LGKF.GLFD
	LKGD.GXLF	LKGD.GXLD
	XFKG.KDXL	XLKG.KDXL
	XFLG.KDGL	XFLG.KDGF
	XKFD.KXDF	XLFD.KXDF
List 2	DXFG.FKDL	LXFG.FKDL
	FDGK.DFLX	FDGK.GFLX
	GDKF.LFXD	XDKF.LFXD
	GDKX.LFXK	GDKX.LGXK
	KGLD.XLGF	KXLD.XLGF
	DFKL.FDXG	DFKL.FDLG
	LXKD.GKXF	LFKD.GKXF
	XGDF.KLFD	XGDF.KLXD
	FGXL.DLKG	FGXK.DLKG
	DKLF.FXGD	DKLF.FXGL
	LKXG.GXKL	LKDG.GXKL
	FKDX.DXFK	FKDX.DXFL
	GLXF.LGKD	GLXD.LGKD
	KLGX.XGLK	KLGX.FGLK
	XLFG.KGDL	XLKG.KGDL
	GXLK.LKGX	GXLK.LKDX
	KGFL.XLDG	KXFL.XLDG
	DKGF.FXLD	DKGF.FGLD

APPENDIX B

Experiment 2: Letter Strings

String Statistics

The overlaps for grammatical versus ungrammatical test strings with training strings were anchor ACS, $M = 1.81$, $SE = .37$ versus $M = 1.79$, $SE = .38$; global ACS, $M = 11.89$, $SE = 2.19$ versus $M = 11.83$, $SE = 2.18$. The overlaps for the high- versus low-ACS test strings were anchor ACS, $M = 3.5$, $SE = .16$ versus $M = 0.10$, $SE = .04$; global ACS, $M = 22.24$, $SE = .41$ versus $M = 1.48$, $SE = .11$. A two-way ANOVA for anchor ACS, with both grammaticality and ACS (high versus low) as between-test-item measures, showed an effect of ACS, $F(1, 44) = 393.35$, $MS_e = .09$, but there was no effect of grammaticality, $F < 1$, and no Grammaticality $\times$ ACS interaction, $F < 1$. A similar ANOVA for global ACS showed a significant effect of ACS, $F(1, 44) = 2330.71$, $MS_e = .55$, but no effect of grammaticality, $F < 1$, and no Grammaticality x ACS interaction, $F < 1$.

TABLE B1
Experiment 2: Training Strings

Match rehearsal/ edit correct string	Distractor 1/ hypothesis test 1	Distractor 2/ hypothesis test 2
DFGD.FDLF	LFGD.FDLF	XFKD.FDLF
DFKD.FDXF	DLKD.FDXF	XFKD.FDLF
DFKX.FDXK	XFKX.FDXK	DFKD.FDLK
DLGX.FGLK	DLGL.FGLK	DLGX.LGLG
DLKD.FGXF	DXKD.FGXF	DLFD.LGXF
DLFD.FGDF	DLFD.FGDL	DXFD.FGDX
GDFG.LFDL	KDFG.LFDL	GDFG.DFGL
GLKX.LGXK	GLKX.FGXK	XLKG.LGXK
GLFD.LGDF	GLKD.LGDF	GXFD.LGDX
GXKG.LKXL	GXLG.LKXL	GXKX.LGXL
GXFG.LKDL	GXFG.LKGL	KXFG.LKGL
GXLG.LKGL	GXLG.LKGD	GDLG.LKGX
KDLG.XFGL	KDLG.XFDL	GDLG.XFDL
KDXK.XFKX	KDXK.XLKX	KGXK.XFKD
KGLK.XLGX	FGLK.XLGX	LGLK.XLKX
KXFD.XKDF	KXFK.XKDF	DXFD.XKDL
KXFK.XKDX	KXFK.XFDX	KXLK.XFDX
KXLG.XKGL	KXLG.LKGL	KGLG.XKXL
FDLK.DFGX	FDLK.DFKX	FDFK.DFGD
FDXF.DFKD	FGXF.DFKD	FKXF.DFKX
FDLF.DFGD	FGLF.DFGD	FDXF.DFGX
FGDF.DLFD	FGDF.DLFK	FKDF.DLFG
FGXK.DLKX	FGXK.DLKG	FGXF.DFKX
FKXF.DXKD	FDXF.DXKD	LKXF.DFKD
LGDF.GLFD	LFDF.GLFD	LGDL.GLKD
LGXK.GLKX	LGXK.XLKX	KGXL.GLKX
LGXF.GLKD	LGXF.DLKD	LKXF.GLFD
LKDL.GXFG	LKDL.GLFG	LKDF.GXKG
LKGL.GXLG	LKGL.GDLG	LKGX.GDLG
LFDL.GDFG	DFDL.GDFG	LGDL.KDFG
XKDL.KXFG	XKDF.KXFG	XKGL.KDFG
XKDX.KXFK	XKDL.KXFK	LKDX.KDFK
XKGL.KXLG	XKDL.KXLG	XKDL.GXLG
XFGL.KDLG	XFDL.KDLG	XFGL.GDFG
XFGX.KDLK	XFGX.KDXK	XFGX.FDXK
XLKX.KGXK	XLKX.KGXF	XFKX.KGXL

TABLE B2
Experiments 2–4: Classification Test Items

	Grammatical	Ungrammatical
High ACS	DFGL.FDLG	DFGL.FDLF
	DLKX.FGXK	DLKX.FDXK
	GLKD.LGXK	GLKD.FGXF
	GXFD.LKDF	GXKD.LKDF
	KGXK.XLKX	KDXK.XLKX
	KXFG.XKDL	GXFG.XKDL
	FDLG.DFGL	FDLK.DFGL
	FGXF.DLKD	FGXF.GLKD
	LKDF.GXFD	LKDF.GXKD
	LKXF.GXKD	LKXF.GXKG
	XFDL.KDFG	XKDL.KDFG
	XFKX.KDXK	XFKX.KGXK
Low ACS	DGKL.FLXG	DGKL.DLXG
	DXGK.FKLX	DXGK.FKLD
	GDKF.LFXD	GDKF.LFXG
	GKFX.LXDK	GKLX.LXDK
	KGFL.XLDG	DGFL.XLDG
	KFLD.XDGF	KFXD.XDGF
	FKLX.DXGK	FKLX.DXGF
	FLXG.DGKL	FLDG.DGKL
	LDGK.GFLX	LDGK.GKLX
	LXDK.GKFX	LXDK.GKLX
	XDGF.KFLD	XDGF.KFXD
	XGKF.KLXD	XDKF.KLXD

APPENDIX C

Experiment 3: Training and Test Strings

String Statistics

The overlaps between the match group training strings and grammatical versus ungrammatical test strings were anchor ACS, $M = 0.53$, $SE = .13$ versus $M = 0.70$, $SE = .17$; global ACS, $M = 5.47$, $SE = .97$ versus $M = 5.46$, $SE = .97$. For similar versus dissimilar test strings the overlaps were anchor ACS, $M = 1.15$, $SE = .15$ versus $M = 0.08$, $SE = .04$; global ACS, $M = 10.04$, $SE = .25$ versus $M = 0.88$, $SE = .07$. A two-way ANOVA for anchor ACS, with both grammaticality and anchor ACS as between-test-item measures, indicated that there was a significant effect of anchor ACS, $F(1, 44) = 50.06$, $MS_e = .27$, but no effect of grammaticality, $F(1, 44) = 1.23$, $MS_e = .27$, and no Grammaticality $\times$ Anchor ACS interaction, $F(1, 44) = 1.23$, $MS_e = .27$. A similar ANOVA for global ACS indicated that there was an effect of global ACS, $F(1, 44) = 1205.38$, $MS_e = .84$, but no effect of grammaticality, $F < 1$, and no Grammaticality x Global ACS interaction, $F < 1$.

The overlaps between the control group training strings and grammatical versus ungrammatical test strings were anchor ACS, $M = 0.67$, $SE = .05$ versus $M = 0.68$, $SE = .05$; global ACS, $M = 4.91$, $SE = .05$ versus $M = 4.86$, $SE = .08$. For similar versus dissimilar test strings the overlaps were anchor ACS, $M = 0.61$, $SE = .06$ versus $M = 0.73$, $SE = .05$; global ACS, $M = 4.90$, $SE = .07$ versus $M = 4.87$, $SE = .07$. A two-way ANOVA for anchor ACS, indicated no effect of grammaticality, $F < 1$, or ACS, $F(1, 44) = 2.32$, $MS_e = .07$, and no Grammaticality $\times$ ACS interaction, $F < 1$. A similar ANOVA for global ACS indicated that there was no effect of grammaticality or ACS and no Grammaticality $\times$ ACS interaction, with $F < 1$ in all three cases.

TABLE C1
Experiment 3: Match Group Training Strings

Rehearsal string	Distractor 1	Distractor 2
DFGD.FDLF	KFGD.FDLF	XKGD.FDLF
DFGX.FDLK	DLGX.FDLK	DFLK.FDLK
DFKD.FDXF	DFXD.FDXF	DFKD.GKXF
DFKD.FDXF	DFKG.FDXF	DFKD.FDLG
DLGD.FGLF	DLGD.KGLF	KLXD.FGLF
DLFD.FGDF	DLFD.FXDF	DGFL.FGDF
DXKD.FKXF	DXKD.FKDF	DXGD.LKXF
DXKD.FKXF	DXKD.FKXG	DXKF.FDXF
GDFG.LFDL	GDFG.LFDX	GDFG.KFXL
GDFG.LFDL	GDFG.LFKL	GDFG.LKDG
GLFG.LGDL	GLFG.LFDL	DLFX.LGDL
GLFG.LGDL	GLFG.XGDL	GKFG.KGDL
GXKG.LKXL	GXKD.LKXL	GXDG.LDXL
GXLG.LKGL	GXKG.LKGL	GXLF.LKFL
GXLG.LKGL	GFLG.LKGL	GXLG.DKGD
KDFK.XFDX	GDFK.XFDX	XDFK.KFDX
KDFK.XFDX	LDFK.XFDX	KGFK.XKDX
KDLK.XFGX	KGLK.XFGX	KDXK.XFDX
KDXK.XFKX	KDFK.XFKX	KDXF.XFKL
KXLK.XKGX	KXLG.XKGX	DXLK.XKGL
KXLK.XKGX	KXLK.DKGX	KFLK.XKDX
FDLK.DFGX	FDLK.DKGX	FDGK.DLGX
FDLF.DFGD	FDLF.DFLD	FDLG.KFGD
FDLF.DFGD	FDLF.DFGX	GDLK.DFGD
FDXK.DFKX	FDXK.DFKL	FDXK.LFKG
FGLK.DLGX	FGLK.DLFX	FKLX.DLGX
FGLK.DLGX	FGLK.DKGX	FGLK.GLDX
LGDL.GLFG	LGDL.DLFG	LKFL.GLFG
LGDL.GLFG	LGDF.GLFG	LGDL.GXKG
LGXL.GLKG	LGKL.GLKG	LKXL.GXKG
LKGL.GXLG	LDGL.GXLG	LKFL.GXDG
LKXL.GXKG	XKXL.GXKG	LKDL.GXDG
XKDL.KXFG	XFDL.KXFG	FKDG.KXFG
XKDL.KXFG	XKFL.KXFG	XKDL.LXFD
XLGX.KGLK	XLGX.KDLK	FLGX.KGLF
XLGX.KGLK	XLGX.KGDK	XLGK.XGLK

TABLE C2
Experiment 3: Control Group Training Items

Rehearsal string	Distractor 1	Distractor 2
DLGK.DGKL	GLGK.DGKL	DLGK.DGFX
DGXD.GXLD	DKXD.GXLD	DGXD.FGLD
DKFL.KGLX	DKGL.KGLX	DKXD.KGLX
DFLG.DXGD	DFLX.DXGD	XKLG.DXGD
DLXK.LFXL	DLXK.DFXL	FLDK.LFXL
DXKD.XLFG	DXKD.XKFG	DGKF.XLFG
GXLF.DFKX	GXLF.DFGX	GXKF.LFKX
GLXL.GLDK	GLXL.GLDF	GLXF.GXDK
GDLX.KDFL	KDLX.KDFL	GDLX.GDKL
GKDL.FLDX	GXDL.FLDX	GKDL.FGDF
GFDK.LGXD	GFLK.LGXD	LFDK.LGKD
GLFK.XLK	GLFD.XLKF	GXFK.XLKD
KLFG.DXFK	KLFG.LXFK	GLFG.DXFD
KXGL.GKFL	KXGL.GXFL	FXGF.GKFL
KGDX.KFLK	KGDX.KFGK	KFDG.KFLK
KDGL.FGLX	KDGL.FGLK	KDFL.FDLX
KGFX.LKDG	LGFX.LKDG	KGFD.LKFG
KFXD.XFDK	KLXD.XFDK	KFXD.LFDL
FKXF.DKFD	FKLF.DKFD	DKXF.DGFD
FLDF.GKLF	FLDL.GKLF	FGDF.GKXF
FXDL.KFGD	FXDL.DFGD	FXGL.KFGK
FKLX.FDLF	FKLX.FKLF	DKLX.KDLF
FDKF.LXDL	FDKF.LXFL	FXKF.LFDL
FGLK.XGDK	FGLK.XGDF	FGDK.XGLK
LGXK.DLKG	KGXK.DLKG	LGXL.DLKX
LKGX.GFDX	LFGX.GFDX	LFGX.KFDX
LFKG.KLXF	LFXG.KLXF	LFDG.KGXF
LXFD.FXKL	LXFG.FXKL	LGFD.FXDL
LXKF.LGFX	LXKF.XGFX	FXKF.LDFX
LDFK.XKGD	LDFK.XLGD	LGFK.XKLD
XDGF.DFXG	XDGF.DFKG	XDKF.DFXD
XGKL.GDXK	XGKL.GDXL	XGKD.GFXK
XKLD.KXDF	FKLD.KXDF	XKLX.LXDF
XFGK.FKXF	XDGK.FKXF	XFLK.FDXF
XLKG.LDGL	XLKG.LDFL	DLKG.LDXL
XFDG.XDKG	XFDG.XDKL	XLDG.XDKX

APPENDIX D

Experiment 5: Training and Test Strings

String Statistics

Across the three sets of test strings (HIHA, LIHA, and LILA) respectively, the mean anchor ACS was 4.89, $SE = 1.15$; 4.65, $SE = 1.10$; and 0.0, $SE = 0$, while the mean global ACS was 28.46, $SE = 6.71$; 28.16, $SE = 6.64$; and 0.0, $SE = 0$. There was no difference in anchor ACS between HIHA and LIHA items, $t(34) = 1.07$, $SE = .22$, while there was a reliable difference between the anchor ACS of the LIHA and LILA items, $t(34) = 29.20$, $SE = .16$. There was no difference in global ACS between HIHA and LIHA items, $t(34) < 1$, $SE = .46$, while there was a difference in global ACS between LIHA and LILA items, $t(34) = 79.51$, $SE = .35$.

TABLE D1
Experiment 5: Training Strings

Set	Rehearsal String	Distractor 1	Distractor 2
1	KDKL.GFFL	DDKL.GFFL	KDGL.GFFL
1	DDGL.GFFL	DDKL.GFFL	KDKL.GFFL
1	DDKD.GFFL	DDKD.GLFL	DDKD.GLFL
1	DDKL.GLFL	XDKL.GLFL	DDKD.GFFL
1	DDKL.GFGL	DDGL.GFGL	DDGL.GFFL
1	DDKL.GFFF	DDKL.GFFL	DDKD.GLFF
2	FGFG.LKDD	FGFG.LKXD	DGFG.LKXD
2	DGFF.LKDD	DGFG.LKDD	DGFG.XKDD
2	DGLG.LKDD	DGLG.LKXD	DGFF.LKDD
2	DGFG.XKDD	DGFG.LKDD	FGFG.XKDD
2	DGFG.LKXD	DGFG.LKDD	DGLG.LKDD
2	DGFG.LKDK	DGFG.LKXK	DGFF.LKXK
3	XKLG.FFLK	DKLG.FFLK	DKLG.LFLK
3	DGLG.FFLK	DKLG.FFLK	DKLG.LFLK
3	DKLG.LFLK	DKLK.LFLK	DKLG.FFLF
3	DKLG.FGLK	DKLG.FFLK	DKDG.FFLK
3	DKLG.FFLF	DKLG.LFLF	DKLG.FGLK
3	DKDG.FFLK	DKLG.FFLK	DKLG.LFLK
4	LFLK.DDGF	LKLK.DDGF	GFLK.DDGX
4	GFLK.XDGF	GFLK.DDGF	GFLK.DDGX
4	FFLK.DDGF	FFLK.XDGF	FFLK.XDGX
4	GFLK.DDGX	GFLK.DDGF	LFLK.DDGF
4	GFLK.DDGF	GFLK.XDGF	FFLK.XDGF
4	GFLK.DDGX	FFLK.DDGX	GFLK.XDGF
5	KLGF.FLKX	KLGF.FLGX	GFGF.FLKX
5	GFGF.FLKX	GLGF.FLKX	KLGF.FLKX
5	GLFF.FLKX	GLGF.FLKX	GFGF.FLKX
5	GLGL.FLKX	GLGF.FLKX	GFGF.FLKX
5	GLGF.GLKX	GLGF.FLKX	GLGF.FLGX
5	GLGF.FLGX	GLGL.FLGX	GLGF.GLKX
6	KXDD.KLFG	KXDD.GLFG	GXKD.KLFG
6	GXKD.KLFG	GXDD.KLFG	KXDD.KLFG
6	GXDD.GLFG	GXDD.KLFG	GXDD.KLFF
6	GXDD.KLFF	GXDD.KLFG	GXDD.GLFG
6	GXDD.KLFL	KXDD.KLFL	GXKD.GLFL
6	XXDD.KLFG	XXDD.GLFG	XXKD.KLFL
7	XDDG.XXKL	KDDG.XXKL	XXDG.XDKL
7	KXDG.XXKL	KDDG.XXKL	XXDG.XDKL
7	KDDK.XXKL	KDDG.XXKL	KXDG.XXKL
7	KDDG.XDKL	KDDK.XDKL	KDDK.XXKL
7	DDDG.XXKL	DDDG.XDKL	XDDK.XXKL

TABLE D1 *Continued*

Set	Rehearsal String	Distractor 1	Distractor 2
7	KDDG.XDKL	KDDG.XXKL	KDDK.XXKL
8	GLFF.LGXD	FLFF.LGXD	KLGF.LGXD
8	FLFF.LGXD	GLFF.LGXD	FLGF.LKXD
8	KLGF.LGXD	KLFF.LGXD	GLGF.LKXD
8	KLFG.LGXD	KLFF.LGXD	KLFF.FGXD
8	KLFF.FGXD	KLFF.LGXD	KLFG.LGXD
8	KLFF.LKXD	KLFF.LGXD	KLFG.LGXD
9	GXXD.GLGL	GXXD.GFGL	XXDD.GLGL
9	XXXD.GLGL	XXXD.KLGL	XXDD.GFGL
9	KXDD.GLGL	KXKD.GLGL	KXXD.KLGL
9	KXKD.GLGL	KXDD.GLGL	KXDD.KLGL
9	KXXD.KLGL	KXXD.GLGL	KXKD.GLGL
9	KXXD.GFGL	KXXD.GLGL	GXXD.GLGL
10	FGXK.LFGF	FGXK.LFGL	FGLK.LFGL
10	FGXK.LFGL	FGXK.LFGF	FGLK.LFGF
10	FGLK.LFGX	FGXK.LFGX	DGLK.LFGL
10	DGXK.LFGX	DGLK.LFGX	DGLK.LFGL
10	LGXK.LFGX	LGXK.LFGL	DGLK.LFGX
10	FGLK.LFGX	FGLK.LFGF	FGXK.LFGF
11	FFFL.KXDG	FFGL.KXDG	FFGL.GXDG
11	FFGX.KXDG	FFGL.KXDG	FFFL.KXDG
11	FFGL.KXDG	FFGX.KXDG	FLGX.KXDG
11	FFGL.GXDG	FFGL.KXDG	FLGL.KXDG
11	FFGL.KDDG	FFGL.KXDG	FFGL.GXDG
11	FFGL.KXDK	FFGL.KXDG	FFGX.KDDK
12	FLGX.DGLF	FLGX.DKLF	GLKX.DGLF
12	FLKD.DGLF	FLKD.DKLF	FLKX.DKLF
12	FLKX.DKLF	FLGX.DKLF	GLKD.DKLF
12	FLKX.DGFF	FLKX.DGLF	FLKX.DKLF
12	GLKX.DGLF	FLKX.DGLF	FLKX.DGFF
12	KLKX.DGLF	KLKX.DKLF	KLGX.DGFF
13	LKLF.GXXK	LFLF.GXXK	LFLF.GXDK
13	LFLF.GXXK	LFLF.GXDK	LKLF.GXXD
13	LGLF.GXDK	LKLF.GXDK	DGLF.GXXK
13	DGLF.GXXK	LGLF.GXXK	LGLF.GXDK
13	FGLF.GXXK	FGLF.GXXD	LGLF.GXDK
13	LGLF.GXXD	LGLF.GXXK	FGLF.GXXK
14	LKLK.XDDK	LKLK.DDDK	LKXK.XKDK
14	LKXK.XDDK	LKDK.XDDK	LKLK.XKDK
14	LKDG.XDDK	LKDK.XDDK	LKXK.XDDK
14	LKDK.DDDK	LKDK.XDDK	LKDG.XDDK
14	LKDK.XKDK	LKDK.XXDK	LKDG.XDDK
14	LKDK.XXDK	LKDK.XKDK	LKDG.XKDK
15	LGFL.FGFF	LFFL.FGFF	LFFG.FGFF
15	LFFL.FGLF	LGFL.FGLF	LGFG.FGLF
15	LFFL.FFFF	LFFL.FGFF	LFGL.FFLF
15	LFFL.FLFF	LFGL.FLFF	LGFL.FGFF
15	LFFG.FGFF	LFFL.FGFF	LFFL.FLFF
15	LFGL.FGFF	LFFL.FGFF	LFFG.FGFF

TABLE D1 *Continued*

Set	Rehearsal String	Distractor 1	Distractor 2
16	XDGX.DKLG	XDGX.XKLG	XDGX.XKDG
16	XDGX.XKDG	XDGX.XKLG	XDKX.XKLG
16	XDKX.XKLG	XDKX.XKDG	XDGX.XKLF
16	XDGX.XKLK	XDGX.XKLG	XDGX.XKDG
16	XDGX.XKLF	XDKX.XKLF	KDGX.XKLG
16	KDGX.XKLG	KDGX.XKLF	XDGX.XKDG
17	XXXX.KDKD	XKXX.KDKD	XKXD.KDKD
17	XKXX.DDKD	XKXX.KDKD	XXXX.KDKD
17	XKXX.XDKD	XKXX.DDKD	XKXX.KDDD
17	XKXX.KDDD	XKXX.XDDD	XKXX.XDKD
17	XKXD.KDKD	XKXX.KDKD	XKXX.KDDD
17	XKXX.KDKD	XKXD.KDKD	XKXK.XDKD
18	XDKD.DKXX	XXKD.DKXX	XXKD.DKXD
18	XXKX.DKXX	XDKX.DKXX	XXKD.DKXK
18	XXDD.DKXX	XXXD.DKXX	XDKD.DKXX
18	XXXD.DKXX	XXKD.DKXX	XXKX.DKXX
18	XXKD.DKXD	XXKD.DKXK	XXKX.DKXK
18	XXKD.DKXK	XXKD.DKXD	XDKD.DKXD

TABLE D2
Experiment 5: Classification Test Items

High item similarity/ high ACS similarity	Low item similarity/ high ACS similarity	Low item similarity/ low ACS similarity
DDKL.GFFL	LFFG.LKDD	DFKG.DXGG
DGFG.LKDD	XXKX.DDGX	DLLX.GKKF
DKLG.FFLK	FFLK.DDKL	DXGG.KFDL
GFLK.DDGF	FGLG.FFFL	GDFK.GGKK
GLGF.FLKX	XKLG.FGLF	GGDF.KKGD
GXDD.KLFG	GFLK.DDKL	GKKF.DLXG
KDDG.XXKL	LKXX.KDDK	KGGK.KGDX
KLFF.LGXD	DDGL.FFLK	KKGD.FXLL
KXXD.GLGL	LKXD.GFFL	KFXG.GDLX
FGXK.LFGX	XDGL.KXDK	FDLL.XGGK
FFGL.KXDG	GFLK.LFGX	FKFD.LXFD
FLKX.DGLF	FLGX.XKLF	FXLD.XFKG
LGLF.GXXK	KLGX.DDGF	LDXL.LDFK
LKDK.XDDK	KDGX.KDKD	LLXF.XLLD
LFFL.FGFF	FFGF.LKXD	LXFX.LLDF
XDGX.XKLG	DGLK.XKDK	XGKK.FDXF
XKXX.KDKD	DKDK.XXKL	XFDL.DFXL
XXKD.DKXX	FGLK.XDGX	XLDX.FKFX

REFERENCES

Abrams, M., & Reber, A. S. (1989). Implicit learning in special populations. *Journal of Psycholinguistic Research,* **17,** 425–439.

Aggleton, J. P., & Shaw, C. (1996). Amnesia and recognition memory: A reanalysis of psychometric data. *Neuropsychologia,* **34,** 51–62.

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 899–912.

Biedermann, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **13,** 640–645.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorisation* (pp. 169–211). New York: Wiley.

Brooks, L. R., & Vokey, 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.

Cleeremans, A. (1993). *Mechanisms of implicit learning.* Cambridge, MA: MIT Press.

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

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

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.

Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology,* **33,** 154–207.

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

Gomez, R. L., & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 396–410.

Hintzman, D. L. (1986). ‘‘Schema abstraction’’ in a multiple-trace memory model. *Psychological Review,* **93,** 411–428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple trace memory model. *Psychological Review,* **95,** 528–551.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language,* **33,** 1–18.

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

Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **25,** 524–531.

Kinder, A., & Shanks, D. R. (2000). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Manuscript submitted for publication.*

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for instances. *Psychological Science,* **3,** 172–179.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 79–91.

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, and Cognition,* **22,** 169–181.

Lewicki, P., & Hill, T. (1989). On the status of nonconscious processes in human cognition: Comment on Reber. *Journal of Experimental Psychology: General,* **118,** 239–241.

Manza, L., & Reber, A. S. (1997). Representing artificial grammars: Transfer across stimulus forms and modalities. In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 73–106). Oxford, England: Oxford Univ. Press.

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

McAndrews, M. P., & Moscovitch, M. (1985). Rule-based and exemplar-based classification in artificial grammar learning. *Memory & Cognition,* **13,** 469–475.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review,* **102,** 419–457.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science,* **283,** 77–80.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review,* **85,** 207–238.

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

Morris, C. C., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behaviour,* **16,** 519–533.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review,* **92,** 289–316.

Neal, A., & Hesketh, B. (1997). Episodic knowledge and implicit learning. *Psychonomic Bulletin and Review,* **4,** 24–37.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorisation relationship. *Journal of Experimental Psychology: General,* **115,** 39–57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14,** 700–708.

Perruchet, P. (1994). Defining the units of a synthetic language: Comment on Vokey and Brooks (1992). *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 223–228.

Perruchet, P., Gallego, J., & Pacteau, C. (1992). A reinterpretation of some earlier evidence for abstractiveness of implicitly acquired knowledge. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology,* **44A,** 193–210.

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. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory,* **2,** 88–94.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General,* **118,** 218–235.

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition,* **6,** 189–221.

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.

Reber, A. S., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition,* **5,** 333–361.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General,* **125,** 123–138.

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

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology,* **50A,** 216–252.

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

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science,* **16,** 1–40.

St. John, M. F., & Shanks, D. R. (1997). Implicit learning from an information processing standpoint. In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 162 – 194). Oxford, England: Oxford Univ. Press.

Treves, A., & Rolls, E. T. (1994). Computational analyses of the role of the hippocampus in memory. *Hippocampus,* **4,** 374–391.

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

Tunney, R. J., & Altmann, G. T. M. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **25,** 1322–1333.

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

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **18,** 328–344.

Vokey, J. R., & Brooks, L. R. (1994). Fragmentary knowledge and the processing-specific control of structural sensitivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 1504–1510.

Whittlesea, B. W. A. (1997a). The representation of general and particular knowledge. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 335–370). Hove, England: Psychology Press.

Whittlesea, B. W. A. (1997b). Production, evaluation, and preservation of experiences: Constructive processing in remembering and performance tasks. In D. Medin (Ed.), *The psychology of learning and motivation* (Vol. 37, pp. 211–264). San Diego, CA: Academic Press.

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

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica,* **98,** 141–165.

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

Wright, R., & Whittlesea, B. W. A. (1998). Implicit learning of complex structures: Active adaptation and selective processing in acquisition and application. *Memory & Cognition,* **26,** 402–420.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition,* **25,** 747–763.

Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation and receiver operating characteristic data. *Neuropsychology,* **12,** 323–339.

Accepted June 27, 2000; published online January 4, 2001