Structural versus Processing Accounts of Implicit Learning

A dissertation submitted for

the degree of Doctor of Philosophy

at the University of London

by

Theresa Johnstone

University College London

September 1999

Abstract

Artificial grammar learning (AGL) experiments are used to investigate the possibility that separate general and specific learning systems exist that store implicit, abstract rule knowledge versus specific, episodic knowledge. Chapter 1 concludes that the most urgent priority is to settle continuing debates about whether memorisation of grammatical examples leads to rule, exemplar, or fragment knowledge. In Chapter 2, it is recommended that a biconditional grammar be used to settle the knowledge debate, as rule, exemplar, and fragment knowledge are inevitably correlated in finite-state grammar generated stimuli.

Chapters 3 and 4 present evidence that memorising grammatical training examples leads to fragment, but not rule or exemplar knowledge. In contrast, active hypothesis testing is required to gain rule knowledge. Chapters 5 and 6 demonstrate that knowledge gained by memorising training examples is explicit according to objective recognition, cued-recall, and subjective confidence tests.

Using a biconditional grammar, there is no support for a dichotomy between general and specific learning systems. The results are best explained by one episodic-processing system that records processing of specific structural aspects of grammatical items in order to meet the demands of the training task. During classification, knowledge may appear to be implicit when participants are unaware of the source of fluent processing (i.e., when they unconsciously use fragment knowledge to classify test items as grammatical or ungrammatical). However, instructions that draw attention to the relevant knowledge (i.e., in cued-recall and recognition tests) reveal that classification knowledge is explicit.

Table of Contents | *Page Numbers*

List of Tables and Figures

Appendices

Preface

The work described in this dissertation was carried out in the Psychology Department at University College London, between 1996 and 1999, under the supervision of Professor David Shanks.

This dissertation has not been submitted in whole or in part for a degree or diploma or other qualification at any other university. Many of the ideas expressed here are those of others, and this has been indicated, as far as possible, by reference to the appropriate sources. I carried out all of the work.

Acknowledgements

The research reported in this dissertation was supported by the Medical Research Council. I am grateful to the Psychology Department at University College London for the provision of research facilities. I would like to take this opportunity to thank Professor David Shanks for supervising this research and in so doing helping me to develop my intellectual skills. It goes without saying that the errors of content and style in this dissertation are mine alone.

People have an impressive capacity for storing information about particular events. This "episodic" memory (Tulving, 1983) allows us to recall the context of specific experiences, such as what we did on our last holiday. We also have an ability to acquire general knowledge, that is, properties 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 abstract, general properties such as the rules of a grammar that are separate from knowledge of specific objects or events.

Cognitive psychology has traditionally dealt with this distinction by assuming separate processes for acquiring specific and general knowledge. Under various terms (e.g., episodic, explicit, declarative), knowledge of specific events is assumed to be distinct from knowledge about general properties (e.g., semantic, implicit, procedural). A puzzle, however, is to explain how we acquire general knowledge as abstract properties themselves are never directly observed (see Whittlesea, 1997a, b). 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, as we are not normally deliberately intending to perform such abstraction, it must be largely an incidental and unconscious process.

Undoubtedly, there is a wealth of evidence consistent with this 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 by asking them to memorise strings of letters generated from the finite-state grammar shown in Figure 1. This grammar specifies rules for ordering string elements such as those that 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.

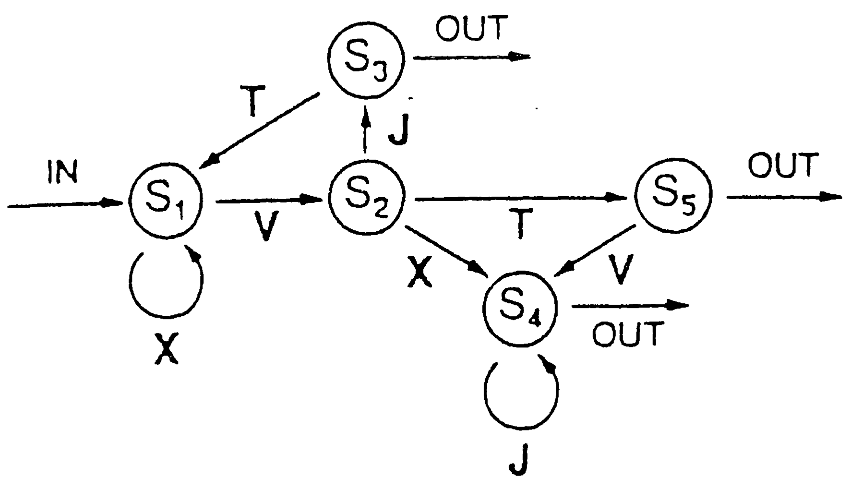

Figure 1. This artificial grammar was used by Knowlton, Ramus, and Squire (1992), and was originally created by Abrams and Reber (1989).

Knowlton, Ramus, and Squire (1992) tested specific knowledge by asking participants to recognise 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 to support the idea of separate "implicit", general and "explicit", specific learning systems.

This dual systems theory of implicit learning is based on four major claims. First, it is claimed that implicit knowledge is acquired when participants observe or memorise representative examples of a complex rule-governed concept, without being told that the examples conform to a set of rules (Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997, Experiment 2b; Reber, 1967, 1989; Reber & Allen, 1978; Reber & Lewis, 1977). Secondly, in these incidental learning conditions, participants are passive "consumers" (Lewicki & Hill, 1989, p.240) of stimulus-driven knowledge (Cleeremans, 1993, p.19). Thirdly, it is claimed that an implicit learning system creates mental representations of abstracted knowledge, in parallel with an explicit system that creates representations of specific whole or partial training items in a separate episodic memory. While the empirical work reported in later chapters focuses on the claim that we can acquire implicit rule knowledge (Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997, Experiment 2b; Reber, 1967, 1989; Reber & Allen, 1978; Reber & Lewis, 1977), there are also claims that implicit knowledge is based on abstract patterns of family resemblance (Mathews, Buss, Stanley, Blanchard-Fields, Cho, & Druhan, 1989), first-order dependencies between adjacent letters (Gomez, 1997), or associative chunk strength (Servan-Schreiber & Anderson, 1990).

Finally, it is claimed that participants lack awareness of the knowledge they use to classify test items at above chance levels as they cannot fully state the rules of the grammar (Reber & Lewis, 1977) and accurate performance is accompanied by a subjective experience of guessing (Dienes, Altmann, Kwan, & Goode, 1995). 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.

In contrast to the dual-system theory, exemplar and fragment accounts assume that above-chance classification performance can be explained solely on the basis of episodic knowledge. The exemplar account claims that participants encode a collection of training examples (Brooks, 1978; Brooks & Vokey, 1991; Neal & Hesketh, 1997; McAndrews & Moscovitch, 1985; Vokey & Brooks, 1992) and at test items that are highly similar to training items (e.g., differing by only one letter), are more likely to be called grammatical than dissimilar items. The letter-fragment account (Dulany, Carlson, & Dewey, 1984; Meulemans & Van der Linden, 1997, Experiment 2a, Perruchet, 1994; Perruchet & Pacteau, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990) suggests that participants use specific knowledge of letter fragments seen in training strings to classify test items. In this case, participants are assumed to classify test items containing fragments seen in training as grammatical and test strings containing novel fragments as ungrammatical.

Whereas the implicit rule abstraction, exemplar, and fragment accounts share an assumption that specific aspects of the structure of training examples (rules, exemplars, or letter-fragments) are acquired in a stimulus-driven manner, the episodic-processing account (Whittlesea, 1997a, b; Whittlesea & Dorken, 1993, 1997; Whittlesea & Wright, 1997; Whittlesea & Williams, 1998, in press, Wright & Whittlesea, 1998) suggests that knowledge acquisition is driven by processing. By this account, participants actively process training strings in order to meet the demands of the training task and in so doing acquire knowledge of the specific aspects of training items (rules, exemplars, or letter fragments) necessary to meet those demands. As a result, episodic representations are created combining knowledge of both the processing carried out and the information used to satisfy the training instructions.

At test, items that overlap with training items on the structural aspects encoded during training cue prior processing episodes and as a result are processed more fluently than dissimilar test items that do not cue prior episodes. The knowledge underlying fluency is neither implicit nor explicit. Instead, when participants are unaware of the relationship between fluency and the information they acquired during training they will respond on the basis of familiarity, whereas when they are aware of the relationship they will respond on the basis of recollection.

The assumptions of the episodic-processing account can be illustrated using an AGL example where a participant is asked to memorise a set of letter strings, such as MXRTMXR. In this example, the participant actively meets the demands of the memorisation task by mentally rehearsing each training string left to right as a series of two- and three-letter fragments. Using this strategy, MXRTMXR is rehearsed for example as the three letter fragments MXR, TM and XR and the episodic-processing system is assumed to create episodic representations of the mental rehearsal process and those specific fragments. By the end of the training phase, this participant will have acquired episodic representations of mentally rehearsing a large number of letter fragments leading to an ability to process efficiently future letter strings containing the same letter fragments.

During a later classification test, the assumption is that test items that are similar to training items on the structural dimension used to process training items (letter fragments in this example) will cue episodic representations of processing similar training items. Cueing prior episodes will result in similar test items being processed more efficiently (fluently) than dissimilar test items. Thus, continuing with the example, a test string containing previously seen letter fragments, such as MXR, TM and XR, would cue episodic representations of processing those fragments, whereas a test string containing only novel fragments would not retrieve any episodic representations.

Finally, the suggestion that participants will respond on the basis of subjective feelings of familiarity if they are unaware, or on the basis of recollection if they are aware of the relationship between the information they acquired during training and test demands can be illustrated by comparing classification and recognition test performance. When the participant, in the example, is asked to classify novel test items as grammatical or ungrammatical, he will not understand the relationship between the letter-fragment knowledge gained in training and fluency of processing test items. Consequently, fluency in processing test items containing training fragments will unconsciously be attributed to grammaticality and accompanied by a subjective feeling of familiarity. In contrast, when the participant is asked to discriminate between fragments seen during training and novel fragments in a recognition test, old/new judgements will be based on conscious recollection of fragment knowledge as the test instructions draw attention to the information acquired during training. Recollection is a separate process to fluency and not simply an attribution based on fluency.

There is therefore a similarity between the episodic-processing account (Whittlesea, 1997a, b; Whittlesea & Dorken, 1993, 1997; Whittlesea & Wright, 1997; Whittlesea & Williams, in press, Wright & Whittlesea, 1998) and the rule-abstraction account (e.g., Reber, 1967, 1989) as both accounts predict that knowledge can be applied implicitly in a classification test. However, these two accounts disagree about the form of knowledge (processing episodes versus rules) and whether knowledge is stored in an implicit form (Reber, 1967, 1989) or stored in a neutral form that can be expressed implicitly or explicitly depending on test instructions (Whittlesea & Williams, in press).

Forms of Knowledge

The first stage in evaluating the four accounts of implicit learning has to be identifying what knowledge is acquired during incidental memorisation conditions, as without reliable evidence about the information used to classify test items it is impossible to test whether that knowledge is implicit or explicit (see Shanks & St. John, 1994).

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. While knowledge of what bigrams are allowable can be a form of rule knowledge, this thesis seeks to establish whether participants can acquire the rules of the specific biconditional grammar created by Mathews et al. (1989) by simply memorising training strings without knowing that those training strings were constructed according to a set of rules.

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 accurate 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; McAndrews & Moscovitch, 1985; 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 that 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 design 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 (i.e., two letter bigrams, three letter trigrams etc.) in the training strings and classify novel test strings as grammatical to the extent that test strings contains fragments that were present in the training strings (Dienes, Broadbent, & Berry, 1991; 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 memorised 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 models that 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 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 exemplar knowledge (e.g., Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986) from only one knowledge base.

Meulemans and Van der Linden (1997) have provided the most convincing rule and fragment 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 in a discussion of the methodological problems with finite-state grammars in Chapter 2.

Evidence for Episodic-Processing Knowledge

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 or not 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 memorised 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 a digit and immediately made a judgement about whether the digit was a low number (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 (in press) put forward a discrepancy-attribution hypothesis which suggests that participants will apply knowledge explicitly or implicitly at test depending on whether or not 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., DAISY, DELICATE), orthographically regular easy to pronounce nonwords (e.g., BARDLE, PLEMIDON) and less regular and hence harder to pronounce nonwords (e.g., LICTPUB, MOLPEOT). 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 or not.

As novel regular nonwords (e.g., HENSION) were 21% more likely to be called old than novel words (e.g., TABLE), it was suggested that participants did not expect nonwords to be processed fluently and as a result unconsciously attributed the surprising fluency of orthographically regular nonwords to those items having been pronounced during training. In contrast, there was no discrepancy between the first impression that 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.

Summary

Despite 30 years of AGL research there are still debates about whether participants who memorise representative examples of a complex rule-governed concept, without knowing that those examples conform to a set of rules, classify on the basis of implicit or explicit knowledge. However, before we can determine whether knowledge acquired in incidental learning situations is implicit or explicit it is necessary to gain a better understanding of the form of knowledge (i.e., rules, exemplars, fragments, or processing episodes) used to classify test items. Chapter 2 identifies weaknesses in the standard finite-state grammars typically used to investigate implicit learning and suggests that the debate about forms of knowledge can be more successfully investigated using a biconditional grammar.

The AGL literature review, in Chapter 1, indicated that despite 30 years of research there is still much debate about what participants who memorise rule-based training stimuli actually learn. In this chapter, research by Meulemans and Van der Linden (1997) is analysed to demonstrate that the standard AGL approach of using finite-state grammars is flawed because these grammars do not allow us to unconfound the contributions of rule- and similarity- (exemplar or fragment) based knowledge in classification test performance.

In their Experiments 2a and 2b, Meulemans and Van der Linden made heroic efforts to unconfound the factors of grammaticality and fragment statistics. Half of the test items were grammatical and half were nongrammatical and, orthogonally, half of the test strings were highly associated with training strings (i.e., contained familiar letter fragments) and half were not. This created four sets of test items: grammatical and associated (GA), grammatical and not associated (GNA), nongrammatical and associated (NGA), and nongrammatical and not associated (NGNA). The degree to which test strings were associated to training strings was measured using a statistic called associative chunk strength (ACS; see Experiment 1, Chapter 3 for a detailed explanation of the ACS calculation). This measures the overlap of letter-fragments (chunks) of two (bigrams) and three (trigrams) letters between test and training strings weighted by the number of times a chunk occurred in the training strings. Meulemans and Van der Linden calculated mean ACS statistics for letter chunks that occurred in anchor positions (initial and terminal chunks) and in global positions (anywhere within a letter string). In both experiments ACS varied between associated and nonassociated items, but was balanced across grammatical and nongrammatical strings. Experiments 2a and 2b differed in two respects. First, test items in Experiment 2b contained chunks that had been seen at least once during training, whereas the test items in Experiment 2a contained some novel chunks that had not been seen during training. However, the number of novel chunks was equivalent across grammatical and nongrammatical strings in Experiment 2a. Secondly, the two experiments differed in terms of the number of training strings (32 versus 125 respectively) and training trials (64 versus 125 respectively) experienced by participants.

Table 1
Mean String Characteristics and Percentage of Test Strings Classified as Grammatical by the Experimental Groups in Meulemans and Van der Linden's (1997) Experiments 2a and 2b.

	Grammatical Associated	Grammatical Nonassociated	Nongrammatical Associated	Nongrammatical Nonassociated
Experiment 2a				
Anchor ACS	4.78	3.31	4.44	3.41
Global ACS	9.69	6.17	9.71	6.30
Novelty	0.00	1.25	0.00	1.50
NCP	0.63	1.75	1.38	2.13
Length	6.88	6.00	6.75	6.63
Classification	65.31	42.81	67.50	45.00
Experiment 2b				
Anchor ACS	18.00	13.38	18.25	12.25
Global ACS	35.44	23.43	34.80	23.80
Novelty	0.00	0.00	0.00	0.00
NCP	0.25	0.25	0.63	1.88
Length	6.50	6.50	6.75	6.63
Classification	67.81	62.81	56.25	60.31

Note. ACS = associative chunk strength. NCP = novel chunk position.

Table 1 summarises the test string characteristics and Meulemans and Van der Linden's key findings. The row labelled "Classification" gives the percentages of strings of each type classified by participants as grammatical. On the basis of analyses of variance (ANOVA), Meulemans and Van der Linden concluded that after the short training phase (Experiment 2a) participants classified test strings on the basis of ACS, with no reliable effect of grammaticality: The classification rates closely parallel the ACS measures. However, when participants received extended training (Experiment 2b), they classified test strings on the basis of grammaticality, with no reliable effect of ACS. In this case the classification rates do not parallel the ACS measures, but instead tend to be higher for grammatical than nongrammatical strings. Meulemans and Van der Linden suggested that these results provide evidence of two independent learning mechanisms, which are brought into operation depending on the number of items presented in the training phase. When fewer training strings are presented, classification judgements are based on chunk frequency and overlap with training items, whereas with longer training, performance is based on knowledge of the rules of the grammar.

The Basis for Reappraising Meulemans and Van der Linden's Conclusions

Meulemans and Van der Linden's findings were based on an assumption that the training strings only provided participants with two types of knowledge: grammatical rules and chunk frequency information. However, the training strings also provided information about legal *locations* of chunks within training strings and this finer level of knowledge was not captured by the global ACS measure.

Redington and Chater (1996) used models to demonstrate that evidence of participants classifying at above chance levels, in prior AGL studies, could be explained solely on the basis of participants calling test items that contained entirely familiar chunks grammatical and items that contained at least one novel chunk ungrammatical. For this reason, Meulemans and Van der Linden balanced the number of novel chunks across grammatical and ungrammatical test strings in Experiment 2a (though not between associated and nonassociated items) and ensured that test strings did not contain any novel chunks in Experiment 2b. But these models, like Meulemans and Van der Linden's ACS measures, are not sensitive to familiar training chunks in novel positions within test strings, and a number of researchers have suggested that participants may acquire knowledge of the legal positions of chunks within letter strings (e.g., Dienes, Broadbent, & Berry, 1991; Dulany, Carlson, & Dewey, 1984). So there is a possibility that participants could become sensitive to chunks that they have seen in training being presented in novel locations in test strings. The following analyses will demonstrate that in Experiment 2b grammaticality was confounded with positional information.

Empirical Evidence That Information about Familiar Training Chunks in Novel Test Positions Was Confounded with Grammaticality.

The issue of familiar training chunks appearing in novel test positions can be illustrated by looking at the letter strings used in Experiment 2b. Within the set of training strings, the trigram VXR occurred ten times (MXRMVXR, MVXR, MVXRVVV, MVXRMXT, MVXRMXR, MVXRVMT, MVXRVV, MVXRV, MVXRM, and MVXRVVM). Within these ten training strings, VXR appeared in only two of the five possible locations: nine times as the second trigram and twice as the last trigram (in one string the second trigram is also the last). This means that when VXR appeared as the fourth trigram in four out of the eight NGA test items (MXRVXRM, MXRVXRV, VMRVXRM, VMRVXRV) it was appearing in a novel location. Perhaps participants call these strings nongrammatical not because they violate the rules of the grammar but because they contain familiar chunks in novel locations.

A new statistic was created to measure how many times bigrams and trigrams occurred in novel chunk positions (NCP) within test items. A terminal chunk was only counted as being in a novel position when it both occupied specific letter positions that it had not been seen in during training and had never occurred as the last chunk in any training string. This can be illustrated using the test string MVXVTRX. First NCP was set to 7 as 3 bigrams (XV, VT, and TR) and 4 trigrams (VXV, XVT, VTR, and TRX) appeared in novel positions. For example, the trigram VXV only appeared in letter positions 1-3 and 4-6 in training strings, but never in locations 2-4 as in this test string. Secondly the NCP value was reduced by 1 to 6 as the final trigram in the test string (TRX) had appeared as the final trigram in some training strings. The mean NCP statistics for each test string type are shown in Table 1, along with Meulemans and Van der Linden's anchor ACS and global ACS measures.

It can be seen that in Experiment 2a, the NCP score was somewhat lower in grammatical test strings than nongrammatical test strings (1.19 versus 1.76) and was markedly lower in associated than nonassociated strings (1.01 versus 1.94). A contrasting pattern is obtained in Experiment 2b with much lower NCP in grammatical than nongrammatical (0.25 versus 1.26) items and somewhat lower NCP in associated than nonassociated (0.44 versus 1.07) strings. In both experiments Meulemans and Van der Linden had carefully manipulated the measures of anchor and global ACS so that there were significant differences overall between associated versus nonassociated strings, but not between grammatical and nongrammatical strings. However, there may also be significant differences in NCP scores. Mann-Whitney tests were used to assess differences in NCP scores between different string types in the two experiments. A significance level of .05, two-tailed, is assumed for all statistical tests, unless the significance level is specifically stated. In Experiment 2a there was a significant difference in NCP scores between the associated and nonassociated test items ($U = 76.5$), but the difference between grammatical and nongrammatical items was not significant ($U = 87$). In Experiment 2b the opposite pattern was observed with a significant difference between the grammatical and nongrammatical strings ($U = 75.5$), while the difference between associated and nonassociated strings was not significant ($U = 122.5$).

These findings indicate that the NCP characteristics of the test strings show the same trends as the classification results found by Meulemans and Van der Linden. This suggests that the results of Meulemans and Van der Linden's two experiments can potentially be explained on the basis of a unitary learning system that is sensitive to both chunk frequency and location within training strings. In order to identify the best predictor of classification performance multiple regression was used to examine which string characteristics account for the highest proportions of the variance in test scores.

What Is the Best Predictor of Grammaticality Classification Rates?

Table 2 shows the correlations between the independent variables and between each independent variable and the dependent variable of classification performance. Importantly, it can be seen that in Experiment 2b grammaticality (i.e., whether a string is grammatical or nongrammatical) is significantly correlated with NCP, suggesting that an apparent effect of grammaticality observed in the ANOVA cannot be taken as evidence that participants abstracted the rules of the grammar.

Table 2

Intercorrelations Between the Predictor Variables in Meuleman's and Van der Linden's (1997) Experiments 2a and 2b.

		Classification	Grammaticality	Anchor TRS	Global TRS	Anchor ACS	Global ACS	Novelty	NCP
2a	Grammaticality	+.07							
	Anchor TRS	+.34	-.51**						
	Global TRS	-.27	-.37*	+.20					
	Anchor ACS	+.47**	-.05	+.32	-.07				
	Global ACS	+.52**	+.02	+.38*	+.00	+.57**			
	Novelty	-.50**	+.06	-.21	-.21	-.10	-.53**		
	NCP	-.04	+.19	+.02	-.06	+.05	-.19	+.61**	
	Length	+.32	+.16	+.05	+.08	+.67**	+.34	+.06	+.26
2b	Grammaticality	-.32							
	Anchor TRS	+.31	-.81**						
	Global TRS	+.35*	-.83**	+.83**					
	Anchor ACS	+.11	-.05	+.07	+.38*				
	Global ACS	+.18	-.01	+.04	+.44*	+.64**			
	Novelty	–	–	–	–	–	–		
	NCP	-.39*	+.37*	-.41*	-.56*	-.18	-.33	–	
	Length	-.39*	+.15	-.33	-.27	-.14	-.16	–	+.07

Note. Grammaticality refers to whether a string is grammatical or nongrammatical. Dashes indicate that there were no novel fragments in Experiment 2b. NCP = novel chunk position. TRS = transition rule strength. ACS = associative chunk strength.
* $p<.05$, ** $p<.01$ (two-tailed, $n = 32$ test strings).

Multiple regression was used to assess which independent variables are the best predictors of classification performance in each of Experiments 2a and 2b. In order to test within-participants predictors against the appropriate error term, the individual regression equation method recommended by Lorch and Myers (1990) was adopted. A separate simultaneous regression was run for each participant with grammaticality, anchor ACS, global ACS, novelty (Experiment 2a only), NCP and string length as predictor variables and the mean number of trials (0, 1, or 2; each string was presented twice in the test stage) on which the participant classified each string as grammatical as the dependent variable. String length was included because test strings varied from 5 to 7 letters in length and this was not controlled across the four test item types.

Table 3

Regression Coefficients from Individual Analyses of Participants' Data in Meulemans and Van der Linden's Experiment 2a.

Participant	Gram	Anchor ACS	Global ACS	Novelty	NCP	Length	Variance Explained	*P*
1	.04405	-.06941	.05454	-.04215	.04486	-.08030	9.2%	.86
2	-.04923	.26239	-.06431	-.07109	.04936	.00367	37.2%	.05*
3	-.02235	.02404	.03832	-.24686	.13560	-.02092	38.6%	.04*
4	-.12264	.10127	-.02502	-.16690	.03110	-.04370	19.4%	.45
5	-.23011	.12021	-.02214	-.10587	-.02217	.12436	41.0%	.03*
6	-.17056	.10612	-.02199	-.15014	.09668	-.01483	22.8%	.33
7	.13955	-.01602	-.03447	-.11803	-.00004	.24587	25.1%	.25
8	-.08144	-.06175	-.02922	-.22968	.14023	.07675	32.2%	.11
9	-.03732	.00448	-.02230	-.15606	.02346	.13065	13.7%	.68
10	-.03732	-.00448	-.02230	-.15606	.02346	.13065	13.7%	.68
11	.09353	.09181	.05403	.03338	.06410	-.01527	29.1%	.16
12	.07975	-.03776	.04044	-.11746	.01218	.13783	19.1%	.46
13	.12740	.00251	.00010	-.12699	.04070	-.03382	12.1%	.75
14	.22163	.10905	-.02000	-.08041	-.00589	-.24864	34.5%	.08
15	-.08543	.08053	.02877	-.10965	.04643	.03001	25.9%	.23
16	.14670	-.01580	-.01153	-.08226	.04375	-.19965	17.8%	.51
17	-.05866	.08216	.01240	-.13898	-.01444	.05916	29.7%	.15
18	-.01445	.09998	-.01105	-.02374	-.01011	.06049	11.6%	.77
19	-.23039	.11991	.03024	-.14405	.04358	-.04405	49.6%	.01**
20	-.01166	.00968	.04010	-.12100	.05645	.00856	20.2%	.42
M	-.01002	.05045	.00073	-.11770	.03996	.01534	25.1%	
SE	.02786	.01800	.00759	.01444	.00998	.02577		
t(19)	0.36	2.80*	.10	8.15**	4.00**	0.60		

Note. Gram = grammaticality, ACS = associative chunk strength, NCP = novel chunk position. * $p < .05$, ** $p < .01$.

This analysis provided one equation for each of the 20 participants in each of the two experiments. The regression coefficients for each equation from Experiments 2a and 2b are shown in Tables 3 and 4, respectively. The mean correlation coefficient was calculated across participants for each predictor variable and a one-sample *t*-test was then used to assess whether each predictor variable differed reliably from zero. The bottom three rows of each table show the results of these tests. In Experiment 2a, anchor ACS, novelty and NCP were reliable predictors of classification performance, while in Experiment 2b no predictors were significant by a two-tailed test; however, length was reliable at the $p < 0.05$ level by a one-tailed test. Grammaticality was a reliable predictor in neither experiment: the lack of a grammaticality effect in Experiment 2b critically challenges Meulemans and Van der Linden's conclusions.

Table 4
Regression Coefficients from Individual Analyses of Participants' Data in Meulemans and Van der Linden's Experiment 2b.

Participant	Gram	Anchor ACS	Global ACS	NCP	Length	Variance Explained	p
1	.14442	-.04465	.01149	.04094	-.21039	44.8%	.01**
2	.05285	.00924	.00649	.01610	.22238	14.4%	.51
3	-.07260	.01628	-.00903	-.05392	.09834	20.0%	.29
4	.25710	-.01576	-.01801	-.03500	.00769	28.8%	.10
5	-.16949	.01879	-.01474	-.11183	-.27459	33.6%	.05*
6	.09077	-.04880	.00772	-.04772	-.20284	32.6%	.06
7	-.09511	.01801	-.01350	-.19245	-.11555	36.0%	.03*
8	.06729	-.01448	-.00110	-.04970	-.24400	20.5%	.28
9	.10819	.04678	.00822	.05853	-.11989	40.7%	.01**
10	-.00121	.00921	-.02870	-.10045	-.13722	21.1%	.26
11	.00202	-.02474	.00008	-.03670	-.01844	7.5%	.83
12	.01367	-.00165	.00315	-.00002	.07252	1.3%	1.00
13	.16153	-.02713	.00453	-.05676	-.06243	24.1%	.18
14	.19640	-.04907	.01707	.04608	-.13343	28.1%	.11
15	-.02245	-.01563	.00720	-.02400	-.02287	4.3%	.94
16	.06123	-.02387	.03984	.12014	.15975	40.4%	.01**
17	.21029	.00654	-.00275	-.00006	-.19560	23.9%	.19
18	.02794	.05347	-.00602	.06852	.07054	25.7%	.15
19	-.22462	.04374	.00114	-.09754	-.24156	41.2%	.01**
20	-.11286	.02123	.00006	-.02352	.07351	11.1%	.67
M	.03477	-.00112	.00066	-.02397	-.06370	25.0%	
SE	.02858	.00687	.00322	.01613	.03266		
t(19)	1.22	0.16	0.20	1.49	1.95		

Note. Gram = grammaticality, ACS = associative chunk strength, NCP = novel chunk position. * $p < .05$, ** $p < .01$.

At first glance, the absence of a reliable effect for NCP in Experiment 2b contradicts the claim that a unitary model based on positional knowledge can largely account for the entire pattern of findings. However, further analysis of the data of Experiment 2b suggests that NCP rather than length is the critical psychological variable. The key observation is that the relationship between length and classification is not linear: Participants classified as grammatical 65.0% of strings 5 letters long, 69.7% of strings 6 letters long, and 61.8% of strings 7 letters long. Inclusion of strings of length 5 is unwarranted from the point of view of the NCP model as there were only two of them. In strings of length 6, 2/9 contained chunks in novel positions while in strings of length 7, 11/21 contained chunks in novel positions. This pattern of chunks in novel positions raises the question of whether classification rates for strings of seven letters may have been depressed by the greater number of chunks in novel positions within such strings, relative to six-letter strings.

To answer this question, a two-way ANOVA was used on the classification rates for strings of 6 and 7 letters, with length and whether strings contained chunks in novel positions or not as between-strings factors. Note that treating NCP as a binary rather than a continuous variable should make little difference as NCP had a value of 0 or 1 in all but three test strings. The analysis revealed a highly significant effect of chunks in novel positions, $F(1, 26) = 17.32$, $MSE = 11.09$, but no effect of length, $F(1, 26) = 1.96$. The interaction between length and whether strings contained novel chunks was marginally significant, $F(1, 26) = 3.09$, $MSE = 3.09$, $p = .091$. These results suggest that NCP is the critical predictor of classification performance and that length only appears to be a predictor because of idiosyncrasies in the distribution of chunks in novel positions across strings of different lengths.

Overall this analysis shows that in both experiments the string characteristics that Meulemans and Van der Linden believed were driving classification performance were confounded with other variables that have greater predictive power. In Experiment 2a, anchor ACS, novelty and NCP were reliable predictors of performance, which agrees with Meulemans and Van der Linden's conclusion that in Experiment 2a participants were responding on the basis of chunk information, though it is now evident that chunk novelty and location were stronger predictors than associative chunk strength. In Experiment 2b, using one-tailed tests, length was a reliable predictor, while grammaticality was not. These results question Meulemans and Van der Linden's conclusion that participants in Experiment 2b were classifying on the basis of grammaticality. It appears that when string length is controlled, NCP is the best predictor of performance and hence the results of Experiment 2b can also be explained on the basis of chunk information.

One final issue is that the significance levels in the final columns of the individual regressions in Tables 3 and 4 indicate that the classification performance of only four participants in Experiment 2a and six participants in Experiment 2b could be predicted reliably. The ranges of individual regression coefficients show the extent to which participants varied in the information they used to classify test strings. These individual participant analyses confirm the claim that grammaticality is not the basis of classification performance in Experiment 2b. The regression analyses for both experiments explained only 25% of the variance in performance. The low levels of variance accounted for also demonstrate that we are a very long way from a full understanding of what participants learn in artificial grammar learning experiments such as these.

What are the implications of using finite-state grammars?

Meulemans and Van der Linden explained differences in performance between their Experiments 2a and 2b on the basis that shorter training led to knowledge of letter chunks, while longer training resulted in knowledge of the rules of the grammar. However, reanalysis of their data suggests that fragment novelty and NCP were stronger predictors of performance than the anchor and global ACS explanation given by Meulemans and Van der Linden. In particular, knowledge of valid letter chunk locations provides the strongest explanation of performance in Experiment 2b.

As well as questioning Meulemans and Van der Linden's explanation of their findings, these analyses provide an opportunity to open up a methodological debate about why researchers continue to use artificial grammars based on finite-state transition rules to investigate implicit learning, as these grammars do not appear to provide a means of convincingly determining whether participants are classifying on the basis of rule knowledge or of distributional statistics.

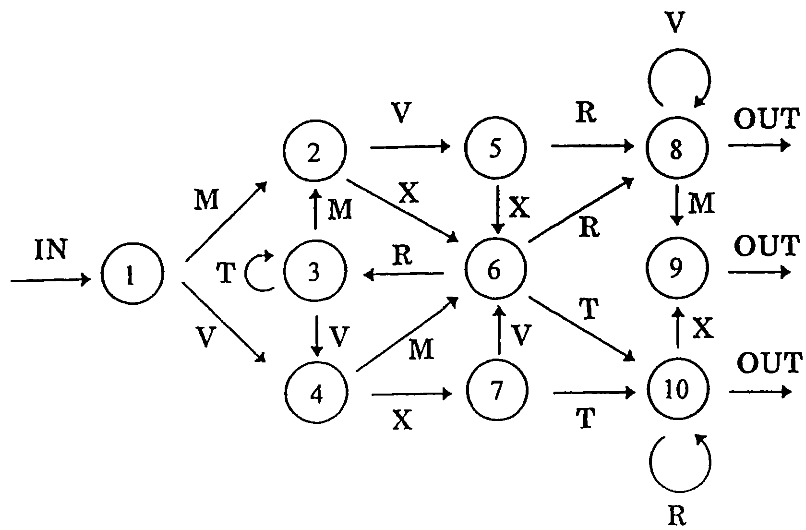

Figure 2. The artificial grammar used by Brooks and Vokey (1991).

The problem with transition-rule-based 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, see Figure 2) start with MV, MX, VM, or VX. But if participants classify test strings on the basis of knowing what letters are legal in the first two positions, it is not clear whether they are doing this on the basis of 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) or whether they are classifying on the basis of bigram knowledge (i.e., all the training strings began with MV, MX, VM, or VX). It is suggested that a partial solution to this problem is to develop new methods of quantifying grammaticality that can then be taken into account in selecting experimental materials. Chapter 3, identifies a more complete solution created by using a biconditional grammar in which rule-structure is unconfounded from the distributional statistics of n-grams.

Can grammaticality be quantified?

Over the last thirty years, there have been a number of different explanations of what participants learn in finite-state grammar experiments (see the literature review in Chapter 1) and during this time there has been a trend to control for and quantify distributional 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 are assumed to exist whenever distributional statistics do not account for all of the variance in test performance.

Although this has not been attempted before, it is possible to quantify precisely how grammatical test strings generated from finite-state grammars are. For example, the string MXRVXRM is intuitively less ungrammatical than the string XXXXX. While, the former can be turned into a grammatical string simply by moving the final M to position 4, the latter cannot be transformed so easily into a grammatical string. One method of measuring grammaticality can be illustrated using the grammar designed by Brooks and Vokey (1991) and the specific letter strings selected from this grammar by Meulemans and Van der Linden. The Brooks and

Vokey (1991) grammar, shown in Figure 2, comprises a set of 17 rules that specify which letter can be added to a string of letters at each transition between nodes of the finite-state diagram, and 6 rules that specify legal terminal nodes for each string. First, the number of times each of these 23 rules occurred in training strings was counted. Repetition of a rule in training strings was deemed to increase the 'transition rule strength' (TRS) in much the same way that the repetition of letter fragments increases associative chunk strength (anchor ACS and global ACS). These 23 measures of transition rule strength were then used to create global TRS measures, which include all of the rules, and anchor TRS measures, which include the first and last rules, in each test string. These rule measures are similar to the global and anchor ACS measures. The TRS calculations were carried out by breaking each test string down into its constituent rules, summing the training transition rule frequency counts of the rules composing a specific string, and then dividing the total rule strength by the number of letters in the string. This can be demonstrated using test string VXVRM from Experiment 2b, in conjunction with the finite-state grammar diagram in Figure 2. VXVRM comprises the following transition rules: V was selected at the transition from node 1 to 4; X was selected at the transition from node 4 to 7; V was selected at the transition from node 7 to 6; R was selected at the transition from node 6 to 8; M was selected at the transition from node 8 to 9; and the terminal transition was at node 9. These six transition rules occurred 64, 33, 17, 45, 19, and 39 times respectively in the 125 training strings. The global TRS for VXVRM was therefore (64 + 33 + 17 + 45 + 19 + 39) / 5 = 43.4. Anchor TRS was quantified by adding the transition rule strength for the first and last rules and dividing by 2. The anchor TRS for VXVRM was (64 + 39) / 2 = 51.5.

In Meulemans and Van der Linden's Experiment 2a, the mean anchor TRS values for grammatical versus nongrammatical items were 9.81 versus 8.00, and for associated versus nonassociated items were 9.56 versus 8.25. A two-way ANOVA with grammaticality (as defined by Meulemans & Van der Linden) and similarity as between-string variables showed an effect of grammaticality, $F(1, 28) = 12.29$, $MSE = 2.138$, and an effect of similarity, $F(1, 28) = 6.45$, $MSE = 2.138$, but no interaction of grammaticality with similarity, $F < 1$. The mean global TRS values for grammatical versus nongrammatical items were 13.02 versus 7.89, and for associated versus nonassociated items were 11.19 versus 9.73. A two-way ANOVA with grammaticality and similarity as between-string variables, showed an effect of grammaticality, $F(1, 28) = 73.98$, $MSE = 2.845$, and an effect of similarity, $F(1, 28) = 6.00$, $MSE = 2.845$, but no interaction of grammaticality with similarity, $F < 1$. This suggests that in Experiment 2a anchor and global TRS were confounded with anchor and global ACS.

In Meulemans and Van der Linden's Experiment 2b, the mean anchor TRS values for grammatical versus nongrammatical items were 37.88 versus 30.88, and for associated versus nonassociated items were 34.31 versus 34.44. A two-way ANOVA with grammaticality and similarity as between-string variables showed an effect of grammaticality, $F(1, 28) = 52.17$, $MSE = 7.513$, and no effect of similarity, $F < 1$, and no interaction of grammaticality with similarity, $F < 1$. The mean global TRS values for grammatical versus nongrammatical items were 47.26 versus 27.38, and for associated versus nonassociated items were 41.54 versus 33.11. A two-way ANOVA with grammaticality and similarity as between-string variables showed an effect of grammaticality, $F(1, 28) = 118.31$, $MSE = 26.734$, and an effect of similarity, $F(1, 28) = 21.26$, $MSE = 26.734$, but no interaction of grammaticality with similarity, $F(1,$

28) = 3.58. This means that in Experiment 2b, global TRS was confounded with anchor and global ACS.

Consistent with the idea that TRS is a measure of grammaticality, its correlation with the standard binary classification of grammaticality, in Experiment 2b, was significant for both anchor ($r = .81$) and global TRS ($r = .83$). Moreover, consistent with the view that participants in Experiment 2b were not abstracting the rules of the grammar, when anchor and global TRS measures were used in the full regression they failed to account for significant proportions of the variance in classification rates.

It therefore appears that despite Meulemans and Van der Linden creating the best set of training and test strings that it is possible to construct with this particular grammar, it was impossible for them to have balanced test strings according to more sensitive TRS measures of grammaticality. As the transition-based rules in Brooks and Vokey's (1991) grammar are common to all finite-state grammars, it is recommended that different types of grammars are used in future that allow grammaticality to be straightforwardly quantified and to provide a sounder basis for manipulating grammaticality measures orthogonal to chunk similarity measures.

Chapter 2 ended by suggesting that the finite-state grammars used for over 30 years to investigate the possibility that there are distinct rule-abstraction and exemplar-based learning systems are methodologically unsound as they do not allow investigators to unconfound the contributions of rule- versus exemplar-based (letter-fragment and whole training exemplar) knowledge in classification performance. In Chapters 3 to 6, a biconditional grammar is used to demonstrate that when rule- and exemplar-based knowledge are satisfactorily unconfounded there is no evidence for a distinct rule-abstraction system.

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 eight 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 other should be an F, where there is a G the other letter should be an L, and where there is a K, the other letter should be an X (e.g., DFGK.FDLX is legal, whereas LFGK.KDLX is not). This grammar has four advantages over transition-rule grammars. 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 long as an F occurs in the associated letter location. Secondly, as the rule-related positions have three intervening letters, it is possible to unconfound rule and fragment knowledge. Thirdly, all strings are eight letters long, meaning that it is not necessary to control for length. Finally, it is straightforward to quantify how grammatical test strings are. All grammatical strings contain four valid rules and in the current studies all ungrammatical test strings contain three valid rules and one illegal letter pairing.

The strings generated from this biconditional grammar allow exemplar and fragment information to be unconfounded from grammaticality more successfully than has been achieved with finite-state grammars. The first aim of the present research, therefore, is to re-evaluate the key assumptions of implicit learning that have driven AGL research over the last 30 years but which have yet to be settled: Do participants who memorise grammatical training strings, without knowing that these strings conform to a set of rules, acquire implicit rule knowledge? Or, is the performance of memorisers better explained by similarity-based exemplar or letter-fragment knowledge?

Rule Learning

In addition to studying the effects of memorising training strings, Shanks et al. (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) a clear dissociation in classification test accuracy was found, 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 exemplar knowledge, suggesting that the mental representations underlying their performance were the rules of the grammar. These results suggest that, as predicted by the episodic-processing approach (Whittlesea, 1997a, b; 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 seven experiments reported in Chapters 3 to 6 is to assess how well the findings fit this episodic-processing approach. Experiment 1 examines whether exemplar 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 exemplar (whole-item) similarity while ensuring that these two factors had minimal overlap with fragment similarity. In keeping with prior research (e.g., Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997; Servan-Schreiber & Anderson, 1990) and Chapter 2, two- and three-letter fragment similarity will be referred to as associative chunk strength (ACS). 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 surface characteristics of the training stimuli by asking them to memorise 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 relational properties of the letter strings, by asking them to hypothesis test 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. A control group was included, participants' awareness of the rules of the grammar was assessed, and the rule letter pairs were counterbalanced across participants. While the match and edit groups trained on the same grammatical training items, a control group was asked to memorise 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. A questionnaire was used fairly exhaustively to assess participants' knowledge of the rules of the grammar. All 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 three letter pairs that can be created from the letters D, F, G, K, L, and X.

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, it was predicted that the match group would perform at the chance level anticipated in the control group. Secondly, as the edit participants had actively sought to identify the rule structure, it was predicted that those who succeeded in identifying the rules would show an effect of grammaticality. Thirdly, 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. 24 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 randomised 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 64 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 memorise 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, and (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 or not. 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 string. 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 upon 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. The questionnaire is shown in Appendix G.

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 (Lists 1 and 2) were counterbalanced for participants 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 given in Chapters 3 to 6 and the appendices 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 one or two rule violations, making them one or two letters different from the grammatical string. For each training string used for the control group, two versions were also created that differed from the original string by one or two letters.

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). There were no relationships of grammaticality or similarity between the test strings and the training strings processed by the control group.

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 List 1 and List 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) / 6) / 2 = 2.02 to the 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 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

Table 5 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, MSE = 44.46, indicating that participants' ability to memorise the training strings improved as training progressed. A two-way ANOVA for the match group, with block as a within-subjects variable and list as a between subjects variable, indicated that there were no effects of block, F (3,18) = 2.46, MSE = 52.73, or list, $F < 1$, and there was no Block x List interaction, $F < 1$. The control and match groups' performance was close to ceiling and participants were performing the memorisation task accurately across training blocks.

Table 5

Mean Percentage of Correct Responses by Training Blocks in Experiments 1 and 2

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

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 5). A two-way ANOVA with block as a within-subjects variable and list as a between-subjects variable yielded an effect of block, $F\,(3, 18) = 6.91$, $MSE = 66.37$, but no effect of list, $F < 1$, and no Block x 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 6 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 6 indicate that this is not significantly different from the percentage correct in the control group, $t < 1$, $SE = 4.69$. The mean percentage correct for the edit group

Table 6

Mean Percentages of Items Classified as Grammatical, and Signal Detection Results in Experiments 1 and 2.

Group	N	Mean Percentage "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	53± 5.1	0.18±0.27	0.01	68±8.1*	1.04±0.63*	0.05
Edit	8	77	63	29	12	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

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

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, $F(1, 21) = 10.43$, $MSE = 1015.27$, and a Group x Grammaticality interaction, $F(2,21) = 5.16$, $MSE = 1015.27$. The main effects of group and similarity, and the Group x Similarity, Grammaticality x Similarity, and Group x Grammaticality x 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 x Similarity interaction, $F(1, 7) = 2.03$, $MSE = 26.70$. In the match group there was no effect of grammaticality, $F(1, 7) = 1.39$, $MSE = 647.01$, or similarity, $F < 1$, nor a Grammaticality x Similarity interaction, $F < 1$. The edit group showed an effect of grammaticality, $F(1, 7) = 8.60$, $MSE = 2342.61$, 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 6 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. In contrast, discrimination in the edit group was well above chance. Participants in all three groups showed a slight bias toward calling strings ungrammatical.

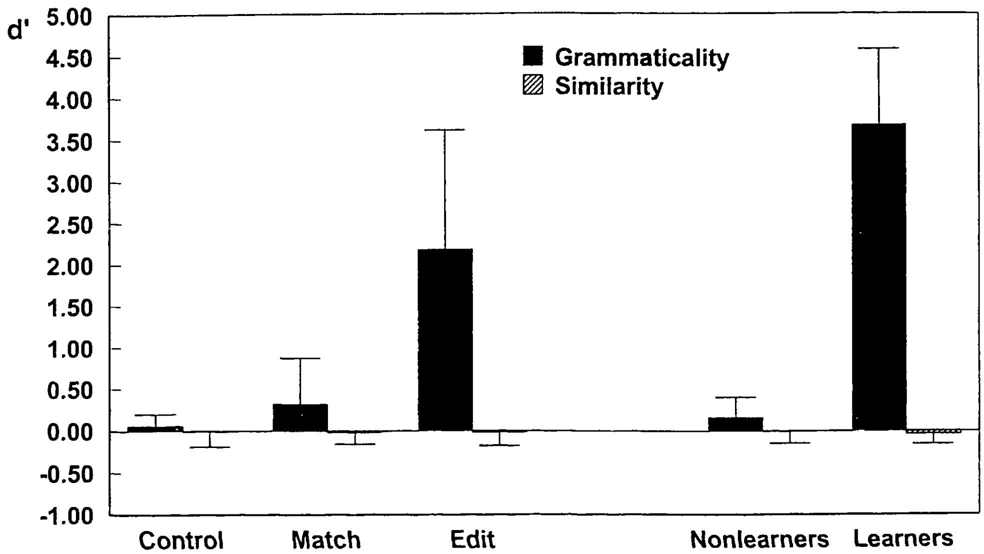

Figure 3. Mean d' for grammaticality-based (d'_g) and similarity-based (d'_s) classification in the control, match, edit, nonlearner, and learner groups of Experiment 1. Error bars represent 95% confidence intervals.

Table 6 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. 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 similarity sensitivity scores (d'_s) for all three groups encompass chance-level sensitivity. For these calculations, grammatical responses for similar (grammatical and ungrammatical) test items are scored as hits whereas grammatical responses for dissimilar (grammatical and ungrammatical) items are scored as false alarms.

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 6). 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 towards calling test strings ungrammatical, whereas the learners had a slight bias towards 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 6), 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 score 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, $F(1, 10) = 2.79$, $MSE = 36.20$, or similarity, $F < 1$, nor a Grammaticality x Similarity interaction, $F < 1$. For the learners, on the other hand, there was a significant grammaticality effect, $F(1, 4) = 328.52$, $MSE = 124.81$, with no effect of similarity, $F < 1$, and no Grammaticality x Similarity interaction, $F(1, 4) = 1.54$, $MSE = 16.02$.

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 99% 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 training 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 largely 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 a whole test items with training items, in Experiment 2 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.

The episodic-processing account (Whittlesea & Dorken, 1993; Wright & Whittlesea, 1998) 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. Secondly, 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. The chunk knowledge acquired during training would create familiarity effects for novel high ACS test strings as there would be a "discrepancy" (Whittlesea & Williams, in press) between the impression that all test strings are novel and the unexpected fluency of processing high ACS test strings. As participants would be unaware of the effect of prior chunk processing on their fluency of processing high ACS test strings, they would attribute increased fluency to grammaticality. Since processing in the training stage did not include explicit analysis of the rule structure, these participants would classify at chance levels in relation to the rules of the grammar and would not be able to say what the rules of the grammar are.

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 of the test strings 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 5 shows the mean percentage of training items on which participants 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, $F(3,21) = 2.13$, $MSE = 23.38$. Performance was close to ceiling and the results show that participants were performing the memorisation task accurately across training blocks. A one-way ANOVA for the edit group yielded a significant effect of block, $F(3, 21) = 3.51$, $MSE = 103.70$. This suggests, as predicted, that the edit group acquired new knowledge as training progressed with successful hypothesis testing.

Table 6 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 53% with confidence interval of 48 to 58% suggesting that these results could have occurred by chance. The mean percentage correct for the edit group was 75% with confidence intervals of 58 to 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, $F(1,14) = 9.22$, $MSE = 1384.97$, and ACS, $F(1,14) = 17.90$, $MSE = 564.35$, and a significant Group x Grammaticality interaction, $F(1,14) = 5.20$, $MSE = 1384.97$. The effects of group, $F(1,14) = 1.42$, $MSE = 208.29$, and Group x ACS, $F(1,14) = 2.85$, $MSE = 564.35$, Grammaticality x ACS, $F < 1$, and Group x Grammaticality x ACS, $F(1,14) = 1.05$, $MSE = 161.79$ 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, $F(1,7) = 17.94$, $MSE = 551.14$, but no effect of grammaticality, $F(1, 7) = 1.85$, $MSE = 213.22$, and no Grammaticality x ACS interaction, $F(1,7) = 1.24$, $MSE = 193.37$. The edit group showed an effect of grammaticality, $F(1, 7) = 7.66$, $MSE = 2556.73$, whereas the effect of ACS, $F(1,7) = 3.16$, $MSE = 577.57$ and the Grammaticality x ACS interaction, $F < 1$, were not significant.

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

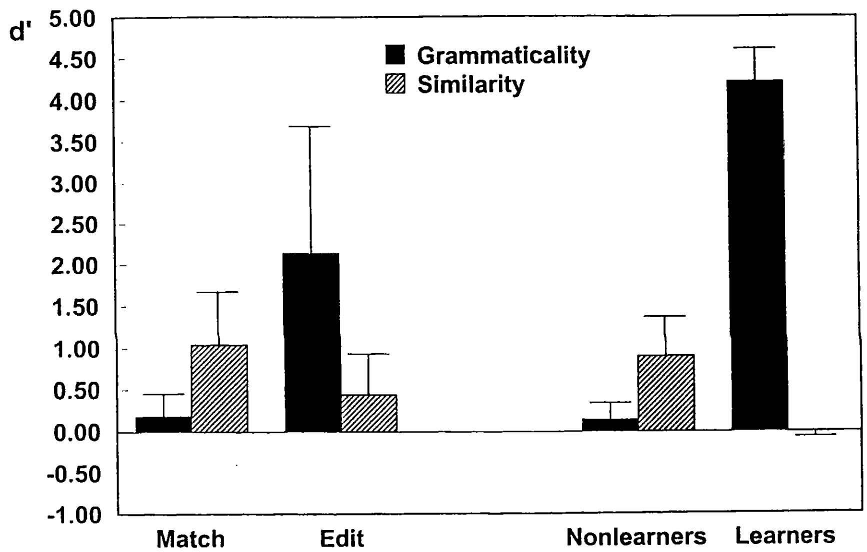

Figure 4. Mean *d'* for grammaticality-based (d'_g) and similarity-based (d'_s) classification in the match, edit, nonlearner, and learner groups of Experiment 2. Error bars represent 95% confidence intervals.

Percent 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 6 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 with confidence intervals of 0.41 to 1.67, while the edit group performed at chance with confidence intervals of -0.05 to 0.93.

Verbal answers to the questionnaire 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 6. 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, $F(1,11) = 25.04$, $MSE = 543.59$, but no effect of grammaticality, $F(1, 11) = 2.23$, $MSE = 165.98$, nor a Grammaticality x ACS interaction, $F < 1$. For the learners, by contrast, there was a significant grammaticality effect, $F(1, 3) = 2933.57$, $MSE = 12.66$, with no effect of ACS, $F < 1$, and no Grammaticality x ACS interaction, $F(1, 3) = 1$, $MSE = 1.09$.

Finally, signal detection measures were calculated for sensitivities to the grammaticality and the 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 analysed 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 the 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. Verbal answers to the questionnaire indicated that edit learners who used rule knowledge to classify test items at above chance levels of accuracy also had explicit knowledge of the rules. In contrast, participants who classified on the basis of fragment knowledge (i.e., the match group and edit nonlearners) performed at chance in relation to the rules of the grammar, could not say what the rules of the grammar are, and reported that they were guessing in the classification test. There was therefore an association between classifying on the basis of the rules of the grammar and being able to verbally report the rules of the grammar.

Chapter 4: Exemplar versus Fragment Knowledge

This chapter focuses on whether participants who mentally rehearse training strings acquire knowledge of whole training exemplars or letter fragments. In Experiment 1 (Chapter 3) and Shanks, Johnstone and Staggs (1997, Experiment 4), 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 groups, in both experiments, showed no effect of whole-item similarity in their classification test performance. The first aim of the experiments reported in this chapter was to provide stronger tests of exemplar theories (Brooks, 1978; Medin & Schafer, 1978; Nosofsky, 1986) by increasing the number of similar training items for each similar test item.

The second aim was to continue to test the episodic-processing prediction that memory preserves records of the processing applied to training items in order to meet the demands of the training task and later uses these records to drive test performance (Whittlesea, 1997b). In Experiment 2, the demands of the match task were to memorise a target string and then to select it from a list of three strings, where one distracter string differed from the target by one letter and the other differed by two letters. Although the two distracters were highly similar to the target item, participants successfully selected the target string on 94% of trials. This high level of accuracy in the match task combined with ACS effects in classification performance suggests that participants had memorised training items as a series of two- and three-letter fragments. Moreover, these findings suggest that as fragment knowledge was sufficient to meet the demands of the training instructions, fragment knowledge will continue to drive the test performance of match participants in experiments with stronger exemplar manipulations.

Experiment 3

This experiment provided a stronger test of the exemplar account as participants memorised six similar training strings for 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 orthogonal to whole training-item similarity. Unlike in the previous experiments, training strings were not constructed according to the biconditional rules.

Participants were asked to memorise the 108 letter strings shown in Appendix C. 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 and 2 and asked to classify test strings as grammatical or ungrammatical. The classification test stimuli comprised 18 high-whole-item similarity/high ACS strings (HIHA) 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 test strings (LIHA) 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 strings (LILA) 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 episodic-processing account and the strong ACS effects found in Experiment 2, it was predicted that participants would continue to classify more high ACS letter strings (LIHA) than low ACS strings (LILA) as grammatical, keeping whole-item similarity constant. In contrast to a strong ACS effect, it was predicted that participants may not acquire knowledge of whole training exemplars, as letter chunking was often sufficient to meet the demands of the match task. Thus participants would be equally likely to classify HIHA and LIHA items as grammatical. The three sets of test strings were designed to allow the contributions of whole-item similarity and ACS in classification performance to be unconfounded. 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. 12 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 Experiment 1 (Chapter 3).

Materials. 324 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 C). Unlike Experiments 1 and 2, double letters (e.g., DD) were used in this experiment. 108 of the training strings were each presented for mental rehearsal in the match task. The remaining 216 strings were the distracters in the list of three strings presented at the end of each match trial. Half of the distracter 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 C). Eighteen high-item/high-ACS strings (HIHA) 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 strings (LIHA) were dissimilar to all training strings, but overlapped with all training items in terms of ACS (low-item/ high-ACS test strings). A third set of low-item/ low-ACS strings (LILA) was dissimilar to all training items and dissimilar to all training string fragments. Appendix C 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 7 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 effect of block, $F\ (3, 33) = 2.38$, $MSE = 32.63$.

Table 7

Mean Percentage Correct Training Responses for Experiments 3 and 4

Experiment	Block 1	Block 2	Block 3	Block 4	Overall Accuracy
3	90	90	87	93	90
4	94	93	95	94	94

Table 8 presents the mean percentage of items classified as grammatical for each of the three test item types (HIHA, LIHA, and LILA). If participants classified on the basis of knowledge of whole training items then they would have called HIHA strings grammatical and LIHA strings ungrammatical; the results for these items were therefore reanalysed taking "grammatical" responses to HIHA and "ungrammatical" responses to LIHA items as "correct". The results of this analysis indicated that participants did not memorise whole items as they performed at chance levels with a mean percentage correct of 50% with confidence intervals of 47.5% to 52.5%. 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. By this analysis, participants classified 71% of strings correctly with confidence intervals of 63% to 79%.

Related *t* tests indicated that there was no difference in the percentages of HIHA and LIHA items classified as grammatical, $t < 1$, but that there was a reliable difference in the percentages of LIHA and LILA items classified as grammatical, $t\ (11) = 5.13$. Further, the mean d'_{item} score of 0.02 revealed chance sensitivity to

Table 8

Mean Percentages of Items Classified as Grammatical, and Signal Detection Result, in Experiments 3 and 4

		Mean Percentage "Grammatical" Responses			Whole-Item Similarity Measures			Associative Chunk Strength (ACS) Measures		
Experiment	N	HIHA	LIHA	LILA	Correct and CI	Sensitivity (d'_{item}) And CI	Bias C	Correct and CI	Sensitivity (d'_{acs}) and CI	Bias C
3	12	73%	73%	31%	50%±2.5	0.02±0.18	-0.68	71%±8.0*	1.30±0.63*	-0.02
4	12	81%	74%	9%	54%±3.7*	0.40±0.35*	-1.01	82%±8.1*	2.28±0.69*	+0.34

Note. HIHA = high item / high ACS similarity, LIHA = low item / high ACS similarity, and LILA = low item / low ACS similarity. CI = 95% confidence interval. * indicates $p < .05$.

whole-item similarity as the confidence intervals were -0.16 to 0.20. In contrast, the d'_{acs} score of 1.30 indicated above chance sensitivity to ACS as the confidence intervals were 0.67 to 1.93. Despite the fact that there were six similar training items for each similar test item, participants classified solely on the basis of fragment knowledge.

Experiment 4

The aim of Experiment 4 was to provide an even stronger test of the exemplar account by increasing the number of training items that overlapped with each similar test item from 6 to 24. Again it was predicted that participants would classify more high ACS letter strings (LIHA) as grammatical than low ACS strings (LILA) and would be equally likely to classify HIHA and LIHA strings as grammatical.

Method

Participants. 12 students at University College London performed the experiment and were paid £5 for taking part.

Procedure. The same match and classification tasks were used as in Experiment 1 (Chapter 3).

Materials. 432 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). 144 of the training strings were each presented for mental rehearsal in the match task. The remaining 288 strings were the distracters in the list of three presented at the end of each match trial. Half of the distracter strings differed from the initial 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). Six high-item/high-ACS strings (HIHA) each overlapped with 24 training strings on six or seven letters and also overlapped with all training strings in terms of ACS. Six low-item/high-ACS strings (LIHA) were dissimilar to all training items, but overlapped with training strings in terms of ACS (low-item/high-ACS test strings). A third set of low-item/low-ACS strings (LILA) was dissimilar to all training items and dissimilar to all training string fragments. 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 and Discussion

Table 7 shows the mean percentage of training items on which participants made correct responses. A one-way ANOVA, with block (36 training trials) as a within-subjects variable, found no overall effect of block, $F\,(3, 33) = 1.09$, $MSE =$ 12.76.

Table 8 presents the mean percentage of items classified as grammatical for each of the three test item types (HIHA, LIHA and LILA), together with the mean percentage correct if participants were classifying on the basis of whole-item knowledge or ACS. In contrast to Experiment 3, the mean percentage correct for whole-item accuracy was 54% indicating that participants may have memorised whole training items, as the confidence intervals were 50% to 58%. This conclusion is reinforced by a d'_{item} sensitivity score of 0.40 with confidence intervals of 0.05 to 0.75. However, there was a much stronger influence of ACS on classification performance as the mean percentage correct was 82% with confidence intervals of 74% to 90%, and the mean d'_{acs} sensitivity score was 2.28 with confidence intervals of 1.59 to 2.97.

A one-way ANOVA, with whole-item similarity as a within-subjects variable, found a marginal difference between the percentages of HIHA and LIHA items classified as grammatical, $F(1, 11) = 4.35, p = .06, MSE = 85.46$, whereas the same ANOVA with ACS found a reliable difference between the percentages of LIHA and LILA items classified as grammatical, $F(1, 11) = 62.24, MSE = 404.98$. The sums of squares generated by the two ANOVAs indicated that the whole-item manipulation accounted for 3% of the variance in performance for HIHA and LIHA items, whereas ACS accounted for 80% of the variance in performance between LIHA and LILA items. In addition, the d'_{acs} scores are significantly greater than the d'_{item} scores, $t(11) = 4.92, SEM = .38$. Overall, these results suggest that ACS knowledge was the major determinant of classification performance.

As the ACS effect was so much stronger than the whole-item similarity effect it is worth considering whether participants had noticed the high repetition of four-letter chunks and were, in fact, learning longer letter chunks than the two- and three-letter fragments measured by ACS. Relative to Experiment 3, the design of this experiment increased the number of training trials (108 to 144), anchor ACS (4.77 to 8.25) and global ACS (28.31 to 40.44). It is therefore possible that the whole-item similarity effect is confounded with knowledge of fragments of four or more letters.

The difference between HIHA and LIHA test items was reassessed using a new fourgram ACS measure, which combined bigram, trigram and fourgram strength. Using this new measure, there was no difference between HIHA and LIHA test items for mean anchor fourgram ACS (7.5 versus 6.45), t (10) = 1.29, SE = 0.819, but there was a difference in mean global fourgram ACS (32.14 versus 28.77), t (10) = 3.49, SE = 0.965. Thus, because there were more training trials in this experiment than in Experiment 3, it is possible that participants may have learned about fourgrams and therefore that fourgram ACS can fully explain the data.

Exemplar Knowledge

The results of Experiments 1 and 3 undermine exemplar accounts which suggest that participants acquire structural knowledge of a collection of whole training items in a stimulus-driven manner (Brooks, 1978; Brooks & Vokey, 1991; McAndrews & Moscovitch, 1985; Vokey & Brooks, 1992, 1994) and instance models (e.g., Hintzman, 1986, 1988; Medin & Schafer, 1978; Nosofsky, 1986, 1988) which suggest that classification responses are influenced by similarity to prior examples. There was no evidence that participants memorised a collection of training exemplars with one (Experiment 1, Knowlton & Squire, 1994; Shanks et al., 1997) or six (Experiment 3) similar training items. Although there was marginal support for exemplar knowledge with 24 similar training items (Experiment 4), there was also the possibility that participants used four-gram knowledge to classify test items.

The results of Experiments 1, 3, and 4, combined with the predictions of the episodic-processing account also undermine Whittlesea's (1987) claim that exemplar knowledge drove performance in a speeded identification task. Whittlesea trained participants on letter strings (e.g., set IIa - FEKIG, FUTEG, PURYG, FYRIP, and KURIT) that each differed from a prototype (e.g., FURIG) by two letters by showing these items for unlimited time and asking participants to write them down. At test, each item was presented for 30 ms followed by a pattern mask and participants were asked to write down the letters they could read in their correct positions. Whittlesea concluded that participants had memorised the training exemplars rather than the prototype as they were more accurate in writing down the letters of old training items than three sets of novel items that also differed from the prototype by two letters (e.g., set IIb - FYKIG, FUTYG, PUREG, FERIP, PURIT; set IIc - FUKIP, PUTIG, FURYT, FYREG, KERIG; and set III - PEKIG, FYTEG, PURYT, FYKIP, KURET).

Unfortunately, these results do not provide sound evidence of exemplar learning as the pattern of results can also be explained by participants being less accurate in writing down novel fragments, than in writing down novel exemplars. While the old test items (set IIa) contained only familiar letter-pairs, the other three sets contained four (set IIb - FYKIG, FUTYG, PUREG, FERIP, PURIT), six (set IIc - FUKIP, PUTIG, FURYT, FYREG, KERIG) and six (set III - PEKIG, FYTEG, PURYT, FYKIP, KURET) novel fragments respectively.

Processing-driven Fragment Knowledge

In contrast, and as predicted, participants demonstrated a strong processing-driven effect of fragment knowledge in Experiments 3 and 4. These findings support Whittlesea's (1997a, b; Whittlesea & Dorken, 1993, 1997; Whittlesea & Wright, 1997) episodic-processing account which suggests that training leads to the acquisition of a set of processing skills that are sufficient to satisfy the demands of the task and which are reused whenever the current context, stimuli and task cue the original processing episodes. On each training trial, match participants were instructed to memorise a letter string in preparation for a short-term memory test, but they were not told how to memorise that string. It can be inferred from the results that participants actively adapted to the training task by applying rehearsal processes and organising each "to-be-remembered" string into a series of letter fragments. Furthermore, it seems likely that they maintained this approach as it was sufficient to meet the demands of the task which was to select each to-be-remembered string from the list of three strings presented in the second part of each match trial.

When participants were subsequently presented with more eight-letter strings at test, the format of the strings cued memory for the operations (mental rehearsal) and training string properties (letter fragments) used to process training items. Participants then used their mental rehearsal processing and fragment knowledge to classify the novel test items. Test items containing letter fragments processed during training would be processed more fluently than those containing novel fragments, as familiar fragments would cue prior training episodes whereas novel fragments would not. More fluent processing would be attributed to a test string being grammatical, whereas less fluent processing would be attributed to a test string being ungrammatical. As a result, more LIHA than LILA test strings were classified as grammatical.

While the form of knowledge (letter fragments) acquired by match participants in Experiments 2, 3, and 4 supports the findings of prior AGL studies (e.g., Dienes, Broadbent, & Berry, 1991; Dulany, Carlson & Dewey, 1984; Knowlton & Squire, 1994, 1996; Perruchet, 1994; Perruchet & Pacteau, 1990; 1991; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990), it should not be assumed that this fragment knowledge was acquired in a stimulus-driven manner. In particular, the results of Experiments 2, 3, and 4 challenge an assumption of Servan-Schreiber and Anderson's (1990) competitive-chunking model that with sufficient training memorisation leads to encoding whole training items in the form of hierarchical networks of letter fragments. For example, this model predicts that repeated encounters with the string GFLK.XDGF will initially result in mental representations of single letters (D, F, G, K, L, and X), progressing to bigrams (GF, LK, XD, and GF), followed by fourgrams (GFLK and XDGF), eventually leading to a representation of the whole string. Thus with fewer training trials participants are expected to encode only letter fragments, but with extended training they are expected to acquire exemplar knowledge.

Counter to the assumptions of the competitive-chunking model, there was no evidence of whole-item similarity effects with 46 training trials (Knowlton & Squire, 1994, Experiment 2b), 72 training trials (Shanks, Johnstone, & Staggs, 1997 and Experiment 1), or 108 training trials (Experiment 3). Though there was marginal support for whole-items effects with 144 training trials (Experiment 4), this effect could also be due to four-gram fragment knowledge. However, evidence of four-gram knowledge supports the competitive chunking model as it suggests that participants were acquiring a hierarchy of successively longer letter chunks which with further training would have eventually led to exemplar knowledge.

Conclusion

In conclusion, there is no evidence that participants can memorise a collection of training exemplars. Where ACS is balanced across similar and dissimilar test items (Experiments 1 and 3; Knowlton & Squire, 1994; Shanks, Johnstone & Staggs, 1997) there is no evidence for whole-item similarity. In Experiment 4, where fourgram ACS was confounded with whole-item similarity, ACS was the stronger predictor of overall performance. The results of Experiments 1, 3, and 4 suggest that exemplar effects reported by Vokey and Brooks (1992) and Whittlesea (1987) were confounded with letter-fragment knowledge.

As the chance rule-based classification performance of the match groups in Experiments 1 and 2 (Chapter 3) and Shanks, Johnstone, and Staggs (1997, Experiment 4) runs counter to a substantial amount of evidence that memorisation leads to implicit, rule knowledge (e.g., Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997; Reber, 1967, 1989; Reber & Allen, 1978; Reber & Lewis, 1977), and is of such crucial significance for the theoretical understanding of implicit learning, the aims of the experiments reported in this chapter were to examine in more detail the form of knowledge acquired by match participants and to use sensitive tests of awareness of the knowledge used to classify test items.

In contrast to the lack of support for implicit learning theories, the fragment effects in the classification performance of memorisers in Experiments 2 to 4 provided strong support for the episodic-processing account (Whittlesea, 1997a, b) as it appears that participants met the demands of the match task by "chunking" training strings into two- and three-letter fragments and this fragment knowledge was sufficient to explain classification test performance. It is suggested that biconditional grammar experiments are more likely than finite-state grammar studies to add to our understanding of the knowledge acquired by incidental learning instructions as rule and fragment knowledge can be tested orthogonally with biconditional but not finite-state grammar generated test strings.

At study by Reber and Allen (1978) illustrates the difficulties of determining what information participants use to classify test items generated from a finite-state grammar. After observing grammatical training strings, participants correctly classified 74% of novel strings as grammatical or ungrammatical, while concurrently justifying their classification decisions. Although participants made explicit justifications such as "you cannot start with an X", "TX cannot occur like that", and "MXR looks wrong there", Reber and Allen concluded that the major predictor of classification accuracy was implicit rule knowledge as participants were only able to justify 821 out of 2000 decisions. Unfortunately free report (i.e., cued recall) was unlikely to extract all that participants knew and the relative contributions of rule and fragment knowledge in classification performance were not quantified.

Using Reber and Allen's (1978) training strings and a more sensitive recognition test, Perruchet and Pacteau (1990) demonstrated that memorisers acquired explicit bigram knowledge as only 3 out of 25 old bigrams were judged less familiar than novel bigrams and there was a significant correlation between recognition scores and the frequency of occurrence of bigrams in training strings. Dienes, Broadbent, and Berry (1991, Experiment 1) also used Reber and Allen's (1978) grammar and extended Perruchet and Pacteau's bigram recognition findings.

In this study, participants classified test items, then verbally reported the rules or strategies used to classify test items, and finally carried out a "sequential letter dependency" (SLD) task. In the SLD task, they specified which letters could follow stems varying in length from zero to five letters. Although 65% of test strings were classified correctly, the verbally reported rules only accounted for 55% accuracy. In contrast, the results of the more sensitive SLD test fully accounted for classification accuracy. Thus, Perruchet and Pacteau (1990) and Dienes et al. (1991) were able to demonstrate that explicit fragment knowledge was also a predictor of classification performance.

Experiment 5

The aim of Experiment 5 was to quantify explicit rule 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, exemplars, or letter fragments. Both groups classified the test strings used in Experiment 2. As one participant in Experiment 1 (Chapter 3) worked out what the rules of the grammar were during the classification test, it was possible that other participants might do the same in this experiment. A cued-recall test was used to divide match participants into those who were aware and unaware of the rules of the grammar and a subjective confidence test was used to provide a further measure of explicit rule knowledge that should correlate with accuracy in the cued-recall test and classification performance.

It was predicted that unaware match participants would replicate the chance rule-based classification performance of the match group in Experiment 2. In contrast, aware participants were expected to show significantly more rule knowledge in classification performance than either the control or unaware match groups. As all match participants had processed the training items in the same left-to-right manner required to memorise letter strings, both the unaware and aware match groups were expected to show significant levels of fragment knowledge in classification performance. Control participants were expected to show chance levels of fragment knowledge as their training strings contained all the bigrams used to create both high and low ACS test strings.

The prediction that aware match participants would show ACS effects in classification performance differed from that of Experiment 2 where the edit learners were expected to show no effect of fragment knowledge. It was assumed that the edit learners in Experiment 2 did not show sensitivity to fragment knowledge as they had processed training strings by scanning backwards and forwards across the central dot to check whether the letters were in accordance with the rules of the grammar. The issue of whether edit participants acquire fragment knowledge is addressed in Experiment 6.

Method

Participants. 99 students at University College London performed the experiment as part of their first 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 1-4. 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.

Match Group Training Strings. The letter strings used in Experiment 2 were modified to remove a whole-item similarity effect, 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 E). 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 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 letter fragments (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 6 letters D, F, G, K, L, and X, without using double letters (e.g., DD), were used to construct the strings (see Appendix E)

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

Questionnaire. After the classification test, participants were asked to complete a two-part questionnaire. The first part informed them that the training strings had been constructed according to rules that governed which pairs of letters could occur in positions 1 and 5, 2 and 6, 3 and 7, and 4 and 8 of letter strings. Six questions attempted to elicit rule knowledge by specifying one letter and asking what letter should be paired with it. 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". The questionnaire is shown in Appendix G.

Results

The mean percentages of training items on which participants in the control and match groups made correct responses are shown in Table 9. 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, $F\,(3, 291) = 11.61$, $MSE = 56.50$, but no effect of group, $F\,(1, 97) = 2.73$, $MSE = 257.77$, and no Group x Block interaction, $F < 1$. The performance of both groups improved across training blocks.

Table 9

Mean Percentage of Correct Responses across Blocks in the Training Phase

Experiment	Group	Block 1	Block 2	Block 3	Block 4	Overall Accuracy
5	Match	86	89	91	92	90
	Control	84	84	89	90	87
6	Match	91	94	91	94	93
	Apply Rules	98	99	99	98	98

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 (out of 15 different sets) they experienced. The answers of all control participants 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 an aware group (N = 16) if they answered four or more questions correctly. This cut-off point was selected as it created mean correct questionnaire scores for the unaware match group ($M = 1.09$, $SEM = 0.13$) that did not differ from the control group ($M = 0.97$, $SEM = 0.16$), $t < 1$, while the aware match group, ($M = 5.06$, $SEM = 0.23$) was reliably more accurate than both the unaware match group, $t\,(68) = 14.72$, and the control group, $t\,(43) = 14.76$.

Table 10 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 performing at just 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, $F\,(2, 96) = 10.74$, $MSE = 829.20$, and ACS, $F\,(1, 96) = 32.57$, $MSE = 415.43$, and significant Group x ACS, $F\,(2, 96) = 13.22$, $MSE = 415.43$, Grammaticality x ACS, $F\,(1, 96) = 5.94$, $MSE = 90.63$, and Group x Grammaticality x ACS, $F\,(2, 96) = 4.95$, $MSE = 90.63$, interactions. There was no effect of grammaticality, $F < 1$, and no Group x Grammaticality interaction, $F\,(2, 96) = 2.59$, $MSE = 111.61$. A two-way ANOVA for the control group, with both

Table 10

Mean Percentages of Items Classified as Grammatical, and Signal Detection Results, in Experiments 5 and 6

Group	N	Mean Percentage "Grammatical" Responses: GH	GL	UH	UL	Grammatical Measures: Correct and CI	Sensitivity (d'_g) and CI	Bias c	ACS Measures: Correct and CI	Sensitivity (d'_{acs}) and CI	Bias c
Experiment 5											
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
Experiment 6											
Match	12	66	44	62	47	50± 4.1	0.00±0.24	0.14	59±7.4*	0.59±0.51*	-0.18
Apply Rules	11	97	94	3	2	97± 1.3	3.86±0.27	0.06	51±1.0	0.05±0.04*	0.02

Note. GH = grammatical and high ACS, GL = grammatical and low ACS, UH = ungrammatical and high ACS, and UL = ungrammatical and low ACS. CI = 95% confidence interval. * indicates $p < .05$.

grammaticality and ACS as within-subjects variables, indicated that there was no effect of grammaticality or ACS, and no Grammaticality x 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, $F(1, 53) = 55.19$, $MSE = 464.94$, but no effect of grammaticality, $F < 1$, and no Grammaticality x ACS interaction, $F(1, 53) = 1.50$, $MSE = 89.91$. Finally, for the aware match group there was a significant effect of ACS, $F(1, 15) = 9.52$, $MSE = 641.13$, a marginal effect of grammaticality, $F(1, 15) = 4.14$, $MSE = 115.67$, $p = .06$, and a Grammaticality x ACS interaction, $F(1, 15) = 9.45$, $MSE = 103.30$. The latter interaction derives from the fact that the aware match participants were less likely to call ungrammatical / similar test strings grammatical (M = 45%, $SEM = 4.84$) than the grammatical / similar strings ($M = 59\%$, $SEM = 4.07$), $t(15) = 3.58$.

Signal detection measures were calculated to assess sensitivity in judging 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 towards calling all test strings grammatical while the match participants showed a bias towards calling strings ungrammatical. The grammaticality effect in the aware match group's classification performance appeared to be based on explicit knowledge as there were positive one-tailed correlations between the percentage of items classified correctly and both the number of explicit rules identified by the questionnaire ($r = .46$), and the subjective confidence in the accuracy of the cued-recall rule questions ($r = .44$).

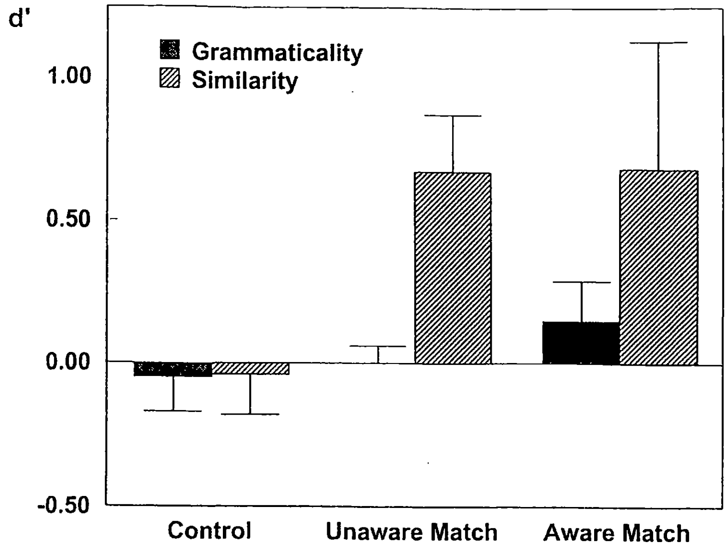

Figure 5: Mean d' for grammaticality-based (d'_g) and similarity-based (d'_s) classification in the control, unaware match and aware match groups in Experiment 5. Error bars represent 95% confidence intervals.

The accuracy of classification performance on the basis of sensitivity to ACS is also shown in Table 10 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 Grammaticality x 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 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 48 test trials indicated that there was no difference in classification accuracy between test block 1 ($M = 53\%$ correct trials, $SEM = 1.85$) and block 2 ($M = 52\%$, $SEM = 1.55$), $t < 1$. This suggests that the aware match group acquired their explicit rule knowledge during training.

Discussion

The classification test results indicated that both aware and unaware match participants used ACS knowledge, but only the aware participants used rule knowledge. The answers to the cued-recall rule questions and the subjective confidence ratings confirmed that the aware match group's rule knowledge was explicit as 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 classifications. Thus, their rule knowledge was explicit by both cued recall and global subjective confidence measures. Experiment 7 (Chapter 6) looks at the relationship between accuracy and subjective confidence on a more detailed trial-by-trial basis.

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

Thirdly, 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 (1997) 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 memorise 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 fragment knowledge in their classification performance, while the aware match group in Experiment 5 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 this experiment will have processed the training strings in the sequential left-to-right manner necessary to meet the demands of the memorization training task leading to knowledge of the distributional statistics of letter fragments (i.e., ACS).

Experiment 6

The results of Experiments 2-5 clearly demonstrate that participants who failed to learn the rules of the grammar (i.e., unaware match and edit nonlearners) used letter fragment knowledge to classify test items. One aim of Experiment 6 was to use the biconditional grammar to examine whether bigram knowledge is explicit. 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. 18 of the 30 bigrams were old, as they had been seen during training, and 12 bigrams were novel. On each trial, participants indicated whether they believed each bigram was old or new. On the basis of prior finite-state grammar research (e.g., Dienes, Broadbent & Berry, 1991; Perruchet & Pacteau, 1990), it was predicted that match participants would have explicit bigram knowledge.

The second aim of Experiment 6 was to test predictions of the episodic-processing account (Whittlesea, 1997a, b) by comparing the bigram recognition test performance of match and apply rules groups. As the match group must mentally rehearse training strings in a sequential left-to-right manner in order to meet the demands of the match task, the episodic-processing account predicts that this group will acquire bigram knowledge and hence be able to discriminate between old training bigrams and novel bigrams in a recognition test.

The episodic-processing account predicts that the match group's recognition accuracy will differ from the chance performance of the apply rules group. As the apply rules participants are told the rules of the grammar at the beginning of the training phase and are instructed to correct ungrammatical strings, they will meet the demands of their training instructions by glancing from letter positions 1 to 5, 2 to 6, 3 to 7, and 4 to 8 in order to identify and correct ungrammatical letter pairs. Consequently, apply rules participants will not process training strings in the sequential left-to-right processing order required to acquire bigram knowledge.

Method

Participants. 24 students at University College London performed the experiment and were paid £5 for taking part. 12 participants were allocated to the Match group and 12 to the Apply Rules group.

Materials. The same training letter strings were used as in Experiment 2. Thirty bigrams were used for the recognition task. 18 of these test bigrams had been used to construct training strings and were therefore "old", while 12 test bigrams had not been seen during training and were therefore "new". Each participant saw the bigram set that matched the specific string set that they saw in training (out of 15 different sets of rules).

Procedure. While the match group memorised training strings in the standard match task, the apply rules group corrected flawed letter strings in a new apply rules task. Both groups carried out a new recognition test, followed by the standard classification test.

Apply Rules Task. The apply rules group was given the rules of the biconditional grammar and asked to identify grammatical flaws in strings of letters such as DFGX.FDLK, which were processed one string at a time over 72 training trials. Each string had between two and four letters that violated the rules of the grammar. Participants were asked to indicate whether 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. The full instructions are shown in Appendix F.

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 judgement on how familiar each letter pair felt to them. The test comprised two blocks of 30 bigrams presented in different random orders across blocks and 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, (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 or not. The precise instructions can be seen in Appendix F.

Results

Table 9 shows the mean percentage of strings correctly selected from the list of three strings by the match group and the mean percentage of letters correctly labelled as grammatical or ungrammatical by the apply rules group across four blocks of 18 trials. The data for one apply rules participant was removed from further analysis as he achieved only 49% accuracy in training. Both groups demonstrated near-perfect accuracy. A one-way ANOVA for the match group, with block as a within-subjects variable, found no overall effect of block, F (3, 33) = 1.05, MSE =

31.18. A similar one-way ANOVA for the apply rules group also found no overall effect of block, $F(3, 30) = 1.27$, $MSE = 3.33$.

The results of the classification test (see Table 10) are presented before those of the recognition test to demonstrate that - by this test - the match and apply rules groups had acquired different types of knowledge during training (letter-fragments versus rules). The match group performed at chance levels in relation to the rules of the grammar with a mean percentage correct of 50% and confidence intervals of 46 to 54%, whereas the apply rules group achieved near-perfect levels of accuracy with a mean percentage correct of 97% and confidence intervals of 96 to 98%. In contrast, the match group performed at above chance levels according to ACS with a mean percentage correct of 59% and confidence intervals of 52% to 67% while the mean percentage correct for the apply-rules group was 51% with confidence intervals of 50% to 52%.

A three-way ANOVA comparing the percentage of items classified as grammatical, with group (match or apply rules) as a between-subjects variable and both grammaticality and ACS as within-subjects variables, found significant effects of grammaticality, $F(1, 21) = 433.75$, $MSE = 117.51$, and ACS, $F(1, 21) = 6.84$, $MSE = 358.16$, and significant interactions of Group x Grammaticality, $F(1, 21) = 417.92$, $MSE = 117.51$, and Group x ACS, $F(1, 21) = 4.36$, $MSE = 358.16$. There was no significant effect of group, $F(1, 21) = 1.09$, $MSE = 623.46$, no Grammaticality x ACS interaction, $F < 1$, and no Group x Grammaticality x ACS interaction, $F < 1$.

Separate two-way ANOVAs comparing the percentage of items classified as grammatical were conducted with both grammaticality and ACS as within-subjects variables. In the match group, there was a significant effect of ACS, $F(1, 11) = 6.15$, *MSE* = 673.50, but no effect of grammaticality, $F < 1$, and no Grammaticality x ACS interaction, $F < 1$. In the apply rules group there was a significant effect of grammaticality, $F(1, 10) = 4841.61$, $MSE = 19.81$, a marginal effect of ACS, $F(1, 10) = 4.23$, p = .067, $MSE = 11.28$, and no Grammaticality x ACS interaction, $F < 1$.

The mean sensitivity d'_g score of zero with confidence intervals of –0.24 to +0.24 indicated that the match group discriminated grammatical from ungrammatical strings at chance levels, whereas the mean sensitivity d'_g score of 3.86 with confidence intervals of 3.59 to 4.13 indicated that the apply rules group discriminated at above chance levels. The mean d'_{acs} sensitivity score of 0.59 with confidence intervals of 0.08 to 1.10 indicated that the match group discriminated at above chance levels, while the mean d'_{acs} sensitivity score of 0.05 with confidence intervals of 0.01 to 0.09 indicated that the apply rules groups' performance was only marginally above chance. In summary, the results of the classification test reveal that during training the match group acquired ACS but not rule knowledge, while the apply rules group acquired almost perfect rule knowledge and only a small amount of ACS knowledge.

As predicted, the match group reliably recognised bigrams as old or new with a mean percentage correct of 63% and confidence intervals of 57 to 69% and a mean d'_{rec} sensitivity score of 0.67 with confidence intervals of 0.16 to 1.1. The criterion of -0.97 (*SEM* = 0.17) indicating a bias towards calling items "old". A related *t* test comparing the percentage of old bigrams correctly recognised as old (*M* = 85%, *SEM* = 4.72) with the percentage of new bigrams incorrectly recognised as old (*M* = 69%, *SEM* = 6.25) confirmed that, as predicted, match participants could reliably discriminate between old and new bigrams, t (11) = 2.41.

Counter to the episodic-processing prediction, the apply rules group also discriminated between old and new bigrams with a mean percentage correct of 59% and confidence intervals of 55 to 63% and a mean d'_{rec} sensitivity score of 0.28 with confidence intervals of 0.04 to 0.52. The mean criterion of –0.72 indicated a bias toward calling bigrams old. A related t test comparing the percentage of old bigrams correctly recognised as old (M = 76%, SEM = 4.79) with the percentage of new bigrams incorrectly recognised as old (M = 68%, SEM = 5.48) indicated that apply-rules participants could reliably discriminate between these two types of bigrams, t (10) = 2.35. In fact, the match and apply-rules groups were equally sensitive to whether bigrams were old or new, t (21) = 1.33, SE = 3.65.

A plausible explanation of the apply rules group's unexpected ability to discriminate between old and new bigrams is that they could not differentiate between the conceptual fluency of the rule-related letters (i.e., D-F, F-D, G-L, L-G, K-X, and X-K) and six old training bigrams based on the same letter pairs (i.e., DF, FD, GL, LG, KX, and XK). To investigate this possibility, the apply rules recognition sensitivity scores were recalculated without the six rule-related bigrams. In contrast to the analysis of all 30 bigrams, the results of this second analysis for the 24 non-rule-related bigrams indicated that the apply rules group performed at chance levels with a mean d'_{rec} score of 0.17 and confidence intervals of –0.08 to 0.42. However, there was still no difference in sensitivity between the match[2] and apply rules groups, $t\,(21) = 1.70$.

Discussion

The results of the match group support the prediction that memorisation leads to fragment knowledge, as the match group were able to discriminate between old bigrams they had seen during training and new bigrams that only appeared in the recognition test. This evidence supports previous studies that used finite-state grammars to demonstrate that participants use fragment knowledge at test (e.g., Dienes, Broadbent & Berry, 1991; Dulany, Carlson & Dewey, 1984; Perruchet & Pacteau, 1990).

Unfortunately, the match group's results do not provide convincing evidence for explicit bigram knowledge as recognition decisions may have been based on an automatic familiarity process rather than conscious recollection (Jacoby, 1991; Hintzman & Curran, 1994). Participants may simply have found old chunks more familiar and on that basis called them old without necessarily consciously recollecting them. This view is supported by recent simulations (e.g., Kinder & Shanks, submitted; Nosofsky & Zaki, 1998) which show that a single familiarity system can explain both classification and recognition performance in both normal and amnesic subjects (see Chapter 7 for a detailed explanation).

An alternative view is that recognition is a separate explicit memory task that is

---

2 A comparison between match sensitivity for all 30 bigrams and apply-rules sensitivity to only the 24 non-rule-related bigrams was warranted as the match group were equally able to discriminate between old and new bigrams when their performance was measured for the sets 30 and 24 bigrams, $t\,(11) = 1.39$.

little contaminated by familiarity (Aggleton & Shaw, 1996; Yonelinas, 1997; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1988). For example, evidence from Knowlton and Squire (1996) indicates that amnesics are significantly worse than normal participants at fragment recognition while performing at equivalent levels to normal participants in classification tests. These results suggest that recognition performance is based on a process of recollection that is separate from a fluency-based familiarity process used in classification performance. Hence, by this account amnesics are impaired in their ability to recollect, yet have normal familiarity-based processing.

The issue of whether fragment-based performance is based on unconscious familiarity or conscious recollection processes continues to be addressed in the next chapter by asking participants to provide subjective confidence ratings on each test trial (see Cheesman & Merikle, 1984; Dienes & Perner, 1996, 1998). The advantage of this method is that it directly compares conscious mental states with accuracy of performance.

Turning to the apply rules group, once the six letter fragments that duplicated the rule pairs were removed from the analysis of their recognition data, there was no evidence of an ability to discriminate between old and new bigrams. Although edit learners in Experiments 1 and 2 classified solely on the basis of rule knowledge, there was always a possibility that they had also learned about letter fragments. The results of this recognition test support the episodic-processing account by demonstrating that, in fact, the apply rules group did not acquire fragment knowledge and that fragment knowledge is only acquired by processing training strings in a sequential right-to-left manner as demanded by match instructions.

Chapter 5 presented evidence that according to objective cued-recall (Experiment 5) and bigram recognition (Experiment 6) tests, participants are aware of the rule and letter-fragment knowledge they use to classify test items. However, there is a second possibility, namely that knowledge can be implicit at a subjective level (Cheesman & Merikle, 1984; Dienes & Perner, 1996, 1998), where there is no relationship between accuracy and subjective confidence.

Dienes and Perner (1996, 1998) distinguished between two different domains of knowledge. The first domain contains facts (e.g., the rules of a grammar or letter fragments seen during training) and supports involuntary classification responses, whereas the second domain contains attitudes towards facts (i.e., "knowing" the rules of a grammar or "recollecting" seeing certain letter fragments in training strings) and supports the conscious, voluntary application of knowledge. The implications of this dual knowledge framework are that: (1) on classification test trials where performance is based solely on involuntary factual representations of the properties of training strings, participants will be accurate, yet have a subjective experience of guessing (i.e., the results will support a "guessing criterion" of implicit knowledge); and (2) because the knowledge driving test performance will vary in the subjective support it receives there will be no overall correlation between accuracy and subjective confidence ratings across all classification trials (i.e., the results will support a "zero correlation criterion" of implicit knowledge). Dienes, Altmann, Kwan, and Goode (1995) discuss the guessing and zero correlation criteria in detail.

An alternative view (Whittlesea & Dorken, 1997) is that one type of mental representation is sufficient to account for both accuracy and subjective experience. By this account, episodic representations are created during training that preserve the processing and organisation imposed on stimuli in order to meet the demands of the training task. At test, the retrieval of episodic-processing knowledge drives both accuracy and confidence judgements. Test items that are similar to training items are processed more fluently than test items that are dissimilar to training items and the relative fluency of processing forms the basis of both accuracy and subjective confidence judgements. In contrast to the dual-knowledge account, this unitary processing approach predicts that because accuracy and confidence are driven by the same information, (1) participants will perform at chance when they say they are guessing (i.e., there will be no evidence to support the guessing criterion), and (2) there will be a correlation between accuracy and confidence (i.e., there will be no evidence to support the zero correlation criterion).

Conflicting results have been reported in five previous AGL studies that tested the guessing and zero correlation criteria using the same-letter set for training and test stimuli. Chan (1992, cited by Dienes, Altmann, Kwan, & Goode, 1995) and Dienes, Altmann, Kwan, and Goode (1995) found evidence for implicit knowledge according to both criteria. Dienes and Altmann (1997, Experiment 2) found evidence for the guessing criterion, but not the zero correlation criterion. Redington, Friend, and Chater (1996) and Whittlesea, Brooks, and Westcott (1994) found no evidence of implicit knowledge by either criteria.

Unfortunately the studies that supported the guessing and zero correlation criteria only measured the grammatical status of test strings and failed to control for the effects of fragment knowledge (ACS) on test performance[1]. As suggested by much previous AGL research, there is evidence to support fully (e.g., Dienes, Broadbent & Berry; 1991; Dulany, Carlson & Dewey, 1984; Johnstone & Shanks, 1999; Perruchet, 1994; Perruchet, Gallego, & Pacteau, 1992; Perruchet & Pacteau, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990) or partially support (e.g., Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997) the view that classification test performance is driven by the overlap of letter fragments between training and test items. Indeed the results of Experiments 1 and 2 (Chapters 3) indicate that when ACS is equated across grammatical and ungrammatical test strings, match participants perform at chance in relation to the rules of the grammar.

Therefore, if the guessing and zero correlation criteria are measuring real effects then these effects should be stronger when assessed according to the true basis of classification decisions (i.e., ACS) than when assessed according to the grammaticality of test strings. The aims of Experiment 7 were to assess the relationship between accuracy and confidence for both rule and ACS knowledge and to test the predictions of the dual knowledge (Dienes & Perner, 1998) and episodic-processing (Whittlesea & Dorken, 1997) accounts of the relationships between accuracy and subjective confidence.

---

[1] For example, the grammatical and ungrammatical test strings used by Dienes and Altmann (1997) differed for both anchor ACS ($\underline{M}$ = 5.71 versus 1.09), $\underline{t}$ (56) = 15.65, $\underline{SE}$ = .30, and global ACS ($\underline{M}$ = 8.83 versus 2.86), $\underline{t}$ (56) = 8.88, $\underline{SE}$ = .67.

Experiment 7

Participants were trained and tested on the stimuli used in Experiment 5, as these strings unconfound the contributions of rule and ACS knowledge in classification performance. On each classification trial, participants were first asked to indicate whether the test string was grammatical or ungrammatical and then asked to specify how confident they were in the accuracy of their initial response. Confidence was rated on a scale from 50 to 100%, where 50% indicated guessing and 51-100% indicated increasing levels of confidence.

Based on the results of Experiments 1-6, it was predicted that classification performance would be at chance in relation to grammaticality, but above chance in accordance with ACS. Support for the dual-knowledge account required that participants classify at above chance levels in the 50% confidence category (guessing criterion), with no relationship between accuracy and confidence for all responses (50-100%) (zero correlation criterion). In contrast, support for the episodic-processing account required that participants perform at chance when they used the 50% confidence rating and that there was a relationship between confidence and accuracy across all responses (50-100%).

Method

Participants. 21 students at University College London performed the experiment and were each paid £5 for taking part.

Procedure. The same match task, letter strings and classification test were used as in Experiment 5, with a minor modification to the classification test. On each trial, participants were first asked to press the Y key if they believed that the string conformed to the rules of the earlier strings or the N key if they believed that the string did not conform to the rules of the earlier training strings. They were then asked to rate how confident they were that their Y or N response was correct on a scale ranging from 50% (complete guessing) to 100% (complete certainty). They were encouraged to use any integer within this range to indicate as accurately as possible their confidence in their Y or N judgement.

Results

The mean percentages of training items across four blocks of training on which participants made correct responses was 89%, 90%, 91%, and 93% with an overall accuracy level of 91%. A one-way ANOVA comparing accuracy with training block (1 to 4) as a within-subject variable, indicated that there was no effect of block, $F\,(3, 60) = 1.51$, $MSE = 31.81$. This indicated that accuracy levels remained at around 90% across all training blocks.

The mean percentage of items classified as grammatical for the four test item types was 70% (GH), 32% (GL), 69% (UH), and 38% (UL). The overall mean percentage correct according to the rules of the grammar (48%, with confidence intervals of 46% to 50%) indicated chance performance. A two-way ANOVA comparing the percentage of items classified as grammatical (Y responses), with both grammaticality and ACS as within-subjects variables, found a significant effect of ACS, $F\,(1, 20) = 31.30$, $MSE = 790.63$, but no effect of grammaticality, $F\,(1, 20) = 2.37$, $MSE = 89.24$, and no Grammaticality x ACS interaction, $F\,(1, 20) = 2.80$, $MSE = 85.28$. The sums of squares calculated for the ANOVA indicated that ACS knowledge accounted for 42% of the variance in classification performance, whereas grammaticality accounted for less than 1% of the variance. These results entirely replicate the findings of Experiments 2-4 where memorisers classified on the basis of ACS, but not rule knowledge.

Signal detection measures were calculated for each participant to assess sensitivity (d'_g) in judging the grammaticality of test strings. The mean d'_g of –0.08 with confidence intervals of –0.18 to 0.02, confirmed that participants did not classify on the basis of rule knowledge. The mean criterion of –0.06 demonstrated that there was little classification bias.

The mean percentage ACS accuracy and mean d'_{acs} scores were calculated by taking grammatical responses to high ACS strings as "correct" and grammatical responses to low ACS strings as "incorrect". According to both of these measures, participants classified on the basis of ACS knowledge as there was a mean percentage correct of 67% with confidence intervals of 61% to 73%, and a mean d'_{acs} score of 1.05 with confidence intervals of 0.64 to 1.46. The mean criterion of –0.09 indicated that there was little classification bias.

The relationship between subjective confidence and classification accuracy was examined by counting the total number of trials on which participants used ratings of 50%, 51-60%, 61-70%, 71-80%, 81-90%, and 91-100% and then calculating the proportion of these trials on which participants were correct in their response. Accuracy was calculated on the basis of both the grammatical and ACS status of test strings. For the latter, classifying a high ACS string as grammatical was "correct", while classifying a low ACS string as grammatical was "incorrect". The frequency counts and proportions correct are shown in Table 11 and the proportions correct are shown graphically in Figure 6.

Table 11

Classification Accuracy by Subjective Confidence Categories for Experiment 7

Trials	50%	51-60%	61-70%	71-80%	81-90%	91-100%	51-100%
Frequency Counts							
Grammatically Correct	110	228	203	182	142	111	866
ACS Correct	134	294	274	260	212	180	1220
Total Number of Trials	239	463	407	394	281	232	1777
Proportions Correct							
Grammatical	.46	.49	.50	.46	.51	.48	.49
ACS	.56	.63**	.67**	.66**	.75**	.78**	.69**
Mean Confidence Ratings	.50	.58	.68	.78	.89	.98	

Note. ** indicates $p < .001$.

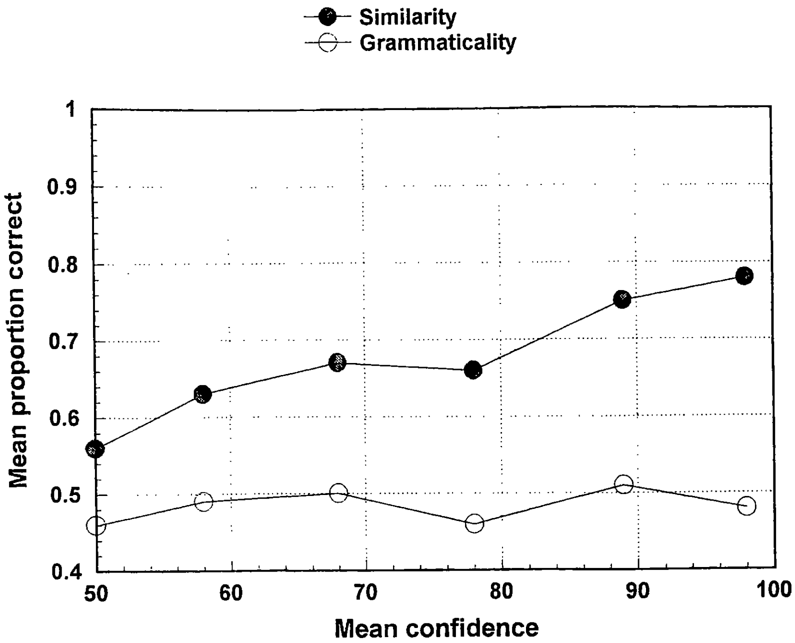

Figure 6: Mean accuracy and confidence scores for Experiment 7

The results for grammatical accuracy suggest that there was no relationship between classification accuracy and subjective confidence as participants performed at chance levels across all confidence categories while confidence rose across those categories. In contrast, the results for ACS accuracy show that accuracy increased with subjeetive confidence as there was a reliable linear trend, t (20) = 2.69. When participants said they were guessing (50% subjective confidence) their performance was reliably at chance for grammaticality (.46), but was marginally significant for ACS (.56) ($z = 1.81, p = .07$), when assessed by binomial tests. However, this latter effect is very weak and disappears if only a single participant – who had 28 out of 39 correct responses in the 50% confidence category - is removed from the analysis ($z = 0.78$). Moreover, the test should be interpreted cautiously because one of its assumptions, strict independence of observations, is violated: the same participants made repeated responses.

Another way of analysing these data is to calculate for each participant the proportion of correct responses in relation to ACS in the 50% confidence category. In this category, 5 participants made no responses, more than half of the responses of 9 participants were correct and more than half the responses of 7 participants were incorrect. For the 16 participants who did respond at the 50% confidence level, the mean proportion correct was .54 with confidence intervals of .39 to .69 which indicates that when participants said they were guessing, they performed at chance levels ($z = 0.49$). Thus the results for the 50% confidence trials provide extremely weak evidence for the guessing criterion. The marginally significant effect in the binomial test is largely attributable to one participant who made 39 of the 239 responses and the analysis of mean level of accuracy across all participants was not significant.

However, these results also provide only weak support for the prediction that participants would perform at chance in relation to ACS when they gave a 50% confidence rating as the confidence limits around the mean proportion correct (CI = .39 to .69) also include the above chance mean ACS proportion correct across all six confidence categories (M = .67). Consequently, the results of this test cannot distinguish between chance and above chance performance when participants report that they are guessing.

Linear regression was used to assess whether knowledge was implicit according to the zero correlation criterion. Up to six confidence/rule or ACS accuracy datapoints (M = 5, range 2 to 6) contributed to grammaticality and ACS regression equations for each participant. Each confidence/accuracy datapoint was weighted by the number of trials that had contributed to that confidence category. Two sets of mean intercept and slope coefficients were created using the 21 grammaticality and 21 ACS equations.

The mean intercept of .47, with confidence intervals of .43 to .51, and slope of .03, with confidence intervals of $-.15$ to .21, indicated that there was no relationship between grammatical accuracy and confidence. In contrast, the mean intercept of .51, with confidence intervals of .41 to .61, and slope of .64, with confidence intervals of .35 to .93, demonstrated a stronger relationship between confidence and sensitivity to ACS. Chance accuracy in relation to both rule and ACS knowledge accompanied 50% confidence ratings.

Discussion

Experiment 7 investigated the possibility that knowledge applied accurately in a classification test might be implicit in relation to subjective confidence as measured by guessing and zero correlation criteria. Two analyses of the mean proportion correct across participants for the 50% confidence category were unable to determine whether participants performed at chance when they said they were guessing. However, regressions across all confidence categories (50-100%) supported the conclusion that there was no evidence for implicit knowledge according to either criterion.

The results of the linear regressions support the unitary episodic-processing account (Whittlesea & Dorken, 1997) and prior studies that found no evidence for the guessing or zero correlation criteria (Redington, Chater, & Friend, 1996; Whittlesea, Brooks, & Westcott, 1994). Evidence that accuracy and confidence, in relation to ACS, are related across the 50-100% confidence range suggests that one type of knowledge (i.e., fluency) can support both classification and confidence judgements.

While prior studies used finite-state grammars provide evidence of implicit knowledge according to subjective criteria (Dienes & Altmann, 1997; Dienes et al., 1995), Experiment 7 was based on a biconditional grammar that allowed the contributions of rule and ACS knowledge to be independently assessed. As in Experiments 1 to 6, unconfounding rule and ACS knowledge indicated that unaware memorisers use ACS rather than rule knowledge.

If ACS is the real basis of classification performance, then measuring the relationship between accuracy and confidence on its true basis (i.e., ACS) should have resulted in greater accuracy in the 50% confidence category than the rule-based 63% calculated by Dienes and Altmann (1997, Experiment 2, same letters group). As the results of Experiment 7 indicated ACS accuracy of 54% in the 50% confidence category the existing data need not undermine the notion of a subjective threshold.

Everyday experience suggests that we have an ability to store two types of knowledge independently. Episodic knowledge allows us to recall the context of specific experiences, such as what we did on our last holiday. In contrast, general knowledge of the properties of classes of objects and events is not tied to specific experiences and enables us to make judgements about novel instances. Thus, we can judge the grammaticality of a sentence we have never heard before and read words in unfamiliar handwriting. Moreover, as we are not normally intending to abstract underlying rules, it appears that we acquire general knowledge in an incidental and unconscious manner.

A good deal of evidence supporting this dual systems account has come from AGL studies. For example, Knowlton, Ramus, and Squire (1992) presented evidence that despite being selectively impaired in making judgements about specific items, amnesics had intact general knowledge of an artificial grammar. Evidence such as this seems to support the notion that we have separate learning systems that acquire implicit, general and explicit, specific knowledge.

However, despite 30 years of AGL research, there is still debate over the form of knowledge acquired in incidental learning situations. While the dual systems account suggests that memorising letter strings, without realising that those letter strings were constructed according to a set of rules, leads to both implicit knowledge of the rules of the grammar *and* episodic knowledge of specific training examples (e.g., Cleeremans, 1993; Lewicki & Hill, 1989; Reber, 1967, 1989), it has also been suggested that behaviour that appears to be rule-based can also be explained by an episodic system that acquires specific knowledge of a collection of training exemplars (Brooks, 1978; Brooks & Vokey, 1991; Neal & Hesketh, 1997; Vokey & Brooks, 1992), the frequency statistics of letter fragments in training items (e.g., Dulany, Carlson, & Dewey, 1984; Perruchet & Pacteau, 1990), or processing training items in particular ways in order to meet the demands of the training task (Whittlesea, 1997a, b; Whittlesea & Dorken, 1993, 1997; Whittlesea & Williams, in press; Whittlesea & Wright, 1997; Wright & Whittlesea, 1998).

As well as varying in assumptions about the form of knowledge acquired and whether we have one or two learning systems, these four accounts also differ in other ways. While the implicit rule-abstraction, exemplar, and fragments accounts suggest that knowledge acquisition is stimulus-driven, the episodic-processing account suggests that knowledge acquisition is driven by the processing required to meet the demands of the training task. Finally, there is also debate over whether incidentally acquired knowledge is always applied implicitly (i.e., the abstraction account suggests that implicit rule knowledge is applied in classification tests), applied explicitly (exemplar and letter-fragment knowledge), or applied implicitly or explicitly depending on whether test instructions disguise or alert participants to the relationship between fluent processing of test items and the information acquired during training (episodic processing account).

In Chapter 2, a reanalysis of a study that supported the dual-systems account (Meulemans & Van der Linden, 1997) demonstrated that finite-state grammars, which dominate AGL research, do not allow us to unconfound the contributions of rule, exemplar and fragment knowledge in test performance. Because these grammars (e.g., Brooks & Vokey, 1991, see Figure 2 in Chapter 2) use transition rules that dictate legal consecutive letters in particular letter string locations (e.g., 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), it is has not been possible to determine whether incidental learning leads to rule, exemplar (e.g., test item MVXTX is similar to training example MVXTR), or fragment knowledge (e.g., all the training strings began with MV, MX, VM or VX) in test performance.

The inadequacies of finite-state grammars were overcome by using a biconditional grammar that generates strings of eight 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 other should be an F, where there is a G the other letter should be an L, and where there is a K, the other letter should be an X. As each of the three rules can occur in any of the letter locations and as rule-related positions have three intervening letters, it was possible to design letter strings that unambiguously test for knowledge of rules versus exemplars (Experiment 1), rules versus fragments (Experiments 2, 5, 6, and 7) and exemplars versus fragments (Experiments 3 and 4).

The first priority was to establish what form of knowledge is acquired by incidental memorisation (Experiments 1 to 4 in Chapters 3 and 4), as without this information it is impossible to determine whether that knowledge is implicit or explicit (Experiments 5 to 7 in Chapters 5 and 6). The results of Experiments 1 to 4 indicated that incidental memorisation leads to knowledge of fragments (Experiments 2 to 4). Only 1 out of 16 memorisers noticed the rules of the grammar (Experiments 1 and 2) and that one participant could state the biconditional rules. There was no evidence of exemplar knowledge when for each test item there had been one

(Experiment 1) or six (Experiment 3) similar training items. However, there was marginal support for exemplar knowledge with 24 similar training items (Experiment 4), though in this experiment exemplar knowledge was confounded with knowledge of four-letter fragments.

As there was strong support for fragment learning (Experiments 2 to 4) and one participant had noticed the biconditional rules (Experiment 1), the experiments reported in Chapters 5 and 6 focused on replicating the rule and fragment effects found in Experiments 1 to 4 and assessing whether rule and fragment knowledge acquired by memorising training items is implicit or explicit. In Experiment 5, the 16 out of 70 participants who used rule knowledge to classify test items were able to specify at least four of the six rule pairs (i.e., D-F, F-D, G-L, L-G, K-X, and X-K) in a cued-recall test. Memorisers could discriminate between old fragments seen during training and novel fragments (Experiment 6) and demonstrated a relationship between fragment-based classification performance and subjective confidence ratings (Experiment 7).

In order to distinguish between stimulus- and processing-driven learning accounts, the classification performance of passive memorisers was compared to that of active hypothesis-testers (Experiments 1 and 2). Evidence that memorisers classified on the basis of letter-fragments while successful hypothesis testers classified on the basis of rules suggests that knowledge acquisition is process- rather than stimulus- driven. These results are consistent with memorisers processing training strings in the sequential left-to-right manner required to encode letter-fragments and rule learners processing training items by glancing from locations 1 to 5, 2 to 6, 3 to 7, and 4 to 8 to search for letter pairs that violated the rules of the grammar. The results of the recognition test in Experiment 6 provide further support for this conclusion as memorisers, but not rule learners were able to discriminate between old letter fragments seen during training and novel letter fragments.

Finally, whether knowledge is stored in an implicit, explicit or neutral form was investigated by comparing performance on tests that did and did not alert participants to the relationship between the knowledge they acquired during training and the demands of the tests. The strongest evidence for implicit performance comes from Experiments 3 and 4 where training strings contained exemplar and fragment but not rule knowledge. Despite their lack of rule knowledge, memorisers did not object to test instructions asking them to classify test items as grammatical or ungrammatical. In fact, their classification results indicate that they unconsciously attributed high processing fluency created by overlapping letter fragments between training and test items to grammaticality, and low fluency created by novel test fragments to ungrammaticality. In contrast, in Experiment 6, memorisers applied fragment knowledge explicitly as they were able to meet the demands of a fragment recognition test that asked them to discriminate between fragments seen during training and novel fragments. In this case memorisers attributed high processing fluency to a test fragment being old and low fluency to it being new. These results suggest that knowledge is stored in a neutral form that can be expressed implicitly (i.e., using fragment knowledge to classify test items as grammatical or ungrammatical) or explicitly (i.e., using fragment knowledge to determine whether test fragments are old or new) depending on whether or not test instructions alert participants to the relationship between processing fluency and the knowledge they acquired during training.

No evidence for dual implicit rule versus explicit exemplar-based learning systems

As the type of knowledge acquired varied between match and edit instructions (Experiments 1 and 2) there was no support for Lewicki and Hill's (1989, p.240) claim that participants are passive "consumers" of knowledge, or Cleeremans' (1993, p.19) claim that the acquisition of rule knowledge is stimulus-driven. As memorisers predominantly used fragment knowledge (Experiments 2, 5, 6, and 7) and the few who acquired rule knowledge could report the rules verbally (Experiments 1 and 2), and in cued-recall tests (Experiment 5) there was no support for a separate learning system that unconsciously abstracts rule knowledge (e.g., Knowlton & Squire, 1994, 1996; Knowlton, Ramus, & Squire, 1992; Meulemans & Van der Linden, 1997; Reber, 1967, 1989; Reber & Allen, 1978; Reber & Lewis, 1977).

No evidence for a system that encodes training exemplars

The lack of exemplar effects in Experiments 1 (one similar training item) and 3 (six similar training items), and only marginal support for exemplar knowledge in Experiment 4 (24 similar training items) challenges claims that memorisers encode a collection of training exemplars in a stimulus-driven manner (e.g., Brooks, 1978; Brooks & Vokey, 1991; McAndrews & Moscovitch, 1985; Neal & Hesketh, 1997; Vokey & Brooks, 1992) and the predictions of instance models (e.g., Hintzman, 1986, 1988; Medin & Schafer, 1978; Nosofsky, 1986, 1988) that test items that are highly similar to training items (e.g., differing by only one letter), are more likely to be called grammatical than dissimilar test items. However, evidence that participants acquired four-gram knowledge supports Servan-Schreiber and Anderson's (1990) competitive-chunking model as it suggests that with longer training participants may eventually encode whole exemplars in the form of a nested hierarchy of successively longer letter fragments.

These results also support Shanks, Johnstone & Staggs's (1997, Experiment 4) and Knowlton and Squire's (1994) findings that when fragment knowledge is controlled across similar and dissimilar test items, there is no evidence for exemplar effects. In accordance with the episodic-processing account, these results suggest that exemplar effects will only occur if participants receive training instructions that can only be met by processing training items in a way that binds together individual features. The strong fragment effects in Experiments 1, 3 and 4 suggest that mentally rehearsing exemplars as a series of letter fragments was sufficient to meet the demands of the training task and as a result only fragment knowledge was acquired.

Support for a single fragment learning system

In contrast to the lack of conclusive evidence for exemplar (Experiments 1, 3, and 4) or implicit rule knowledge (Experiments 1, 2, 5, 6, and 7), there was strong support for claims that participants use fragment knowledge (Experiments 2 to 7) to classify test strings as grammatical or ungrammatical (Dienes, Broadbent, & Berry, 1991; Dulany, Carlson, & Dewey, 1984; Johnstone & Shanks, 1999; Perruchet, 1994; Perruchet & Pacteau, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson,

1990) that undermines prior suggestions that fragment knowledge cannot fully account for classification performance (Gomez & Schvaneveldt, 1994; Knowlton & Squire, 1994, 1996; Mathews et al., 1989; Meulemans & Van der Linden, 1997; Reber & Allen, 1978).

Unfortunately the results of Experiments 2 to 7 do not allow us to determine whether fragment knowledge was acquired actively in order to meet the demands of the test instructions or in a passive, stimulus-driven manner. A comparison of the classification performance of the match and apply rules groups in Experiment 6 suggests that knowledge acquisition depends on the active processing carried out during training. The match group mentally rehearsed training strings in a sequential left-to-right manner and subsequently showed fragment, but not rule effects in their test performance. In contrast, the apply rules group processed their training items by glancing from letter locations 1 to 5, 2 to 6, 3 to 7, and 4 to 8, in order to correct ungrammatical letter pairings. As a result, the apply rules group showed an effect of rule knowledge and no effect of fragment knowledge in their test performance. In contrast, the results of Experiment 2 suggest that fragment knowledge can be acquired in a passive, stimulus-driven manner as edit participants who failed to discover the rules of the grammar by active hypothesis-testing showed an effect of fragment knowledge in their classification performance.

It also appears that fragment knowledge is held in a neutral form that can be applied implicitly or explicitly as participants unconsciously attributed high ACS to grammaticality in classification tests (Experiments 2 to 7), while attributing high ACS to "old" in a recognition test (Experiment 6). Evidence that fragment knowledge is stored in a neutral form challenges earlier suggestions that fragment knowledge is explicit (e.g., Dulany, Carlson, & Dewey, 1984; Perruchet, 1994; Perruchet & Pacteau, 1990) or implicit (Servan-Schreiber & Anderson, 1990).

However, the conclusion that fragment-based fluency drives grammaticality decisions runs counter to the conclusions of Buchner (1994) who found that faster identification of letter strings in a perceptual clarification task was not a predictor of later classifying a string as grammatical. After memorising grammatical letter strings, participants were told that the training strings conformed to a set of rules and asked to carry out two-stage test trials. First, they were asked to press a button as soon as they could identify a letter string obscured behind a solid black mask that gradually clarified. Then, depending on test instructions, they indicated whether they had seen the string during training (50% of test trials) or whether the string was grammatical (50% of test trials). The set of test strings comprised 20 old grammatical strings, 20 new grammatical strings and 20 new ungrammatical strings, that were each presented on a recognition and a classification trial.

While Buchner found that identification speed predicted recognition responses, as test items identified more quickly were more likely to be called old than items identified more slowly, there was no such relationship between identification speed and grammaticality judgements. This evidence suggests that grammaticality judgements are not based on processing fluency and hence conflicts with the earlier conclusion that in Experiments 2 to 7, memorisers classified on the basis of fragment-based fluency.

However, there are significant differences between the test stimuli, instructions and procedures used in Experiments 2 to 7 and in Buchner's study that according to the discrepancy-attribution hypothesis (Whittlesea & Williams, in press) would lead to different outcomes. In Experiments 2 to 7, participants were only tested on novel items and only made classification responses. In contrast, Buchner's participants saw old and new test items and made both classification and recognition responses. According to the discrepancy-attribution hypothesis, test performance is only based on implicit familiarity when test instructions do not refer to knowledge acquired during training. Thus in Experiments 2 to 7, when participants were told that all test strings were novel and asked to classify these strings as grammatical or ungrammatical, they were not alerted to the fact that their incidentally acquired fragment knowledge would lead to variations in processing similar and dissimilar test stimuli. As a result, these conditions created a discrepancy between the expectation that all strings were novel and surprising variations in the fluency of processing test strings. As participants were unaware of any other source of fluency, they implicitly attributed high fluency to a test item being grammaticality and low fluency to a test item being ungrammatical.

In contrast, Buchner's participants received training instructions that alerted them to information acquired during training as they were informed that they would be tested on old as well as new test stimuli. In this case the discrepancy-attribution hypothesis predicts that participants will use recollection and not fluency to make all test decisions. On recognition trials, old strings contained only familiar fragments whereas new strings contained familiar and unfamiliar fragments. Thus old items were identified more quickly than new items and there was a relationship between speed of identification and old/new status. However, grammatical test items included both old items containing only familiar training fragments and new items containing a mixture of old and new training fragments. Thus there was no clear-cut relationship between speed of identification and calling a string grammatical.

Future research could investigate whether identification speed predicts grammaticality and recognition judgements by using the same training procedure and stimuli as Experiment 5. Half of these test strings are grammatical and half are ungrammatical. Orthogonal to the grammaticality manipulation half of the strings comprise letter fragments from training items (high ACS items) and half contain largely novel fragments (low ACS items). In a between-subjects experiment, all participants would memorise grammatical training strings, be tested on novel test strings, and carry out two-stage test trials. One group would receive test instructions telling them that they were to see old and new test items and that after identifying each string they were to indicate whether that string was old or new. A second group would be told that the training strings conformed to a set of rules, that all test items were novel and that after identifying each string they were to indicate whether it was grammatical or ungrammatical.

As the classification group would have been told that all test strings were novel and as their test instructions would not alert them to the relationship between the fragment knowledge they acquired during training and variations in the fluency of processing test items, it is predicted that participants would classify in accordance with the discrepancy-attribution hypothesis (Whittlesea & Williams, in press). That is, high ACS test items would be processed and identified faster than low ACS items and faster processing/identification would be attributed to grammaticality, whereas slower processing/identification would be attributed to ungrammaticality.

As all test strings would be novel, the recognition group would show the same ACS-based pattern of responding as the classification group, but would instead base their responses on recollection of fragments seen during training. Thus both groups would use ACS knowledge acquired during training with the result that high ACS strings would be called grammatical/old and low ACS strings would be called ungrammatical/new. Plus in both cases items called grammatical or old would be identified faster than items called ungrammatical or new. These predictions are based on evidence that participants are unlikely to learn training exemplars (Experiments 1, 3, and 4) and that they classify on the basis of ACS knowledge (Experiments 2 to 7).

Support for the episodic-processing account

The episodic-processing account (Whittlesea, 1997a, b; Whittlesea & Dorken, 1993, 1997; Whittlesea & Williams, in press; Whittlesea & Wright, 1997; Wright & Whittlesea, 1998) provides the strongest explanation of the results of Experiments 1 to 7. The knowledge acquired was consistent with the processing applied and the dimensions of test items processed in order to meet the demands of the training instructions. Thus, as a result of processing training strings in the sequential left-to-right manner required to mentally rehearse training strings, memorisers used fragment knowledge to classify test items (Experiments 2 to 7) and could discriminate between old and new fragments in a recognition test (Experiment 6). Similarly, as a result of processing the letters within training items in the order 1-5, 2-6, 3-7, and 4-8 in order to check for rule violations, edit learners and the apply rules group classified test items on the basis of rules (Experiments 1, 2, and 6) and could not discriminate between old and new fragments in a recognition test (Experiment 6).

The results also support the claim that episodic representations of processing (mental rehearsal of letter fragments or checking for rule violations) and specific structural aspects of training items (letter fragments or biconditional rules) are stored in a neutral form that can be applied implicitly or explicitly at test depending on whether test instructions disguise or alert participants to the relationship between the knowledge they acquired during training and the demands of the test. Memorisers appear to have unconsciously attributed the processing fluency of test items containing old letter fragments to grammaticality (Experiments 2 to 7), while also being able to express this knowledge explicitly in a recognition test by discriminating between old and novel fragments (Experiment 6).

Finally, there was evidence that knowledge from one episodic system is sufficient to explain classification performance (Whittlesea & Dorken, 1997), as there was a relationship between memorisers' fragment-based classification accuracy and subjective confidence ratings (Experiment 7). These results conflict with evidence that classification accuracy depends on implicit rule knowledge whereas subjective confidence is based on separate explicit representations (Dienes, Altmann, Kwan, & Goode, 1995; Dienes & Altmann, 1997). For example, Dienes and Altmann (1997, Experiment 2) found that participants classified accurately in relation to the rules of a finite-state grammar when believing that they were guessing.

However, before any strong conclusions can be reached about the relationship between accuracy and subjective confidence, it is necessary to replicate Experiment 7 and Dienes and Altmann's (1997, Experiment 2) study. It is necessary to repeat Experiment 7 as there was marginal support from a binomial test for accurate fragment-based classification when participants believed they were guessing, though this effect disappeared when the contributions of one participant were removed. The Dienes and Altmann study needs to be replicated as the original analysis only compared rule-based classification accuracy with subjective confidence ratings. If, as so many AGL studies now show, memorisers classify on the basis of fragment rather than rule knowledge (e.g., Experiments 2 to 7; Dulany, Carlson, & Dewey, 1984; Perruchet, 1994; Perruchet & Pacteau, 1990; Redington & Chater, 1996), the results of the Dienes and Altmann study should show greater support for the guessing criterion when accuracy is calculated on the basis of the overlap of fragment knowledge between training and test items than when it is calculated on the basis of the grammar. This is because the percentage accuracy in the 50% guessing category is likely to increase when calculated on the true basis of classification decisions (i.e., ACS).

Why do these results conflict with transfer studies?

Many people have suggested that transfer studies in which the letter-set is changed at test provide evidence of abstract rule knowledge (e.g., Gomez & Schvaneveldt, 1994; Knowlton & Squire, 1996; Manza & Reber, 1997; Reber, 1969; Reber & Lewis, 1977). But, it is difficult to understand how can this be the case when experiments with the same letter-set at test fail to yield evidence of rule abstraction (Experiments 1, 2, 5, 6, and 7). In fact, there are two similarity-based accounts that can fully explain transfer effects without requiring rule abstraction at study.

Brooks and Vokey (1991) showed 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 could be seen as similar to 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.

Redington and Chater (1996) also demonstrate that above-chance transfer performance can be explained by non-abstractionist models that learn surface fragments of training strings and classify test items that contain familiar fragments as grammatical and test items that contain novel fragments as ungrammatical. When test items are instantiated in a new letter set, the model attempts to map the test letter set onto the training letter set.

For example, if test stimuli generated from the grammar created by Brooks and Vokey (1991, see Figure 2 in Chapter 2) were based on a different letter set from training items it would be relatively easy to map the test letter set onto the training set as certain letters and combinations of letters can only appear in particular locations in grammatical strings. Using the letter-set M, R, T, V, and X, grammatical items can only begin with single letters M or V, and bigrams MV, MX, VM, and VX. Furthermore, MV can never be followed by T whereas MX, VM and VX can. Similarly VX can never be followed by R, but the other three bigrams can. As certain letters and combinations of letters are location dependent it is relatively easy to map a new letter set (e.g., B, F, K, L and N) onto the old letter set (M, R, T, V and X) by examining the first three letters of test strings. By learning the surface fragments in training items and mapping new letter sets onto old letter sets, Redington and Chater's (1996) models were able to simulate transfer effects in studies by Altmann, Dienes and Goode (1995), Brooks and Vokey (1991), Gomez and Schvaneveldt (1994), and Whittlesea and Dorken (1993, Experiment 5).

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 biconditional grammar findings that memorisers do not acquire implicit rule knowledge.

Support for rule- versus similarity-based processing strategies

Evidence that memorisation leads to fragment-based classification performance (match participants in Experiments 2 to 7), while successful hypothesis-testing (edit learners in Experiments 1 and 2) leads to rule-based classification performance supports prior research that identified two strategies for categorising stimuli using for example artificial grammars (e.g., Mathews et al., 1989; Reber et al., 1980), geometric stimuli (e.g., Nosofsky, Clark, & Shin, 1989), and cartoon animals (e.g., Allen & Brooks, 1991; Regehr & Brooks, 1993). One obvious question to ask is whether similarity- and rule-based skills can be acquired by one mechanism or whether they require separate systems.

Experiments 1 to 7 suggest that a unitary episodic-processing system that records the processing applied to training items and the specific knowledge used to meet the demands of the training instructions is sufficient to explain similarity- and rule-based performance. On this basis, memorisation would lead to episodic representations of mentally rehearsing letter strings in a sequential left-to-right manner (letter locations 1 to 2 to 3 to 4 to 5 to 6 to 7 to 8) and fragment knowledge, whereas successful hypothesis-testing would lead to representations of checking letters in the order 1 to 5, 2 to 6, 3 to 7, and 4 to 8 in order to look for rule violations and knowledge of the biconditional rules that each rule-related pair should contain D and F, G and L, or K and X.

Is it possible for one computational model to acquire exemplar and rule knowledge?

Because simple recurrent network (SRN) models perform at equivalent levels to human memorisers in AGL experiments (e.g., Dienes, Altmann, & Gao, 1999; Redington & Chater, 1996), yet lack an ability to hypothesis-test, it appears that different models are required for exemplar- and rule-based learning. However, by focusing solely on the processing carried out after successful hypothesis-testing (e.g., learners in Experiments 1 and 2), or after participants have been told the rules of the grammar (e.g., the apply rules group in Experiment 6), it can be seen that it is as easy to train neural network models to classify according to rules as it is to train them to classify on the basis of exemplar knowledge.

If learning in AGL experiments and learning models can be described solely in processing terms, then models such as SRNs can be taught to classify on the basis of rule- or exemplar-based knowledge by manipulating the order that letters are fed to the input nodes of the model. For example, when the model is being trained to memorise grammatical examples, such as DGKL.FLXG, then the input nodes will receive the eight letters in the order D-G-K-L-F-L-X-G, with an end of string delimiter after the eighth letter. In contrast, when the same model is being trained on the biconditional rules then the string DGKL.FLXG will be processed in the order of the four rule-related letter pairs as D-F, G-L, K-X and L-G. By the end of the training phase, the exemplar-processing model will have "memorised" training examples, whereas the rule-learning model will have memorised the six rule-related pairs D-F, F-D, G-L, L-G, K-X and X-K.

However, explicit human learning goes far beyond what SRN models are capable of as humans have an ability to actively hypothesis test by attending to a variety of possible relationships between letters before settling on one set of rules that explain the structure of all training exemplars. The model suggested above would not be able to learn rules such as "The second half of the string is the first half backwards", or the letter in position 2 is two letters further along in the alphabet than the letter in position 1" (e.g., If there is a B in position 1 then position 2 should contain a D).

How do Experiments 1 to 7 relate to AGL research on amnesia?

When the AGL performance of patients with anterograde amnesia is compared with that of normal adults it appears that amnesics are impaired on recognition tests yet are able to classify as well as normal adults (e.g., Knowlton, Ramus, & Squire, 1992). Such findings have been taken as support for dual learning systems where classification performance is based on general rule knowledge while recognition performance depends on acquiring information about specific training examples.

One problem for AGL evidence in favour of dual learning systems (Knowlton, Ramus & Squire, 1992; Knowlton & Squire, 1994, 1996) is that it is based on finite-state grammars. As discussed at length in Chapter 2, it is difficult to measure the relative contributions of rule- and exemplar-based knowledge in finite-state grammar classification performance as information about transition rules is inevitably confounded with ACS. In fact, when this problem was overcome by using a biconditional grammar in Experiments 1 to 7 there was no evidence that normal adults acquire implicit rule knowledge.

A second problem for dual-system accounts is that Nosofsky and Zaki (1998) were able to create dissociations in the classification and recognition performance of dot patterns using a single model. When Nosofsky and Zaki trained their Generalised Context Model (GCM) model on stimuli from one category (as is the case in AGL training), the processes of classification and recognition were identical. That is, the probability of a test item being classified as a member of the training category or recognised as old was based on the mean similarity of the test item to all stored training items. However, when they varied a sensitivity parameter that represents the ability to discriminate between distinct training exemplars, classification performance was only slightly reduced, while recognition accuracy was much reduced. This simulation indicates that dissociations in the classification and recognition performance of normal adults and amnesic patients can be produced by one learning mechanism.

Recent simulations by Kinder and Shanks (submitted) have replicated and extended Nosofsky and Zaki's (1998) findings using AGL stimuli rather than dot patterns. This time an SRN model was trained and tested on the stimuli used by Knowlton, Ramus, and Squire (1992) and the learning rate was manipulated to simulate normal and amnesic performance. The results replicated Nosofky and Zaki's findings as recognition performance was more affected than classification performance by reductions in the learning rate, just as was observed in the behavioural data.

Kinder and Shanks added to our understanding by demonstrating that the dissociation was due to differences in the ACS of the test strings used in the classification and recognition tests, as well as differences in learning rates. While the same novel ungrammatical test items were used in both tests in Knowlton, Ramus, and Squire's experiment, old training items were used in the recognition test and new grammatical items were used in the classification test. The dissociation was created by there being a larger ACS difference between the old grammatical versus new ungrammatical items used in the recognition test than between the new grammatical versus new ungrammatical items used in the classification test. Overall, these findings suggest that there is currently little evidence to suggest that general and specific information is acquired by separate systems.

What do these studies contribute to our understanding of concept learning?

AGL research is only one way of investigating how we learn to group objects and events into categories. Theories of concept learning have suggested that we determine whether a novel object is a member of a category by comparing it to a prototype (e.g., Rosch, Simpson & Miller, 1976), a collection of exemplars (e.g., Medin & Schaffer, 1978) or by applying rules to the relationships between features of the novel object (e.g., Allen & Brooks, 1991).

The results of Experiments 1 to 7 suggest that the information used to categorise stimuli depends on the demands of the training task. While Experiments 2 to 7 demonstrated that memorisers used information about surface features (i.e., letter fragments) to classify novel test items into categories of grammatical and ungrammatical letter strings, Experiments 1 and 2 showed that successful hypothesis-testers could categorise the same novel test items on the basis of a set of rules about the relationships between pairs of single letters.

The demands of the training task also determined the extent of the information encoded. Although it appears that the results of memorisers in Experiments 2 to 7 support feature and not exemplar or prototype learning accounts, the results actually show that fragment knowledge was sufficient to satisfy the demands of the training task. That is memorising two- and three-letter fragments was sufficient to be able to select the to-be-remembered letter string from a list of three presented in the second half of each training trial. It seems reasonable to assume that with appropriate instructions participants could extend what they learn to exemplars or prototypes.

Finally, different effects of fragment and rule knowledge do not indicate that we have separate similarity- and rule-based learning systems. Instead these results show that one processing system can account for the acquisition of different types of knowledge. These findings support the effects of selectively attending to some aspects of the structure of training items and not others (e.g., Goldstone. 1994).

Overall Summary

By using a biconditional grammar and orthogonal designs that set the contributions in classification performance of rule versus exemplar, rule versus fragment, and exemplar versus fragment knowledge in opposition to one another, it has been possible to identify the knowledge acquired from an artificial grammar more successfully than any prior AGL experiments.

As a result of being able to identify the form of knowledge used to classify test items it was possible to demonstrate that knowledge acquisition is not a matter of passively absorbing structural aspects of training items. Instead, the form of knowledge acquired is driven by the active processing applied to training items in order to meet the demands of the training task. Thus, the results can be explained by one episodic-processing system that encodes both the processing applied to training items and the specific aspects of training items processed. There was no evidence for dual implicit rule- and exemplar-based learning mechanisms.

Again, as a result of being able to identify the form of knowledge used to classify test items, it was also possible to demonstrate that episodic-processing knowledge is stored in a neutral rather than an implicit or explicit form. Consequently, this neutral form of knowledge can be expressed implicitly or explicitly at test depending on whether participants are made aware of the relationship between the knowledge they acquired during training and the demands of the test. Thus, given appropriate test instructions that refer to the knowledge acquired during training, all knowledge is explicit.

Appendix A: Experiment 1 Strings Statistics and Letter Strings

Table A1

The Means and Ranges for Item Characteristics within List and Training Group

Group	List	Item Characteristic		GS	GD	US	UD
Match/Edit	1	Anchor ACS	Mean	0.44	0.28	0.44	0.32
			Range	0.25-0.75	0.00-0.50	0.25-0.75	0.00-0.50
		Global ACS	Mean	2.33	2.37	2.33	2.33
			Range	1.44-2.90	0.00-0.50	1.82-2.85	1.74-2.95
		Whole-item Similarity	Mean	1.00	0.00	1.00	0.00
			Range	1.00-1.00	0.00-0.00	1.00-1.00	0.00-0.00
Match/Edit	2	Anchor ACS	Mean	0.46	0.21	0.51	0.33
			Range	0.25-1.00	0.00-0.50	0.25-0.75	0.00-0.75
		Global ACS	Mean	2.28	2.16	2.43	2.18
			Range	1.60-3.04	1.68-2.94	1.58-3.12	1.37-2.93
		Whole-item Similarity	Mean	1.00	0.00	1.00	0.00
			Range	1.00-1.00	0.00-0.00	1.00-1.00	0.00-0.00
Control		Anchor ACS	Mean	0.63	0.75	0.60	0.68
			Range	0.25-1.00	0.25-1.50	0.25-1.00	0.25-1.25
		Global ACS	Mean	4.93	4.74	4.86	4.73
			Range	3.54-5.79	3.77-5.50	3.71-5.33	4.01-5.43
		Whole-item Similarity	Mean	0.06	0.11	0.11	0.06
			Range	0.00-1.00	0.00-1.00	0.00-1.00	0.00-1.00

Note. GS = grammatical and similar, GD = grammatical and dissimilar, US = ungrammatical and similar, UD = ungrammatical and dissimilar, and ACS = associative chunk strength. Whole-item similarity indicates how many training items overlap with each training item on seven letters.

Table A2

Experiment 1 Match and Edit Group Training Strings

Match	Rehearsal	Distracter 1	Distracter 2
Edit	Correct String	Hypothesis test 1	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 thesis are reported using the three rules that D is paired with F, G with L, and K with X. In practice, in all seven experiments, 15 different sets of these three rules were used. Rule sets were matched across participants in the Match and Edit groups.

Table A3

Control Group Training Strings used in Experiments 1 and 5

Rehearsal String	Distracter 1	Distracter 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 A4

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: Experiments 2 and 6 String Statistics and Letter Strings

Table B1

The Means and Ranges for Item Characteristics

Item Characteristic		GH	GL	UH	UL
Anchor ACS	Mean	1.75	0.06	1.75	0.04
	Range	1.50-2.25	0.00-0.25	0.50-2.25	0.00-0.25
Global ACS	Mean	11.15	0.74	11.08	0.74
	Range	9.14-12.17	0.50-1.14	8.98-12.13	0.50-1.21

Note. ACS = associative chunk strength, GH = grammatical and high ACS, GL = grammatical and low ACS, UH = ungrammatical and high ACS, and UL = ungrammatical and low ACS.

Table B2

Training Strings for Experiments 2 and 6

Match	Rehearsal String	Distracter 1	Distracter 2
Edit	Correct String	Hypothesis test 1	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 B3

Classification Test Items used in Experiments 2, 5, 6, and 7

	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 Strings Statistics and Letter Strings

Table C1

The Means and Ranges for Item Characteristics

Item Characteristic		HIHA	LIHA	LILA
Anchor ACS	Mean	4.89	4.65	0.00
	Range	3.75-6.00	3.75-6.00	0.00-0.00
Global ACS	Mean	28.46	28.16	0.00
	Range	25.69-30.93	25.39-30.87	0.00-0.00
Whole-item Similarity	Mean	6.00	0.00	0.00
	Range	6.00-6.00	0.00-0.00	0.00-0.00

Note. ACS = associative chunk strength, HIHA = high item/high ACS similarity, LIHA = low item/ high ACS similarity, and LILA = low item/ low ACS similarity.

Table C2

Experiment 3 Training Strings

Group	Rehearsal String	Distracter 1	Distracter 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
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
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 C3

Experiment 3 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

Appendix D: Experiment 4 String Statistics and Letter Strings

Table D1

The Means and Ranges for Item Characteristics

Item Characteristic		HIHA	LIHA	LILA
Anchor ACS	Mean	8.42	8.08	0.00
	Range	6.00-10.50	7.25-10.25	0.00-0.00
Global ACS	Mean	41.05	39.83	0.00
	Range	38.68-45.82	38.24-42.37	0.00-0.00
Whole-item Similarity	Mean	24.00	5.17	0.00
	Range	24.00-24.00	3.00-6.00	0.00-0.00

Note. ACS = associative chunk strength, HIHA = high item/high ACS similarity, LIHA = low item/ high ACS similarity, and LILA = low item/ low ACS similarity.

Table D2

Experiment 4 Training Items

Group	Rehearsal String	Distracter 1	Distracter 2
1	KDGL.FGFF	XDGL.FGFF	KDKL.FFFF
1	XDGL.FGFF	DDGL.FGFF	KDGF.FGFF
1	DDKL.FGFF	DDKL.FGFG	DDGL.FGLF
1	DDGF.FGFF	DDGL.FGFF	XDGL.FGFF
1	DDGL.FFFF	DDGL.FGFF	DDGL.FGLF
1	DDGL.FGLF	DDKL.FGLF	DDKL.FGFF
1	DDGL.FGFG	DDGF.FGFG	DDKL.FGLG
1	DDGL.FGFX	DDGL.FGFF	DDFG.FGFF
1	KDKL.FGFF	KDKL.FFFF	XDGL.FGFF
1	KDGF.FGFF	KDGL.FGFF	KDGL.FGFG
1	XDGL.FGLF	XDGF.FGLF	XDKL.FGFF
1	XDGL.FGFG	XDGL.FFFG	XDGL.FFFF
1	DDKL.FGFX	DDGL.FGFX	DDGF.FGFX
1	DDKL.FFFF	DDKL.FGFF	DDKL.FGFX
1	DDGF.FGLF	DDGF.FGLG	DDKL.FGLF
1	DDGF.FGFG	DDGL.FGFG	DDGL.FGFX
1	DDGL.FFFX	DDGL.FGFX	DDKL.FGFX

1	DDGL.FFFG	DDGL.FGFG	DDGF.FGFG
1	DDGL.FGLG	DDGL.FGLF	XDGF.FGLG
1	DDKL.FGLF	DDGL.FGLF	DDGL.FGFF
1	DDGL.FFFG	DDKL.FFFG	DDGL.FGFX
1	XDGL.FGFG	XDGF.FGFG	XDGF.FGFX
1	DDGF.FGFX	DDGL.FGFX	DDKL.FGFX
1	KDGL.FGFX	KDGL.FGFG	KDGL.FFFF
2	DKDK.XXLG	LKDK.XXLG	DKLK.XXDG
2	LKDK.XXLG	LKDK.XXDG	DKLK.XXLG
2	GKLK.XXLG	GKDK.XXLG	GKDK.XXDG
2	GKDK.XXDG	GKLK.XXDG	GKLK.XXLG
2	GKDK.XXLK	GKDK.XXLF	LKLK.XXLK
2	GKDK.XXLF	GKDK.XXLK	GKDK.XXDG
2	DKLK.XXLG	DKLK.XXDG	DKDK.XXLF
2	DKDK.XXDG	DKLK.XXDG	DKLK.XXLG
2	DKDK.XXLK	DKDK.XXDK	DKDK.XXDG
2	LKLK.XXLG	LKLK.XXDG	LKDK.XXDG
2	LKDK.XXDG	LKDK.XXDK	LKLK.XXLG
2	LKDK.XXLF	LKLK.XXLF	LKDK.XXDK
2	GKLK.XXLK	GKDK.XXLK	LKDK.XXLK
2	GKLK.XXLF	GKLK.XXLK	GKLK.XXDG
2	GKLK.XXDG	GKDK.XXDG	DKDK.XXDG
2	DKDK.XXDG	DKDK.XXDK	DKLK.XXLG
2	GKLK.XXDG	DKLK.XXDG	GKLK.XXLF
2	GKDK.XXDK	GKLK.XXDK	GKLK.XXLK
2	DKDK.XXLK	DKDK.XXDK	GKLK.XXLK
2	LKDK.XXLK	DKDK.XXLF	LKLK.XXLF
2	GKLK.XXLK	GKDK.XXLK	GKDK.XXDK
2	LKDK.XXLF	GKLK.XXLF	LKLK.XXLK
2	DKDK.XXLF	LKDK.XXLF	DKLK.XXLK
2	LKDK.XXLF	LKDK.XXLK	LKLK.XXLG
3	GLKX.LFXD	XLKX.LFXD	GLKX.LFXL

3	XLKX.LFXD	GLKX.LFXD	KLKX.LKXD
3	KDKX.LFXD	KDKX.LKXD	KLFX.LFXD
3	KLFX.LFXD	KLFX.LFXL	KLKX.LKXD
3	KLKX.LKXD	KLFX.LKXD	KLFX.LFXD
3	KLKX.LFXL	KLKX.LFXD	KLKX.LKXD
3	KLKX.LFXX	KLFX.LFXX	GLFX.LFXX
3	GLKX.LFXX	GLKX.LKXX	GLFX.LKXX
3	GLKX.LKXD	KLKX.LKXD	GLFX.LKXX
3	GLFX.LFXD	GLFX.LKXD	KLKX.LFXD
3	XLKX.LFXX	GLKX.LFXX	KLFX.LFXX
3	XLKX.LKXD	XLKX.LFXD	GLKX.LFXD
3	XDKX.LFXD	KDKX.LFXD	XLKX.LKXD
3	KDKX.LFXX	KDKX.LKXX	KLKX.LKXX
3	KDKX.LFXL	KLKX.LFXL	KDKX.LKXD
3	KDKX.LKXD	KDKX.LFXD	KLFX.LKXD
3	GLFX.LFXD	GLKX.LFXD	KLKX.LFXD
3	KLFX.LKXD	KLFX.LFXD	KLKX.LFXD
3	KLFX.LFXL	KLFX.LFXD	KLFX.LKXX
3	GLKX.LKXD	KLKX.LKXD	GLKX.LFXL
3	KLFX.LKXD	KLFX.LFXD	KLFX.LFXL
3	KDKX.LFXL	KLKX.LFXL	KLFX.LFXL
3	KLFX.LFXL	KLKX.LFXL	KDKX.LFXD
3	KLKX.LKXX	KLFX.LKXX	KLFX.LKXD
4	DGFF.GKDK	DGFF.GKDG	FGFF.GKDD
4	LGFF.GKDK	DGFF.GKDK	LGLF.GKLK
4	FFFF.GKDK	FFFF.GKLK	FGFF.GKDD
4	FGLF.GKDK	FGFF.GKDK	FGLF.GKLK
4	FGFF.GKLK	FGFF.GKDK	FGFF.GKDD
4	FGFF.GKDD	FGFF.GKDG	FGFF.GKLK
4	FGFF.GKDG	FFFF.GKDG	LGFF.GKDD
4	DGFF.GKDD	FGFF.GKDD	DGLF.GKDG
4	DGFF.GKLK	DGFF.GKDK	LGLF.GKLK

4	LGLF.GKDK	LGLF.GKLK	DGFF.GKDK
4	LGFF.GKLK	LGLF.GKLK	LGFF.GKDG
4	FFFF.GKDG	FGFF.GKDG	DGFF.GKDG
4	FFFF.GKLK	FFFF.GKDK	FGFF.GKDK
4	DGLF.GKDK	DGLF.GKLK	FGFF.GKDK
4	FGLF.GKDD	FGFF.GKDD	DGLF.GKDK
4	FGLF.GKLK	FGFF.GKLK	FFFF.GKLK
4	LGFF.GKLK	LGFF.GKDK	LGFF.GKDD
4	FFFF.GKDD	FFFF.GKDG	FGLF.GKDD
4	LGFF.GKDD	FGFF.GKDD	LGFF.GKLK
4	FFFF.GKDG	FGFF.GKDG	FFFF.GKLK
4	FGLF.GKDG	FGLF.GKDD	FFFF.GKDG
4	FFFF.GKDD	FFFF.GKDK	FFFF.GKLK
4	FGLF.GKDG	FGFF.GKDG	FGFF.GKDK
4	FFFF.GKLK	FGFF.GKLK	FGLF.GKLK
5	GFXD.DDGL	FFXD.DDGL	FFXX.DDGL
5	FFXD.DDGL	GFXD.DDGL	FFXX.DDKL
5	LKXD.DDGL	LKXD.DDGK	FFXD.DDGL
5	LFXX.DDGL	LFXD.DDGL	LFXD.DDKL
5	LFXD.KDGL	LFXD.DDGL	LFXD.DDGF
5	LFXD.DDKL	LFXD.KDKL	LKXD.KDKL
5	LFXD.DDGK	LFXD.DDGF	FFXX.DDGK
5	LFXD.DDGF	FFXD.DDGF	LFXD.DDKL
5	GFXD.DDGF	LFXD.DDGF	GFXD.KDGL
5	GFXX.DDGL	GFXX.DDKL	LFXD.DDGL
5	FFXD.DDKL	FFXD.DDGL	FFXX.DDGL
5	FFXD.KDGL	LFXD.KDGL	FFXD.DDKL
5	LKXD.KDGL	LFXD.KDGL	FFXD.KDGL
5	LKXD.DDGK	LKXD.KDGK	LKXD.KDGL
5	LFXX.DDKL	LFXX.DDGL	LFXX.DDGK
5	FFXX.DDGL	LFXX.DDGL	LFXD.DDGL
5	GFXD.KDGL	FFXD.KDGL	FFXD.DDGL

5	LFXD.KDGF	LFXD.KDGL	LFXD.DDGL
5	FFXD.DDKL	FFXD.DDGL	FFXD.KDGL
5	GFXD.DDKL	FFXD.DDKL	GFXX.DDGL
5	LFXD.KDGK	LKXD.KDGK	GFXD.DDGK
5	LFXX.DDGK	LFXD.DDGK	LFXD.DDGF
5	LFXD.KDGF	LFXD.DDGF	LFXX.DDGF
5	LFXX.DDGF	LFXX.DDGL	LFXD.KDGF
6	KXLG.KLKX	FXLG.KLKX	KXDG.KLKX
6	FXLG.KLKX	KXLG.KLKX	XXDG.KLKX
6	XXDG.KLKX	XXDG.KDKX	XXLG.KDKX
6	XXLG.KDKX	XXLG.KLKX	XXDG.KLKX
6	XXLG.KLFX	XXDG.KLFX	XXLG.KDKX
6	XXLG.KLKD	KXLG.KLKD	XXDG.KLKD
6	XXLG.KLKL	XXLG.KLKX	XXLG.KLFX
6	KXDG.KLKX	KXDG.KLKL	KXDG.KDKL
6	KXLG.KDKX	KXDG.KDKX	FXLG.KLKX
6	FXLG.KDKX	FXDG.KDKX	FXDG.KLKX
6	FXLG.KLKD	FXLG.KLKX	FXLG.KDKX
6	XXDG.KLFX	XXDG.KLKX	FXLG.KLFX
6	XXDG.KLKL	XXLG.KLKL	XXDG.KDKX
6	XXLG.KDKL	XXDG.KDKL	XXDG.KLKL
6	XXDG.KDKX	XXDG.KLKX	XXLG.KDKL
6	KXLG.KLFX	KXLG.KLKX	KXDG.KLKX
6	FXLG.KLFX	KXLG.KLFX	FXLG.KLKD
6	XXDG.KLKD	XXLG.KLKD	XXDG.KDKL
6	XXLG.KDKD	XXLG.KDKL	XXDG.KLKD
6	KXLG.KLKL	KXLG.KDKL	XXLG.KLKX
6	XXDG.KLKL	KXDG.KLKL	XXDG.KDKD
6	FXDG.KLKX	FXLG.KLKX	KXLG.KLKX
6	XXLG.KDKL	XXLG.KLKL	XXDG.KLKL
6	XXDG.KLKX	XXDG.KLKX	XXLG.KLKX

Table D3

Experiment 4 Test Items

High Item Similarity High ACS Similarity	Low Item Similarity High ACS Similarity	Low Item Similarity Low ACS Similarity
DDGL.FGFF	DDKX.LFXL	DXGX.KFLD
GKDK.XXLG	KDKL.FGFX	GDLL.DFKG
KLKX.LFXD	KDKL.FGLF	XKGD.LXFK
FGFF.GKDK	FFXD.KDKX	LXGX.FKFD
LFXD.DDGL	LFXX.DDKX	FLDX.KGDF
XXLG.KLKX	XXDG.KDGL	KGXF.LDXK

Appendix E: Experiments 5 and 7 String Statistics and Letter Strings

Table E1

The Means and Ranges for Item Characteristics

Item Characteristic		GH	GL	UH	UL
Anchor ACS	Mean	1.75	0.06	1.75	0.04
	Range	1.50-2.25	0.00-0.25	0.50-2.25	0.00-0.25
Global ACS	Mean	11.15	0.74	11.08	0.74
	Range	9.14-12.17	0.50-1.14	8.98-12.13	0.50-1.21

Table E2

Training Strings for the Match Groups in Experiments 5 and 7

Rehearsal String	Distracter 1	Distracter 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

Appendix F: Instructions used in Experiments 1 to 7

Match Task Training Instructions for the Match (Experiments 1 to 7) and Control Groups (Experiments 1 and 5). In the first part of this experiment we will be looking at how good your short-term memory is for strings of letters like DFGK.FDLX. There will be 72 trials on each of which you will see a letter string. The string will stay on the screen for seven seconds during which time you should mentally rehearse it so that you can remember it. The screen will then go blank for two seconds, then a list of three strings will appear. You will be asked to type the number (1-3) of the string matching the one you are holding in memory.

Please try to be as accurate as possible, but if you really cannot remember then select the string that seems most familiar.

After you have typed in your choice, press RETURN. The program will tell you whether you were correct or not, and if you were wrong the correct choice will be displayed. Then you should press the X key to go on to the next trial.

Edit Training Instructions (Experiments 1 and 2). In this experiment you will be shown strings of letters such as DFGK.FDLX. Each string is made out of the six letters D, F, G, K, L and X. The computer knows a set of rules for putting letters into acceptable orders and your task is to try to work out what these rules are.

Each of the strings you see will have between two and four letters that violate the rules, in terms of the relationships between the letters. You will be asked to indicate whether you feel that each letter conforms to or violates the rules, by putting a Y below letters that you believe conform to the rules and an N below letters that you believe violate them. Please type a full stop under the central dot. When you have made your decision the program will tell you which letters actually violate the rules.

At the beginning of the experiment you will not know any of the rules so you will have to guess whether each letter is acceptable or not. Soon, though, you will begin to discover what the rules are.

Apply Rules Training Instructions (Experiment 6). In this experiment you will be shown 72 strings of letters such as DFGK.FDLX. Each string is made out of the six letters D, F, G, K, L and X and should conform to the following rules. Letter position 1 is linked to position 5, 2 is linked to 6, 3 with 7, and 4 with 8. Where one letter of each pair is D the other must be F, where one is G the other letter must be L, and where one letter is K the other letter must be X.

You will be shown flawed letter strings that violate the rules, in terms of the relationships between the letters. Your task is to put a Y below letters that you feel conform to the rules and an N below letters that you believe violate the rules. Please type a full stop under the central dot. When you have made your decision the program will tell you which letters actually violate the rules.

Classification Test Instructions (Experiments 1 to 7)

Initial Classification Instructions for the Match (Experiments 1 to 7) and Control Groups (Experiments 1 and 5). You may not be aware of it but the letter strings you were asked to memorise in the first part of this experiment were generated from a set of rules. Don't worry if you didn't notice this, as the task was designed to make it very difficult to notice the rules.

Initial Classification Instructions for the Edit Groups (Experiments 1 and 2). In the first part of the experiment you used a hypothesis testing strategy to try to learn the rules of the grammar. Do not worry if you do not feel completely confident in your understanding of the rules.

Initial Classification Instructions for the Apply Rules Group (Experiment 6). None.

Final Classification Instructions for all Groups (Experiments 1 to 7). In this final phase you will be asked to classify 144 new strings according to whether you think they conform to the rules of the strings you saw earlier or not.

Each string will be presented in turn and your task is to rate how well it conforms to the rules on a scale of 1 to 6. The points on the scale indicate the following: (1) certain the string obeys the rules; (2) fairly certain the string obeys the rules; (3) guess that the string obeys the rules; (4) guess that the string does not obey the rules; (5) fairly certain the string does not obey the rules; (6) certain the string does not obey the rules. You do not need to memorise these instructions, as they will be repeated on each screen

Note that in this phase of the experiment you will NOT be told whether your responses are correct or not.

Appendix G: Questionnaires (Experiments 1, 2 and 5)

Questionnaire Completed by the Experimenter for Experiments 1 and 2

Subject Number:..................... Group:......................... Date:.................................

Q1: Did you adopt any particular strategy, in the test phase, to determine if the strings conformed to the rules or not?
....
....
....

Q2: Did you notice any rules in the construction of the training strings?
....
....
....

Q3: There were rules linking letters in the first half of the string to corresponding letters in the second half of the string. Can you tell me any of these rules?
....
....
....

Q4: There were three rules that dictated which letters could appear in location five depending upon what letter was in location one. Can you tell me what those rules were?
....
....
....

Q5: Repeat the above question three times more for location pairs two and six, three and seven and four and eight.
....
....
....

Questionnaire Completed by each Participant in Experiment 5

Please answer the following questions as accurately as you can.
If your answers are 100% correct then you will win a £20 book voucher.
If more than one person is 100% correct, we will throw a coin to determine who gets the prize.

Question 1: The letter strings you saw in the experiment were constructed from the six letters D, F, G, L, K and X. Each string comprised eight letters with a dot in the middle and was created according to a set of rules. These rules governed which pairs of letters could occur in positions 1 and 5, 2 and 6, 3 and 7 and 4 and 8. For example if the letter in position 1 was D the rule may have been that position 5 must contain an X. Alternatively D may have been paired with F, G, K or L.

Can you tell me what these letter pair rules were in the strings you saw?
If you do not know the answer then please guess.

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 a L in position 7, what letter appeared in position 3?.....................................

If there was a X in position 8, what letter appeared in position 4?.....................................

Question 2: Please indicate on the line below how accurate you think you were in specifying the rules. The line represents a scale of 0 to 100, where 0 indicates that you feel you do not know any of the rules and 100 indicates that you are certain you have answered all six parts of question 1 correctly.

----

0 I do not know any rules	100 I am certain that all my answers are correct

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.

Allen, S. W., & Brooks, L. R. (1991). Specialising the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3-19.

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.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch and 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,* 316-23.

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

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

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

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

Dienes, Z., Altmann, G., & Gao, S. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science, 23,* 53-82.

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.

Dienes, Z., & Perner, J. (1996). Implicit knowledge in people and connectionist networks. In G. Underwood (Ed.), *Implicit Cognition.* Oxford University Press.

Dienes, Z., & Perner, J. (In press). A theory of implicit and explicit knowledge. *Behavioural and Brain Sciences.*

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.*

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1985). On consciousness in syntactical learning and judgement: A reply to Reber, Allen, and Regan. *Journal of Experimental Psychology: General, 114,* 25-32.

Goldstone, R. (1994). The role of similarity in categorisation: providing a groundwork. *Cognition, 52,* 125-127.

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

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 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. (Submitted). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming.

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,* 172-79.

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.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 149-157.

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 University 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 and Cognition, 13,* 469-475.

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.

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 relationships between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700-708.

Nosofsky, R. M., Clark, S., & Shin, H. (1989). Rules and exemplars in categorisation, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 282-304.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorisation and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247-255.

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,* 264-75.

Perruchet, P., & Pacteau, C. (1991). Synthetic grammar Learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 120,* 112-16.

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, 218-35.

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 re-evaluation. *Journal of Experimental Psychology: General, 125,* 123-138.

Redington, M., Friend, M., & Chater, N. (1996). Confidence judgements, performance, and practise, in artificial grammar learning. In G. W. Cottrell (Ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 649-654). Mahwah: Erlbaum.

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122,* 92-114.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human, Learning and Memory, 2,* 491-502.

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.

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

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,* 1504-10.

Whittlesea, B. W. A. (1987). Preservation of specific experiences in the representation of general knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 3-17.

Whittlesea, B. W. A. (1997a). The representation of general and particular knowledge. In K. Lamberts & D. Shanks (Eds.), *Studies in cognition: 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, 37,* (pp. 211 – 264). San Diego, CA: Academic Press.

Whittlesea, B. W. A., Brooks, L. R., & Westcott, C. (1994). After the learning is over: factors controlling the selective application of general and particular knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 259-274.

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., & Dorken, M. D. (1997). Implicit learning: Indirect, not unconscious. *Psychonomic Bulletin & Review, 4,* 63-67.

Whittlesea, B. W. A., & Williams, L. D. (In press). The source of feelings of familiarity: Fluency, discrepancy and coherence.

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 and Cognition, 26,* 402 – 420.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and association information: The contribution of recollection and familiarity. *Memory and 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.