Implicit Learning of Natural Language Syntax

Patrick Rebuschat

Research Centre for English & Applied Linguistics
University of Cambridge

and

Trinity Hall, Cambridge

Dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Cambridge.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. No parts of this dissertation have been submitted for any other qualification.

The dissertation does not exceed the regulation length, including footnotes, references and appendices but excluding the bibliography.

....

Para os meus avós, Marta e Sebastião
Für meine Großeltern, Helene und Gerhard

5 August 2009

Acknowledgements

Firstly, I would like to thank my supervisor, John Williams, for his guidance and support during the course of my studies at the University of Cambridge. My research has benefited immensely from John's expertise, advice and enthusiasm for everything implicit, and I could not have wished for a better supervisor.

I would also like to thank John Hawkins, Brechtje Post, Martin Rohrmeier and Ian Cross, all of whom have commented on my work on many occasions throughout the past three years. The quality of my PhD has improved considerably as a result of their thoughtful comments and suggestions.

I am grateful to Susan Rolfe and Barbara Jones for ensuring that the administrative side of my PhD ran smoothly and to my fellow PhD students for many interesting discussions and entertainment.

The results of the experiments reported in this dissertation were presented at meetings of the German Applied Linguistics Society (Koblenz, September 2005), the American Association of Applied Linguistics (Montréal, June 2006), the Cognitive Science Society (Vancouver, July 2006) and the German Cognitive Science Society (Saarbrücken, March 2007). The results were also presented at the Second Language Research Forum (New York, October 2005), the Cambridge-Essex Workshop on Language Acquisition (LALA IV: Cambridge, May 2006, and LALA VII: Cambridge, May 2008) and at the RCEAL Tuesday Colloquia Series (Cambridge, July 2007). I would like to thank the researchers who commented on my work at these events, in particular Zoltán Dienes, Nick Ellis and Robert DeKeyser, whose extensive comments and suggestions were particularly helpful.

Finally, I am most indebted to my wife, Michelle, for all her love and support. Signing up for a formal hall exchange to Homerton College in October 2004 certainly turned out to be the best decision in my life. I would also like to thank my parents, my brother and our families in Germany, Portugal and Canada for their continued emotional support.

The research carried out for this dissertation was financially supported by Trinity Hall, Cambridge, the Economic and Social Research Council (ESRC), the Cambridge European

Trust, the Fundação para a Ciência e a Tecnologia, the Research Centre for English & Applied Linguistics, and the Allen, Meek, & Reed Trust. The support of these institutions is gratefully acknowledged.

Abstract

The present dissertation focuses on the question of how humans acquire syntactic knowledge without intending to and without awareness of what they have learned. The aim is to apply the theoretical concepts and the methodological framework provided by implicit learning research to the investigation of language acquisition. The results of six experiments are reported. In terms of design, all experiments consisted of (i) a training phase, during which subjects were trained on a miniature linguistic system by means of different exposure conditions, (ii) an unexpected testing phase, during which learning and awareness were assessed, and (iii) a debriefing session. A semi-artificial grammar, which consisted of English words and German syntax, was employed to generate the stimulus material for experiments 1, 2, 3, 5 and 6; in the case of experiment 4, nonsense syllables were used instead of English words. The linguistic focus was on verb placement rules. Native speakers of English with no background in German (or any other V2 language) were recruited to take part in the experiments.

Participants in experiments 1-5 were exposed to the semi-artificial system under incidental learning conditions by means of different training tasks. In experiments 1 and 2, an auditory plausibility judgment task was used to expose participants to the stimulus sentences. In experiment 3, elicited imitations were used in addition to the plausibility judgment task. The training phase in experiment 4 consisted solely of elicited imitations, while training in experiment 5 consisted of a classification task which required participants to identify the syntactic structure of each stimulus item, followed by plausibility judgments. Participants in experiment 6, on the other hand, were exposed to the semi-artificial grammar under intentional learning conditions. These participants were told that the word order of the stimulus sentences was governed by a complex rule-system and instructed to discover syntactic rules. After training, participants in all six experiments took part in a testing phase which assessed whether learning took place and to what extent they became aware of the knowledge they had acquired. Grammaticality judgments were used as a measure of learning. Awareness was assessed by means of verbal reports, accuracy estimates, confidence ratings and source attributions. Control participants did not take part in the training phase.

The results of the experiments indicate that adult learners are able to acquire syntactic structures of a novel language under both incidental and intentional learning conditions, while processing sentences for meaning, without the benefit of corrective feedback and after short

exposure periods. That is, the findings demonstrate that the implicit learning of natural language is not restricted to infants and child learners. In addition, the experiments also show that subjects are able to transfer their knowledge to stimuli with the same underlying structure but new surface features. The measures of awareness further suggest that, in experiments 3 to 6 at least, learning resulted in both conscious and unconscious knowledge. While subjects did not become aware of all the information they have acquired, it was clear that higher levels of awareness were associated with improved performance.

The findings reported in this dissertation have several implications for our understanding of language acquisition and for future research. Firstly, while the precise form of the knowledge acquired in these experiments is unclear, the findings provided no evidence for rule learning in the vast majority of subjects. It suggests that subjects in these types of experiments (and perhaps in natural language acquisition) do not acquire linguistic rules. The results support Shanks (1995; Johnstone & Shanks, 2001), who argues against the possibility of implicit rule learning. Secondly, while adults can acquire knowledge implicitly, the work reported in this dissertation also demonstrates that adult syntactic learning results predominantly in a conscious (but largely unverbalizable) knowledge base. Finally, from a methodological perspective, the results of the experiments confirm that relying on verbal reports as a measure of awareness is not sufficient. The verbal reports collected at the end of the experiment were helpful in determining what aspects of the semi-artificial grammar subjects had consciously noticed. At the same time, verbal reports were clearly not sensitive enough to assess whether subjects were aware of the knowledge they had acquired. Confidence ratings and source attributions provided a very useful method for capturing low levels of awareness and to observe the conscious status of both structural and judgment knowledge. Future experiments on language acquisition would benefit from the introduction of this relatively simple, but effective way of assessing awareness.

Table of Contents

1. Introduction

How humans derive information from the environment implicitly, i.e. without the intention to learn and in absence of awareness of what was learned, is one of the central questions within the cognitive sciences (Shanks, 2005).[1] The process of implicit learning is assumed to underlie the human ability to extract knowledge from complex, rule-governed stimulus domains and, as such, appears to be an elementary and ubiquitous process of human cognition (Frensch & Rünger, 2003; Lewicki, Hill, & Czyzewska, 1992; Perruchet & Pacton, 2006; Reber, 1993). Everyday life offers many examples of implicit learning. Language acquisition (Berry & Dienes, 1993; Winter & Reber, 1994), socialization (Lewicki, 1986), music perception (Dienes & Longuet-Higgins, 2004; Rohrmeier, Rebuschat, & Cross, submitted) and many other learning episodes are widely thought to proceed in an implicit fashion.

In the case of language acquisition, for example, it seems clear that infants do not intentionally set out to acquire words and grammatical rules. Yet, in the 16 years after the onset of word learning, the average native speaker of English will acquire approximately 60,000 words – an average of 3,750 words per year and more than ten words per day (Bloom, 2001)![2] More importantly, all individuals will have developed a uniquely human and highly complex rule-system (*syntax*) that allows them to combine words into sentences. They will be competent users of this system – understanding, planning and producing complex utterances in less than a second – but nonetheless remain largely unaware of how the system actually works. For example, every native speaker of English will know intuitively that sentence (1) "John has a cup of coffee and a slice of buttered toast for breakfast." is in accordance with English word order rules, whereas sentence (2) "John a has cup of and coffee a of slice toast for buttered breakfast." does not. However, most individuals would not be able to accurately describe the rules that generate grammatical sentences in English or provide an account for why (1) is an acceptable sentence whereas (2) is not.

The present dissertation focuses on the question of how humans acquire syntactic knowledge without intending to and without awareness of what they have learned. The aim is to apply the

[1] While implicit learning is frequently characterized as learning that occurs incidentally and without awareness of the learning outcome, there are, in fact, multiple meanings associated with the concept of implicit learning. Frensch (1998) provides a useful discussion on the definition of implicit learning.

[2] The rapid, incidental acquisition of words has been extensively studied within developmental psychology (see Bloom, 2000, 2001, for overviews). Bloom & Markson (1998) include a brief commentary on word learning rates.

theoretical concepts and the methodological framework provided by implicit learning research to the investigation of language acquisition. The first chapter will provide a brief overview of implicit learning as an area of research and review a selection of studies that investigated the processes underlying incidental language learning. The subsequent chapters (2-7) then report and discuss the results of six experiments that investigated the acquisition of syntactic knowledge under different learning conditions. A general conclusion (chapter 8) summarizes the key findings of these experiments and outlines directions for future research.

1.1. What is implicit learning?

The term "implicit learning" was first employed by Reber (1967) to differentiate the process of incidental, unaware learning from that of "explicit learning", a more conscious process where the individual deliberately forms and tests hypotheses about the stimulus domain (see Reber, 1993, p. 10). Implicit learning as a field of research began with experiments conducted by Reber and his associates (Reber, 1967, 1969, 1976; Reber & Allan, 1978; Reber & Lewis, 1977; Reber & Millward, 1968, 1971) and developed into a major topic in the landscape of experimental psychology (see Berry & Dienes, 1993; Cleeremans et al., 1998; Frensch & Rünger, 2003; Perruchet, 2008; Reber, 1989, 1993; Seger, 1994; and Shanks, 2005, for general overviews).[3] Implicit learning experiments usually involve three components, namely (i) exposure to a rule-governed stimulus domain under incidental learning conditions, (ii) a measure of learning, and (iii) a measure of awareness (Cleeremans et al., 1998). There are several paradigms that follow this conceptual design (e.g., sequence learning, control of complex systems, and probability learning), but the most influential paradigm is arguably Artificial Grammar Learning (Dienes, in press; Pothos, 2007).

In Artificial Grammar Learning (AGL), participants are first exposed under incidental learning conditions to a stimulus domain that consists, on the surface level, of letter, syllable or tone sequences generated by finite-state grammars (Chomsky, 1956; Chomsky & Miller,

[3] Frensch (1998) found that the number of articles in the PsycLIT database that contained the term "implicit learning" in the title increased substantially since the mid-1980s, and Shanks (2005) reports a similar pattern in his analysis of the Science and Social Sciences Citation Indices. Reber (1993, Chapter 1) includes a brief historical overview of early research (1860-1950) on unconscious learning and an interesting account of the scientific climate in which implicit learning was established as an area of research within cognitive psychology. For a review of studies on unconscious cognition conducted under different research traditions (implicit memory, subliminal perception, automatic processing), see Reber (1993) and, especially, Dienes & Berry (1993, chapter 1).

1958) such as the ones reproduced in figures 1 and 2.[4] Stimulus sequences (e.g., TPTS, VXVPS, XMVTRXM, VVTTRMTM) are generated by following any path of arrows leading from the initial state to the final state of the grammar. After the exposure phase, participants are then tested to establish whether learning took place and to determine to what extent they are conscious of the knowledge they have acquired. Typically, learning is measured by observing the performance on a classification task that requires participants to distinguish sequences that were generated by means of the same rule-system from sequences that were not (e.g., Reber, 1967, 1969). Awareness, on the other hand, has traditionally been assessed by prompting subjects, in post-hoc questionnaires, to verbalize any rules they might have noticed while taking part in the experiment (e.g., Allen & Reber, 1980; Reber & Allen, 1978; Reber & Lewis, 1977). The results commonly indicate that experimental subjects perform significantly above chance on the classification task (at around 65% accuracy), despite remaining unable to describe the rules of the grammar in verbal reports. Untrained controls tend to perform at chance level.

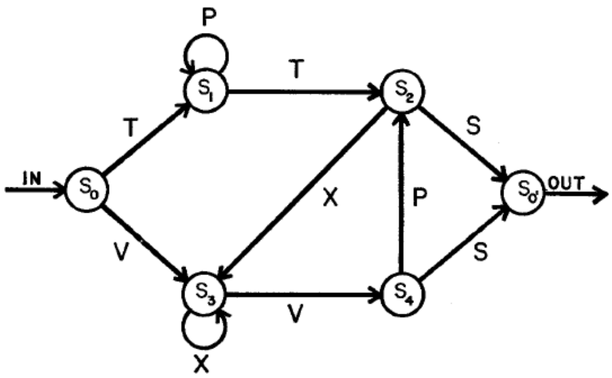

Figure 1. The finite-state grammar used by Reber (1967).

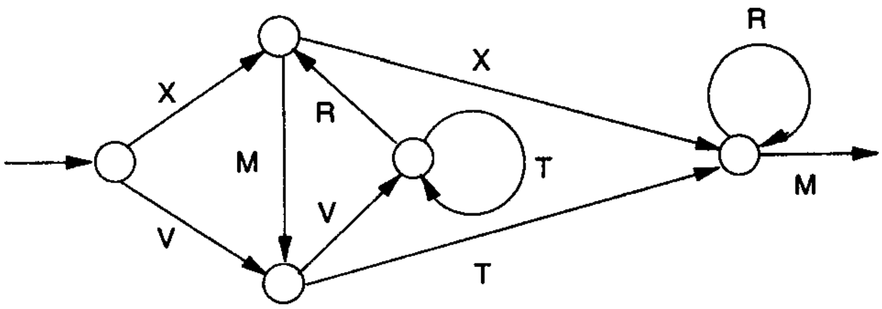

Figure 2. The finite-state grammar used by Dienes, Altmann, Kwan, & Goode (1995, Experiments 1-4).

[4] The term “finite-state grammar” was first used by Chomsky (1956, p. 114). Chomsky & Miller (1958) provide a description of the formal properties of finite-state languages.

In his seminal study, Reber (1967) exposed subjects to letter sequences by means of a memorization task. In experiment 1, subjects were presented with letter sequences and asked to memorize them. One group of subjects was given sequences that were generated by means of a finite-state grammar, while the other group received randomly constructed sequences. The results showed that grammatical letter sequences were learned more rapidly than random letter sequences. Reber suggested that this memorization advantage reflected increasing sensitivity to grammatical structure in the former group. Experiment 2 consisted of two parts. In the first part of the experiment (the learning phase), subjects were presented with letter sequences that had been generated by means of a finite-state grammar and simply instructed to memorize the sequences. Subjects were not told that the letter arrangement followed the rules of a grammar. In the second part of the experiment (the testing phase), subjects were informed that the previous letter sequences had been formed by a set of grammatical rules. Subjects were then given new letter sequences, only half of which followed the same grammar, and instructed to judge whether the sequences were grammatical or not. (There were no control subjects.) The results showed that participants judged 78.9% of all letter sequences correctly, which indicated that simple memorization of grammatical strings was sufficient for subjects to derive information about their construction and that this information could be applied in a novel task. Reber suggested that, during the learning phase, subjects built an underlying representation that mirrored the structure intrinsic to the stimulus domain. Reber's initial assumption was that subjects learned the syntactic structure of the finite-state grammar as such (Dienes, in press; Reber, 1989).

The past decades have resulted in a relative consensus on several characteristics of implicit learning (Berry & Dienes, 1993; Cleeremans et al., 1998; Dienes & Berry, 1997). For example, Reber's (1967) basic finding that subjects can exploit the structure inherent to the stimulus environment incidentally has been frequently replicated and appears very robust (e.g., Dienes, Altmann, Gao, & Goode, 1995; Dienes & Scott, 2005; Mathews, Buss, Stanley, Blanchard-Fields, Cho, & Druhan, 1989; Reber, 1969; Tunney & Shanks, 2003). It is widely accepted that implicit learning gives rise to a sense of intuition, i.e. "people do not feel that they actively work out the answer" but rather make "particular responses because they „feel' right" (Berry & Dienes, 1993, p. 14). Several studies also show implicitly acquired knowledge to be more robust in the face of neurological disorder (e.g., Knowlton, Ramus, & Squire, 1992; Squire & Frambach, 1990; though see Channon, Shanks, Johnstone, Vakili, Chin, & Sinclair, 2002) and in the face of time, i.e. it appears longer lasting than explicit knowledge (Allen & Reber, 1980). At the same time, however, considerable controversy surrounds several key questions, including the following two: (i) How is implicitly acquired knowledge represented? and (ii) Is implicitly acquired knowledge unconscious? (Berry, 1996; Pothos, 2007).

1.1.1. How is implicitly acquired knowledge represented?

Reber (1969) investigated the effect of changing the lexicon or the grammar of the artificial language during the learning episode. The study consisted of two parts. The first part was a learning task that required subjects to memorize strings generated by a finite-state grammar. The second part consisted of a transfer task. Subjects were asked to continue memorizing letter sequences, but the letter sequences were modified without warning. For one group of subjects, the sequences were made up of the same letters as those in the first part, but a different finite-state grammar was used to generate them (old lexicon, new rules). For another group of subjects, the rules were the same, but the actual letters used to represent the grammar were changed (old rules, new lexicon).[5] Reber found that changing the rules had a disruptive effect on subjects' memorization performance, but changing the lexicon had no detrimental effect. The memorization advantage, observed in Reber (1967, Experiment 1), was maintained as long as the rules remained the same. That is, subjects could "transfer" the knowledge acquired while memorizing one set of letter sequences to the memorization of a different set of letter sequences, even though both sets featured different lexicons. Reber concluded that implicit learning results in an abstract representation of the structure displayed in the stimulus environment.

The transfer effect has been frequently replicated (e.g., Altmann, Dienes, & Goode, 1995; Brooks & Vokey, 1991; Gómez & Schvanefeldt, 1994; Gómez, 1997; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Manza & Reber, 1997; Mathews et al., 1989; Tunney & Altmann, 2001; Whittlesea & Dorken, 1993). Several studies have shown transfer to be limited to certain conditions (e.g., Berry & Broadbent, 1984, 1988; Berry, 1991), but its existence seems largely noncontroversial. What remains contentious, however, is how to explain the underlying process. Reber's (1967, 1969) initial assumption was that subjects acquire

[5] Two control groups were also run. In the first group of controls, both the grammar and the lexicon were changed at the beginning of the second session (new lexicon, new rules). In the second control group, nothing was changed (old lexicon, old rules).

abstract, rule-based knowledge during AGL experiments (see Reber, 1989, p. 114).[6] According to the abstractionist account, the mental representation established during AGL consists of a symbolic structure which is independent of the original surface form of the training materials (Manza & Reber, 1997; Reber, 1989; Reber & Lewis, 1977). The transfer effect is explained by assuming that subjects learn rules which capture the structure of the training stimuli and then use this knowledge when judging whether test stimuli follow the same rules or not (e.g., Knowlton & Squire, 1996; Marcus et al., 1999: Manza & Reber, 1997). Manza & Reber (1997, p. 75) summarize the abstractionist position as follows:

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

Support for the abstractionist view comes from a study conducted by Marcus et al. (1999). In the study, seven-month old infants were habituated to a set of three-word sequences (*ga ti ga*, *li na na*, etc.) constructed from a finite-state grammar. After two minutes of exposure, the infants were then tested on novel three-word sequences that were composed of a different lexicon (*wo fe wo*, *de ko ko*, etc.) but generated by the same grammar. The test sequences varied as to whether they were consistent or inconsistent with the grammar of the training sentences. Because none of the test syllables appeared in the training phase, infants could not distinguish the test sequences based on transitional probabilities, and because the test sequences were of the same length and computer-generated, infants could not distinguish them based on cues such as syllable frequency or prosody. Infants were tested with an adapted version of Jusczyk & Aslin's (1995) familiarization preference procedure. According to Marcus et al., if infants could abstract the underlying structure and generalize (i.e., transfer) it to novel words, they should attend longer during the presentation of ungrammatical items than during presentation of grammatical items. Marcus et al. found this to be the case in 15 out of 16 infants, i.e. most infants were able to discriminate between grammatical and ungrammatical syllable sequences, even though they were unlikely to be aware of the rule

[6] Reber (1989, p. 58) later changed his view, arguing that the "functional representation of the mental content of a finite-state grammar should be regarded as an ordered set of bigrams or larger chunks [...] and not as a formal Markovian system."

structure of the stimuli presented to them. Marcus et al. (1999, p. 79) concluded that infants derive "abstract algebra-like rules that represent relationships between placeholders (variables)" and that they were able to generalize these rules to novel instances (see also Gómez & Gerken, 1999; Saffran, Pollak, Seibel, & Shkolnik, 2007, replicated Marcus et al.'s results with a lexicon consisting of nonlinguistic stimuli).

The notion that implicit learning results in abstract, rule-based knowledge has been frequently challenged. One of the earliest criticisms was proposed by Brooks (1978). According to the exemplar-based account, the knowledge acquired during AGL consists of relatively unprocessed representations of training strings (Brooks, 1978; Brooks & Vokey, 1991; Vokey & Brooks, 1992, 1994; Vokey & Higham, 2005). In the training phase, subjects do not abstract rules but merely store stimuli as whole items in memory. In the testing phase, subjects then classify test strings on the basis of similarity to the encoded study instances. If a test string resembles a previously encountered item, it will be endorsed; if not, it will be rejected. Thus, in a standard AGL experiment, subjects should accept a given test string (e.g. DXCCCA) if it resembles a training string (e.g. DBCCCA). In the case of transfer to changed letter-set sequences, subjects are thought to rely on a process of "abstract analogy" (Brooks & Vokey, 1991). The test sequence DBCCCA, for example, can be seen as analogous to the training string PJRRRT. Analogous test items are endorsed, the remaining items are rejected. The exemplar-based view thus not only opposes the notion that subjects learn rules, but it also suggests that the process of abstraction occurs during the testing phase and not during training.

Several studies support the view that much of transfer can be explained by abstract similarity between training and test sequences (Brooks & Vokey, 1991; Gómez, Gerken, & Schvanefeldt, 2000; Tunney & Altmann, 1999). However, the notion that subjects store exemplars has been frequently criticized (e.g., Perruchet, 1994; Perruchet & Pacteau, 1990; Johnstone & Shanks, 2001; Shanks, Johnstone, & Staggs, 1997). Fragment-based accounts (Dulany, Carlson, & Dewey, 1984; Perruchet & Pacton, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990) suggest that subjects accumulate knowledge of fragments or "chunks", such as bigrams or trigrams, weighted by their frequencies of occurrence over the training set. Perruchet & Pacteau (1990, Experiment 1), for example, compared the performance of participants trained on grammatical letter sequences with those trained on the bigrams used to construct the grammatical training strings. One group was asked to memorize sequences generated by a finite-state grammar (e.g., MTVRXV), while the other group was asked to memorize the constituent bigrams of the same strings (e.g., MT, TV, VR, RX, XV). In the testing phase, both groups were asked to judge novel letter sequences, half of which were generated by the same finite-state grammar. Perruchet and Pacteau found that both groups performed significantly above-chance in the classification task, which suggested that above-chance performance in AGL experiments does not necessarily imply the acquisition of abstract rule knowledge (see Mathews, 1990, 1991, and Reber, 1990, for commentaries on Perruchet & Pacteau, 1990; see Perruchet & Pacteau, 1991, for a reply).

The fragment-based view further suggests that, in the testing phase, subjects classify novel test strings as grammatical to the extent to which a test string contains fragments that were present in the training strings. That is, familiarity rather than grammaticality might control performance. Johnstone & Shanks (2001) examined this assumption in a series of AGL experiments. In the study, subjects were exposed to training strings generated by a biconditional grammar (Mathews et al., 1989) and then tested on novel strings that where balanced for grammaticality and familiarity. Subjects classified four types of test items: familiar grammatical, familiar ungrammatical, unfamiliar grammatical, and unfamiliar ungrammatical. Their results showed that subjects displayed no tendency to endorse grammatical items more than ungrammatical ones; instead, they tended to endorse items as grammatical if they were composed of familiar bigrams, regardless of their grammaticality (see also Channon et al., 2002; Kinder & Assmann, 2000; Kinder, Shanks, Cock, & Tunney, 2003). Johnstone and Shanks argue that these findings cast doubt over Reber's (1967, 1969) initial results. According to Shanks (2005, p. 209), Reber's results were "an artefact caused by the fact that his grammatical items tended to be made of familiar bigrams while his ungrammatical ones tended to be made of unfamiliar ones." A number of computational models, most notably Servan-Schreiber & Anderson's (1990) Competitive Chunking model and Perruchet & Vinter's (1998) PARSER model, provide further support for fragment-based views of implicit learning (see Cleeremans & Dienes, 2008, for an overview of computational models of implicit learning).

An important contribution to the debate on the form of implicitly acquired knowledge was made by Whittlesea and his associates (Whittlesea, 1997; Whittlesea & Dorken, 1993; Wright & Whittlesea, 1998). In contrast to the previous proposals, their episodic-processing account does not focus solely on the stimulus-driven acquisition of structural knowledge (i.e., rules, exemplars, or fragments) but emphasizes the importance of the encoding episode as a primary mechanism for learning. According to this view, knowledge of both structure and processing is encoded during training. What type of structural knowledge will be acquired during the training phase is highly dependent on the instructions that subjects receive (Wright & Whittlesea, 1998). In addition, the episodic-processing account proposes that subjects can apply the same knowledge implicitly or explicitly, depending on whether they understand the association between processing fluency and the knowledge they acquired by processing training items in particular ways (Whittlesea & Williams, 1998: Discrepancy-Attribution Hypothesis). Whittlesea & Dorken (1993, p. 230) summarize the central claims as follows:

> Briefly, the account assumes that different tasks performed on the same stimuli will cue different operations (encoding variability), that memory will preserve the specific experience of the stimuli given by those operations (specific and distributed representation), and that operations conducted later receive support to the extent that they can make use of products of operations performed earlier (transfer-appropriate processing). This principle is meant to apply whether the operations are performed deliberately or incidentally, and denies any additional, default memory operations conducted without being cued by the demands of the task.

Evidence that knowledge of both structure and processing is encoded during training was provided by Whittlesea & Dorken (1993). In experiment 1, for example, subjects were instructed to either pronounce or spell letter sequences (e.g., ELFENAD, ONROLET, etc.) that were generated by two finite-state systems: grammar A generated sequences that were to be pronounced; grammar B generated sequences that were to be spelled. In the testing phase, subjects were told about the two grammars and presented with novel sequences. Half of the test sequences followed grammar A and the other half grammar B. Subjects were instructed to classify sequences according to the grammar that was used to construct them. However, prior to each classification decision, subjects were also required to spell or pronounce the test item. Subjects were informed that, in the testing phase, this activity was completely unrelated to whether the item conformed to the rules of grammar A or grammar B. In fact, half the letter sequences that were generated by grammar A would have to be spelled, and vice versa. Whittlesea and Dorken found that test performance was only above chance when the training and test processes were identical. Under matched conditions, subjects were able to distinguish grammar A strings from grammar B strings reliably. However, when items were spelled in training and pronounced at test or pronounced in training and spelled at test, subjects classified at chance levels. Whittlesea and Dorken concluded that, in addition to structural aspects of the stimuli, the knowledge acquired during training included details of processing and that classification performance was successful to the extent that the test instructions cued prior processing episodes.

One of the predictions of the episodic-processing account is that structural aspects of training stimuli (e.g., rules, exemplar, fragments) will only be encoded to the extent that such information enables subjects to meet the demands imposed by the training instructions. Successful test performance then depends on test instructions reinstating the processing context experienced during the training phase. Johnstone & Shanks (2001, Experiments 1 and 2) evaluated this assumption by comparing the incidental and the intentional acquisition of a biconditional grammar. One group of subjects was exposed to the grammar under incidental learning conditions by means of a simple memorization task. The other group was instructed to discover the rules that determined the letter arrangements. After the exposure, subjects were tested by means of a standard grammaticality judgment task. Johnstone and Shanks found that incidental learners ("memorizers") performed at chance levels, while intentional learners ("hypothesis-testers") classified at near-perfect levels. Importantly, memorizers classified on the basis of episodic knowledge of fragments, in contrast to hypothesis-testers, who classified on the basis of rule knowledge. The results thus suggest that incidental learning results in knowledge of fragments, while rule abstraction depends on active, conscious efforts to identify the rules of the grammar (see also Shanks, Johnstone, & Staggs, 1997). In this sense, there is no implicit learning of rules, as originally claimed by Reber (1969, 1989).

In sum, the studies reviewed in this section suggest that the knowledge acquired during standard implicit learning experiments has the following characteristics. Firstly, there is extensive evidence that, in most AGL experiments, subjects acquire fragmentary repesentations of the training stimuli, e.g. bigrams and trigrams (Johnstone & Shanks, 2001; Perruchet & Pacteau, 1990; Redington & Chater, 1996; Servan-Schreiber & Anderson, 1990). This is also the view currently expressed by Reber (1993, p. 58). Secondly, if rule knowledge can be acquired under incidental learning conditions, this would not correspond to formal Markovian rules. As Reber (1993) suggests, the rule knowledge acquired during AGL experiments is probably best described as a partial but representative subset of the rules employed to generate the stimuli, e.g. micro-rules such as “The X cannot occur at the beginning.”, “The T can be immediately repeated after an R.”, etc. (Dulany, Carlson, & Dewey, 1984; Reber, 1993). Thirdly, implicitly acquired knowledge is more surface-dependent than originally anticipated by Reber (1967, 1969). Several studies have shown that transfer subjects often perform worse than non-transfer subjects, which indicates that knowledge is at least partially tied to specific perceptual features of the training stimuli (Altmann et al., 1995; Dienes & Altmann, 1997; Manza & Reber, 1997). Also, the process of abstraction might take place in the testing phase, when old and novel strings are compared, and not during training (Brooks & Vokey, 1991; Redington & Chater, 1996). Implicitly acquired knowledge could be seen as abstract in the sense that sequences such as ABACCA and XYXZZY could be similarly represented as 121331 and not as deeper abstract representations (e.g. in the form of rewrite rules such as A → B|C, B → A, etc.). Finally, the representation of implicitly acquired knowlege is likely to depend on the processing demands of the training phase. Whittlesea & Dorken (1993) suggest that, in principle, subjects can acquire rules, exemplars, or fragments while performing on the training tasks. What type of structural knowledge will be acquired, depends on the particular demands of the encoding episode (though see Johnstone & Shanks, 2001, who argue against the possibility of implicit rule learning).

1.1.2. Is implicitly acquired knowledge unconscious?

It is widely accepted that the knowledge acquired during standard implicit learning experiments is the result of an incidental learning process. However, the question of whether this knowledge is actually “implicit” is highly controversial (Berry, 1996; Stadler & Roediger, 1997). The case for unconscious learning depends on both our definition of awareness and on the validity of tests used to assess unaware learning. Learning is often described as unconscious in different ways, depending on whether the focus is on the learning process, the retrieval process, or on the product of learning itself (Cleeremans et al., 1998; Frensch, 1998; Frensch & Rünger, 2003; see Dienes & Berry, 1997, for a discussion of criteria of implicitness). In this discussion, learning will be presumed unconscious if participants are unaware of what they have learned, i.e. implicit learning is defined in terms of the product of learning rather than the properties of the learning process. Proposals for measuring awareness in this sense include subjective measures, such as verbal reports and confidence ratings, as well as objective measures, such as fragment completion and cued generation tasks. Dienes & Berry (1997), Gaillard, Vandenberghe, Destrebecqz, & Cleeremans (2006), and Seth, Dienes, Cleeremans, Overgaard, & Pessoa (2008) provide overviews of measures of awareness. For alternative definitions on unconscious cognition, see Jacoby (1991), Merikle & Joordens (1997), and Greenwald (1992).[7]

Verbal reports. One of the most common procedures for assessing awareness is to prompt subjects to verbalize any rules they might have noticed while performing on the experimental tasks (e.g., Abrams & Reber, 1988; Allen & Reber, 1980; Berry & Broadbent, 1984; Dienes, Broadbent, & Berry, 1991; Lewicki, Hill, & Bizot, 1988; Reber & Allen, 1978; Reber & Lewis, 1977; see Adams, 1957, Ericsson & Simon, 1980, 1984; Nisbett & Wilson, 1977; and Payne, 1994, for reviews). Learning is thought to be unconscious when subjects show an effect of training (e.g. above chance classification performance), despite being unable to describe the knowledge they have acquired. Several studies have provided evidence for a dissociation between task performance and verbalizable knowledge. In an early control-task study, Broadbent (1977) required subjects to control the number of passengers using buses and the number of empty car parking spaces in the transport system of a fictitious city.[8] Subjects had to reach a target value by varying the time interval between buses and the car parking fee. Afterwards, they completed a multiple-choice questionnaire to determine the extent of verbalizable knowledge. Broadbent found that subjects improved in the ability to control the transport system but not in the ability to describe the relationships within the system. Broadbent & Aston (1978), Broadbent, Fitzgerald, & Broadbent (1986), and Berry & Broadbent (1984) reported similar dissociations. In Berry & Broadbent's (1984) study, for example, subjects were required to control either a sugar production plant or a simulated

---

[7] In implicit learning research, the view that knowledge can be unconscious when it is applied without conscious intention has gained popularity. See Jacoby (1991) and Jacoby, Ste-Maries, & Toth (1993) for an introduction to the Process Dissociation Procedure as well as Curran (2001), Destrebecqz & Cleeremans (2001), Fu, Fu, & Dienes (2007), and Wilkinson & Shanks (2004) for applications to implicit learning research. Greenwald (1992) remarks that the term "unconscious" can also be used as a synonym for "unattended". In this sense, stimuli that are processed outside focal attention are unconscious.

[8] In control-task studies (e.g., Broadbent, 1977; Broadbent & Aston, 1978; Berry & Broadbent, 1984, 1988; Broadbent, Fitzgerald, & Broadbent, 1986), subjects learn to control a complex, interactive system, e.g. an imaginary factory or a simulated person, incidentally. Subjects may be told, for example, that they are managing a sugar production plant and asked to reach and maintain specified production levels. The control is implemented by making adjustments in one or more production factors, e.g. wages, labor, number of workers, etc. After each interaction, the resulting state of the system is produced on the screen. Subjects then continue their manipulations until reaching a satisfactory level. Subjects are not told, however, that the state of the system is calculated by means of an equation that relates input and output variables. Typically, production-control experiments show that subjects are able to achieve a good level of control of the system even though they remain unable to describe how the system works.

person. Under both conditions, practice improved the ability to control these systems but had no effect on the ability to verbalize knowledge. Interestingly, providing a different group of subjects with instructions on how to control the complex systems significantly improved their ability to answer questions about the systems but had no effect on control task performance. The findings of control-task studies suggest that task performance and verbalization draw on separate cognitive systems, which would explain why the acquired knowledge might be inaccessible to free recall (Berry & Dienes, 1993).

The view that learning is unconscious when subjects fail to express the knowledge they have acquired has been criticized for a variety of reasons (see Shanks & St. John, 1994, for an overview). To begin with, the dissociation between acquired knowledge and its verbalizability is not as clear as anticipated (Dienes & Fahey, 1995; McGeorge & Burton, 1989; Stanley, Mathews, Buss, & Kotler-Cope, 1989). Stanley et al. (1989), for example, reported that, given sufficient practice, subjects are often capable of providing verbal descriptions of the knowledge they acquired. Moreover, verbal reports constitute a relatively insensitive and incomplete measure of awareness (Dienes & Berry, 1997; Ericsson & Simon, 1980; Reingold & Merikle, 1988: Exhaustiveness Assumption; Shanks & St. John, 1994: Sensitivity Criterion). Berry & Dienes (1993) observed that failure to verbalize knowledge could simply reflect the problem of having to retrieve large amounts of low confidence knowledge. When subjects are given the option of not responding during verbal reports, then conscious knowledge, though present, may simply not be detected. Erdelyi & Becker (1974, cited in Dienes & Berry, 1997) also report that subjects who are unable to verbalize knowledge on their first attempt are often able to do so when prompted again at a later point in time. Verbal reports might thus not be sensitive enough to capture all of the relevant conscious knowledge.

Finally, it is often unclear whether the information assessed by verbal reports is, in fact, responsible for the performance on the measure of learning (Shanks & St. John, 1994: Information Criterion). Experimenters usually employ verbal reports to determine whether subjects know the rules that generated the stimulus sequences. However, several authors have shown that above-chance classification performance in AGL tasks does not necessarily require knowledge of the rules of the underlying grammar. Knowledge of fragments or chunks seems to be sufficient (Perruchet, Gallego, & Savy, 1990; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). Others have pointed out that subjects build “correlated grammars”, i.e. sets of conscious rules that are of limited scope and validity but which still suffice to classify test strings significantly above chance (Dulany, 1962; Dulany, Carlson, & Dewey, 1984). Subjects often omit to report any knowledge because verbal reports fail to elicit the actual knowledge used in the classification task. In cases such as these, assuming that learning was unconscious would be erroneous.

Objective tests. Several authors have advocated the use of objective tests as more exhaustive measures of awareness (e.g., Eriksen, 1960; Holender, 1986; Reed & Johnson, 1994; Stadler, 1989; Willingham, Nissen, & Bullemer, 1989). An "objective test", as described by St. John & Shanks (1997, p. 168), is a forced-choice test that encourages subjects to access all relevant conscious knowledge in order to perform. If subjects fail to display knowledge on an objective test while showing evidence of learning on a different task, then knowledge is assumed to be unconscious. Objective tests have been employed in both AGL and sequence learning.[9] In the testing phase of a standard AGL experiment, Dulany, Carlson, & Dewey (1984) required subjects to classify letter sequences according to their grammaticality and to underline those items that made each sequence either grammatical or ungrammatical. They found that performance on the objective measure of awareness (the underlining task) correlated highly with subjects' performance on the measure of learning (the classification task). Perruchet & Pacteau (1990, Experiment 3) used a recognition task, a standard measure of explicit memory (Buchner & Wippich, 2000), in order to assess whether subjects had acquired conscious knowledge of bigrams. They reported that performance on the recognition task correlated highly with performance on the grammaticality judgment task.

In the case of sequence learning, Perruchet & Amorim (1992) used a free generation task in addition to a recognition task as measures of awareness. In the former task, subjects were instructed to generate an entire sequence of trials. In the latter task, participants were presented with a sequence of elements for a number of trials and asked to decide whether or not they had seen that particular sequence fragment before. Cleeremans & McClelland (1991, Experiment 2) and Jiménez, Méndez, & Cleeremans (1996) employed a cued generation task

[9] In sequence learning (e.g., Nissen & Bullemer, 1987; Lewicki, Czyweska, & Hoffman, 1987; Cleeremans & McClelland, 1991), subjects are usually presented with visual elements of a sequentially structured sequence and required to respond to each stimulus. Lewicki's (1986) matrix scanning task, for example, requires subjects to scan a complex matrix of numbers arranged in four quadrants on a computer screen. The matrix contains a single target number whose location is changed from trial to trial. The task is to identify the quadrant in which the target is located as quickly as possible by pressing a corresponding key. Subjects are not told that the position of the target stimulus is determined by a rule-system. The typical finding is that, after thousands of trials, subjects become faster in locating the target, despite being unable to make overt predictions (based on the sequence rule) and unable to articulate the rules that determine the position of the target stimuli (Lewicki et al., 1988).

in order to determine whether subjects had acquired conscious knowledge. The task required subjects to predict the next element of a sequence. These three studies assessed learning by measuring how quickly subjects responded to stimuli presented in the context of Nissen & Bullemer's (1987) serial reaction time task. Cleeremans & McClelland (1991), Jiménez et al. (1996) and Perruchet & Amorim (1992) found that subjects acquired some conscious knowledge about the sequential structure of the training material.

Objective tests are generally accepted to be more sensitive to conscious knowledge because they facilitate recall by providing retrieval cues and because a forced-choice procedure is employed (Reingold & Merikle, 1990; St. John & Shanks, 1997). Yet, the evidence provided by these measures of awareness is often seen as inconclusive. A central criticism to objective tests is that they lack exclusivity, i.e. they do not solely measure what they are supposed to. As Reingold & Merikle (1988, 1990) argue, objective tests are inadequate measures of conscious knowledge because they may be contaminated by unconscious knowledge. When objective tests indicate greater than zero sensitivity, it is unclear whether performance is driven exclusively by conscious knowledge or not. Thus, any approach based on objective measure runs the risk of underestimating the influence of unconscious knowledge (Merikle, Smilek, & Eastwood, 2001).[10]

Subjective tests. In subliminal perception research, Cheesman & Merikle (1984, 1986) have suggested that the boundary between conscious and unconscious knowledge should be defined in terms of subjective rather than objective thresholds. In a typical subliminal perception experiment, subjects are required to respond to visual targets that are flashed on a computer screen (e.g., Marcel, 1983; Greenwald, Klinger, & Liu, 1989; Schmidt, Crump, Cheesman, & Besner, 2007). According to Cheesman & Merikle (1984, p. 391), the subjective threshold can be defined as "the detection level at which subjects claim not to be able to discriminate perceptual information at better than a chance level" and the objective threshold as "the detection level at which perceptual information is actually discriminated at a chance level." That is, performance is below the subjective threshold when subjects claim to be unable to detect perceptual information even though their detection performance is above chance. Performance is below the objective threshold when detection performance is at chance. Thus, the subjective threshold is the point at which subjects lack metaknowledge

[10] The "contamination" problem has also arisen in implicit memory and in subliminal perception research (see Buchner & Wippich, 2000; Reingold & Merikle, 1990).

(they do not know that they know); the objective threshold is the point at which subjects do not know that a stimulus was presented (Dienes & Berry, 1997). According to Cheesman and Merikle, perception is unconscious when it is under the subjective threshold (see also Kihlstrom, 2004; Kihlstrom, Barnhardt, & Tataryn, 1992; Merikle & Daneman, 1998).

In implicit learning research, Dienes and his collaborators (Dienes, 2004, 2008; Dienes et al., 1995; Dienes & Berry, 1997; Dienes & Perner, 1996, 1999; Dienes & Scott, 2005) have advocated the use of the subjective threshold as a criterion of implicitness. In AGL, for example, performance can be said to be below an objective threshold if subjects are classifying test strings at chance level. Performance is below the subjective threshold when classification performance is above chance even though subjects claim to be guessing, i.e. learning is unconscious if subjects lack metaknowledge. Dienes et al. (1995) suggested two ways in which subjects can lack metaknowledge. Subjects might lack metaknowledge in the sense that they believe to be guessing when their classification performance suggests that they are not. Dienes et al. called this the *guessing criterion*. Subjects might also lack metaknowledge when their confidence is unrelated to their accuracy. This criterion was originally introduced by Chan (1992) and labelled *zero correlation criterion* by Dienes et al. Several studies have provided evidence of unconscious learning by these criteria (Allwood, Granhag, & Johansson, 2000; Channon et al., 2002; Dienes & Altmann, 1997; Dienes et al., 1995; Dienes & Longuet Higgins, 2004; Dienes & Perner, 2003; Tunney & Altmann, 2001).

One way of dissociating conscious and unconscious processes is to collect confidence ratings (e.g., Cheesman & Merikle, 1984; Dienes et al., 1995; Kunimoto, Miller, & Pashler, 2001; see Redington, Friend, & Chater, 1996, for an alternative view). In AGL, for example, subjects can be asked to report, for each grammaticality judgment, how confident they were in the decision. Chan (1992, cited in Berry & Dienes, 1993) asked for confidence ratings after each judgment. One group of subjects was trained under intentional learning conditions by means a rule search task. Another group was trained incidentally by means of a memorization task. Chan found that, in the intentional group, confidence was strongly related to accuracy, indicating that knowledge was represented more explicitly. In contrast, the incidental group was just as confident in correct decisions as in incorrect ones, suggesting that knowledge was represented unconsciously. Confidence ratings can be taken after each grammaticality decision or concurrently. Dienes et al. (1995, Experiments 1, 2, 3 and 5), Dienes & Longuet-Higgins (2004), Dienes & Scott (2005, Experiment 2), Kuhn & Dienes (2005, Experiment

1b), Tunney & Shanks (2003, Experiment 1b) and Ziori & Dienes (2008), for instance, prompted subjects to enter a confidence rating after each classification decision. Tunney & Shanks (2003, Experiments 1a, 2 and 3), on the other hand, collected confidence ratings concurrently. After receiving training on an artificial grammar, subjects made their grammaticality and confidence judgments at the same time by pressing one of four buttons: "yes conforms to rules – more confident", "yes conforms to rules – less confident", "no does not conform to rules – less confident", and "no does not conform to rules – more confident". A review of the literature shows that most confidence ratings are taken after the grammaticality decision.

A different methodological issue concerns the selection of the type of confidence scale (see Dienes, 2008, for discussion). Scales can be classified according to whether they are continuous (50-100%) or noncontinuous (e.g., binary) and according to the type of categories they employ: numerical categories (50%, 51-59%, etc.) or verbal categories (e.g., guess, somewhat confident, and very confident). Different types of scales have been used in implicit learning research. Dienes et al. (1995, Experiments 1 and 5), Dienes & Scott (2005), and Ziori & Dienes (2008), for example, relied on continuous scales that ranged from 50% (complete guess) to 100% (complete certainty). In Dienes et al. (1995, Experiments 2 and 3), subjects placed their confidence levels on a more limited scale that ranged from 1 (complete guess) to 5 (complete certainty) instead. Channon et al. (2002), on the other hand, used a scale with three response options (0 = guess, 1 = fairly certain, 2 = certain), while Tunney & Shanks (2003, Experiments 1a, 1b, and 4) employed a binary confidence scale (high vs. low confidence). The selection of the confidence scale is not without controversy, and a recent debate focused on the appropriateness of different scales as measures of awareness. In a standard AGL experiment, Tunney & Shanks (2003, Experiment 4) compared the use of binary and continuous confidence scales. In the experiment, subjects first made their grammaticality and confidence decisions concurrently. After this, subjects were prompted to enter a continuous confidence rating (50%-100%). Tunney and Shanks found that, although binary and continuous scales measure the same aspect of awareness, binary judgments appeared more sensitive to low levels of awareness than continuous scales. In a follow-up study, Tunney (2005) reached the same conclusion. More recently, however, Dienes (2008) compared the sensitivity of different types of confidence scales and found no indication that any scale was more sensitive than the other. The matter has yet to be settled.

The use of subjective measures, including confidence ratings, has been criticized because of the problem of response bias (Eriksen, 1960; Reingold & Toth, 1996; see Dienes, 2004, 2008, for discussion). The problem is that subjects may systematically claim to be guessing when, in fact, they possess a small degree of awareness.[11] Subjects set their own criterion for reporting knowledge: more conservative subjects may indicate that they are guessing on their grammaticality judgments unless they are absolutely sure, while more liberal subjects may consistently report high levels of confidence at the slightest intuition. In subliminal perception research, Kunimoto, Miller, & Pashler (2001) recently proposed that the problem of response bias in subjective measures could be solved by combining confidence ratings with signal detection analysis (specifically, type II *d'*). Kunimoto et al. asked subjects, after each discrimination response, to place their confidence levels on a binary scale (high vs. low confidence). If subjects were aware of the information they were using to discriminate between stimuli, they should report high confidence levels for correct decisions and low confidence levels for incorrect decisions, i.e. accuracy and confidence should correlate. If subjects had no awareness of the information they were using in the discrimination task, high and low confidence responses should be indiscriminately assigned to correct and incorrect decisions. The confidence ratings can then be categorized in terms of signal detection theory (SDT; Green & Swets, 1966), which ensures that the sensitivity of the measure is unaffected by subjects' own report criteria (see Macmillan & Creelman, 1991, and Wickens, 2002, for introductions). The SDT measure of sensitivity (*d'*) can be easily calculated and serve as a bias-free index of awareness (though see Evans & Azzopardi, 2007).[12] In the field of implicit learning, Kunimoto et al.'s binary confidence technique has been introduced by Tunney & Shanks (2003; see also Rebuschat & Williams, 2006).

A criticism that can be levelled specifically at the use of confidence ratings concerns the type of knowledge that is assessed by this measure. Consider the case of natural language acquisition (see Dienes, 2008; Dienes & Perner, 1999; Seth et al., 2008). As mentioned above, language acquisition is considered a prime example of implicit learning. All cognitively unimpaired adults are able to discern grammatical sentences of their native language from ungrammatical ones, even though they are unable to report the underlying rule-system.

[11] In perceptual judgment tasks, for example, subjects are often systematically underconfident in their decisions (Juslin, Winman, & Olsson, 2000).

[12] Correct high confidence judgments are categorized as "hits", correct low confidence judgments as "misses", incorrect high confidence judgment as "false alarms", and incorrect low confidence judgments as "correct rejections". The *d'* value corresponds to the z-value of the hit rate minus the z-value of the false alarm rate. A *d'* of zero indicates no awareness; a *d'* of greater than zero indicates awareness (Kunimoto et al., 2001, p. 303)

However, if asked how confident they are in their grammaticality decisions, most native speakers will report high confidence levels, as in: "„John bought an apple in the supermarket.' is a grammatical sentence and I am 100% confident in my decision, but I do not know what the rules are or why I am right." Since in these cases accuracy and confidence will be highly correlated, does this mean that language acquisition is not an implicit learning process after all? Probably not. Dienes (2008; Dienes & Scott, 2005) proposed a convincing explanation for this phenomenon, based on Rosenthal's (1986, 2005) Higher-Order Thought Theory (see Rosenthal, 2000, 2005, 2008 for introductions).

Dienes suggested that, when subjects are exposed to letter sequences in an AGL experiment, they learn about the structure of the sequences. This *structural knowledge* can consist, for example, of knowledge of whole exemplars, knowledge of fragments or knowledge of rules (e.g., "A letter sequence can start with an M or a V.") In the testing phase, subjects use their structural knowledge to construct a different type of knowledge, namely whether the test items shared the same underlying structure as the training items (e.g. "MRVXX has the same structure as the training sequences."). Dienes labelled this *judgment knowledge*. Both forms of knowledge can be conscious or unconscious. For example, a structural representation such as "An R can be repeated several times." is only conscious if it is explicitly represented, i.e. if there is a higher-order thought such as "I {know/think/believe, etc.} that an R can repeated several times." Likewise, judgment knowledge is only conscious if there is a corresponding higher-order thought (e.g., "I {know/think/believe, etc.} that MRVXX has the same structure as the training sequences.") The guessing and the zero correlation criteria measure the conscious or unconscious status of judgment knowledge, not structural knowledge.[13]

Dienes & Scott (2005) assume that conscious structural knowledge leads to conscious judgment knowledge. However, if structural knowledge is unconscious, judgment knowledge could still be either conscious or unconscious. This explains why, in the case of natural language, people can be very confident in their grammaticality decisions without knowing why. In this case, structural (linguistic) knowledge is unconscious while (metalinguistic) judgment knowledge is conscious. The phenomenology in this case is that of intuition, i.e. knowing that a judgment is correct but not knowing why. If, on the other hand, structural and

[13] Dienes & Perner (1999, p. 753) propose the following division: "We suggest that implicit learning is a type of learning resulting in knowledge which is not labelled as knowledge by the act of learning itself. Implicit learning is associative learning of the sort carried out by first-order connectionist networks [...] Explicit learning is carried out by mechanisms that label the knowledge as knowledge by the very act of inducing it; a prototypical example is hypothesis testing. To test and confirm a hypothesis is to realize why it is knowledge."

judgment knowledge are unconscious, the phenomenology is that of guessing. In both cases the structural knowledge acquired during training is unconscious. Dienes and Scott proposed that, in the case of AGL experiments, the conscious status of both structural and judgment knowledge can be assessed concurrently by adding source attributions to the confidence ratings of the testing phase. That is, in addition to asking subjects how confident they were in their grammaticality judgments, one also prompts them to report the basis (or source) of their judgments. As Dienes and Scott suggest, subjects could be asked whether their judgment was based on guessing, intuition, conscious rules or memory for training strings.

In sum, the case for unconscious learning critically depends on our conception of awareness and the method employed to detect the conscious state of knowledge. If learning is thought to be "implicit" when subjects are unaware of the learning content, then the evidence for unconscious learning is somewhat mixed. There is little doubt that subjects are often unable to verbally express the knowledge they have acquired, even when the latter is assessed more rigorously by means of structured questionnaires (e.g., Berry, 1984, cited in Berry & Dienes, 1993). Thus, if verbal reports are taken as a measure of awareness, learning can result in unconscious knowledge (Lewicki et al., 1992; Reber, 1989, 1993). As has been pointed out, however, the use of verbal reports is highly controversial as subjects might fail to report for reasons that are unrelated to the potentially unconscious state of their knowledge. The use of objective tests (old/new recognition, fragment completion, etc.) has been proposed as a more stringent measure but is itself not without problems. Objective tests are more exhaustive measures but suffer from the problem of contamination, i.e. performance on an objective test might be driven by both conscious and unconscious knowledge (Reingold & Merikle, 1988, 1990; see Butler & Berry, 2001, for a discussion). The observation of a subjective threshold by means of confidence ratings or source attributions represents an alternative approach to conceptualizing and measuring awareness. If lack of awareness is taken as performance below the subjective threshold, as suggested by Cheesman & Merikle (1984), there is general agreement about the existence of unconscious cognitive processes operating in perception (Erdelyi, 2004; Greenwald, 1992; Kihlstrom, 1987; Merikle & Daneman, 2000) and in learning (Dienes et al., 1995; Dienes & Scott, 2005).

1.2. Implicit learning and language acquisition

Despite the widespread recognition that language acquisition constitutes a prime example of implicit learning (e.g., Cleeremans et al., 1998; Frensch & Rünger, 2003; Reber, 1967, 1993; St. John & Shanks, 1997), relatively little effort has been made, within linguistics or cognitive psychology, to investigate natural language acquisition under the theoretical framework provided by implicit learning research (Winter & Reber, 1994). It is somewhat ironic therefore that early research on implicit learning processes was directly influenced by previous work in theoretical linguistics and language acquisition research. For example, Chomsky (1956, 1957; Chomsky & Miller, 1958) specified the type of rule-system (the finite-state grammar) that would become the cornerstone of much of implicit learning research, and Reber's (1967, 1969) methodology was much inspired by that of Miller (1958), whose research focused on the cognitive processes underlying language acquisition. As mentioned above, there are nonetheless relatively few studies that investigate language acquisition within the framework of implicit learning.

1.2.1. Statistical learning and infant artificial language acquisition

Roughly around the same time as Reber conducted his first studies on implicit learning (Reber, 1965, 1967), several researchers began employing finite-state grammars in order to investigate language acquisition (see Braine, 1963, 1965, 1966; Moeser & Bregman, 1972; Segal & Halwes, 1965, 1966; Smith, 1966). Braine (1963), for example, reported the results of six experiments in which children were first trained on the word order patterns of an artificial language and then tested to see whether they could generalize the patterns to novel instances. In the most basic variation of Braine's paradigm (Experiment 1), subjects were exposed to two-word sequences. There were two classes of words: A words (e.g., KIV, JUF) always occurred in first position and P words (e.g., BEW, MUB) always in second. Children were told that they would learn a new language by means of a word game, i.e. learning was not incidental. The training task consisted of sentence completion problems that required them to remember the position of each word. Word were printed on cards and placed in front of the subjects, either to the left (1st position) or to the right (2nd position), depending on the word class. For each trial, the experimenter would lay down one word card and leave the other position vacant. Subjects were then given two word cards, one of each class, and asked to choose one to complete the sentence. When subjects chose the correct card, they won a poker chip (eight of which were worth a chocolate); when they made a mistake, they were told the correct answer. After the training phase, subjects received generalization problems. New words (e.g., FOJ, YAG) were either placed to the left or to the right, and subjects had to pick one word out of two alternatives (e.g., KIV, BEW) to complete the sentence. Braine found that children were able to complete the sentences appropriately, i.e. if the new word was placed on the left side, children would select a P word to complete the sentence, and vice versa. Braine took this as evidence for *contextual generalization*.[14]

The early focus of artificial language research was on the acquisition of linear structure. Several studies (e.g., Braine, 1963, 1966; Smith, 1966) showed that subjects of artificial language experiments could readily learn the positions of word classes but, crucially, they were unable to learn dependencies between classes (Smith, 1966; Segal & Halwes, 1965, 1966). Moeser & Bregman (1972), using an artificial phrase-structure language, investigated the effect of semantic information on the acquisition of interclass dependencies. Adult subjects were exposed to the artificial language under intentional learning conditions. All subjects were presented with four-word sequences, e.g. FET YOW ZOR NAK. In three groups of subjects (Experiments 2-4), the words were associated with different reference fields (see figure 3). In experiment 3 (Class Correspondence), for example, each word class was associated with a specific geometric form (e.g. class A words were always associated with coloured rectangles). In experiment 4 (Syntax Correlation), subjects received information about linguistic dependencies in addition to the information about word classes. For instance, one class of words was represented as geometric forms, while a different class of words, whose appearance was dependent of the first, was represented as border variations of the same geometric forms (e.g., class A words were represented as coloured rectangles and class D words as border variations of rectangles). Subjects were instructed to discover the grammatical structure underlying the word sequences and to learn what picture each word referred to. Moeser and Bregman found that all subjects learned some rules of the grammar but that only subjects in the Syntax Correlation condition (Experiment 4) were able to learn the linguistic dependencies between word classes. They interpreted this as evidence for the extensive semantic mediation in syntactic learning and suggested that "it is only when the elements in the reference field mirror the syntactic constraints of the language that complex

[14] "For verbal learning, contextual generalization may be defined informally as follows: when a subject, who has experienced sentences in which a segment (morpheme, word, or phrase) occurs in a certain position and context, later tends to place this segment in the same position in other contexts, the context of the segment will be said to have generalized, and the subject to have shown contextual generalization." (Braine, 1963, p. 323)

grammatical relations are easily acquired." (Moeser & Bregman, 1972, p. 769; see also Moeser & Bregman, 1973; Mori & Moeser, 1983)

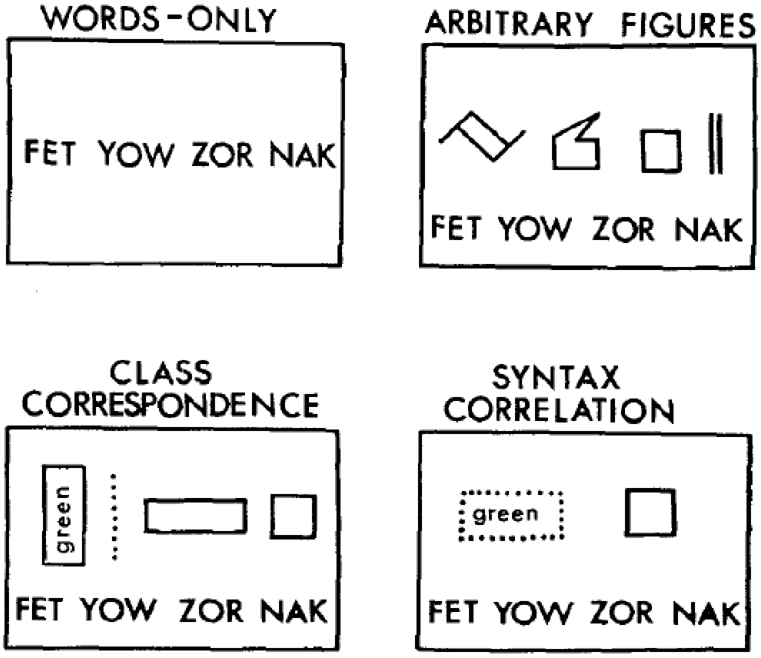

Figure 3. Example of how sentences appeared in the different reference conditions (Moeser & Bregman, 1972, p. 762)

Morgan & Newport (1981) proposed an alternative view on the acquisition of interclass dependencies. As Chomsky (1956, 1957, 1965) pointed out, human language is simultaneously organized into linear and hierarchical structures. As an hierarchical system, human language contains several cue systems that demarcate phrase structure, e.g. intonational and morphological cues. In their study, Morgan and Newport explored whether linguistic dependencies could be acquired under conditions where constituent organization is a more salient part of the artificial language. Morgan and Newport expected that the successful acquisition of hierarchical grammars requires linguistic input that consists of words grouped into phrases, and not of ungrouped word strings. The study adapted the basic methodology of Moeser & Bregman (1972) but employed an artificial language that included a set of correlated cues to constituent organization. Morgan and Newport found that all subjects learned the lexicon and the unconditional regularities of word classes, i.e. word-image associations. However, only subjects provided with additional cues to constituent structure in their reference world learned the constituent structure as well as the dependencies that hold between them. Subjects whose input lacked these cues failed to acquire a phrase-structure representation. Morgan and Newport proposed that natural languages contain phrase bracketing devices such as prosody, concord morphology, and function words in part because they facilitate language acquisition (see also Morgan, 1986: Bracketed Input Hypothesis).

Building on Morgan and Newport's research, Saffran (2001, 2002) investigated the use of distributional information for discovering linguistic phrase structure. Saffran (2001) examined whether *predictive dependencies*, e.g. the relationship that exists between word classes within a phrase, could serve as a statistical cue for the discovery of phrase boundaries. Subjects were exposed to an artificial language that consisted of a phrase-structure grammar and nonsense syllables that were associated with different word classes (e.g. A words: *biff*, *hep*, *mib*; D words: *klor*, *pell*; etc.). In the artificial language, the predictive dependencies between words that occurred in the same phrase were the only cue for phrasal grouping. In one type of phrase, for instance, A words occurred without D words, but the presence of D words always predicted the presence of A words.[15] In experiment 2, children and adults were exposed to sequences of nonsense syllables under incidental learning conditions by means of a picture drawing task. Subjects drew an illustration while nonsense sequences were played in the background. Afterwards, all subjects performed on a variety of tests. Saffran found that both adults and child learners could acquire the phrase-structure grammar by using predictive dependencies as a grouping cue. In a subsequent study, Saffran (2002) showed that adults and children were better at acquiring artificial languages that contain predictive dependencies than those that do not. Interestingly, Saffran also found that the same constraint on learning emerges in tasks that use nonlinguistic materials such as tones or geometric shapes. Saffran (2002, 2003) took this as evidence for a domain-general learning mechanism for the induction of language (see Saffran & Thiessen, 2007, for an overview).

The majority of research in the statistical learning tradition (see Aslin, Saffran, & Newport, 1999; Gómez, 2005, 2006; Perruchet & Pacton, 2008; Saffran, 2003, for overviews) has focused on the acquisition of adjacent relationships between stimuli. However, as Williams (in prep) points out, an important characteristic of natural languages is that they contain non-adjacent dependencies. Examples of long-distance dependencies in natural languages include morphosyntactic relationships such as number agreement (e.g., *The book*s *purchased by the*

[15] In English, for example, the presence of a determiner usually signals the subsequent occurrence of a noun, but a noun need not necessarily be preceded by a determiner.

department were *very expensive*.) and auxiliary-participle agreement (e.g., *He* was *eat*ing). Gómez (2002) examined whether transitional probabilities could be used as cues in the acquisition of long-distance dependencies. Gómez employed two artificial languages, each of which generated three-word sequences (e.g., *pel wadim rud*, *vot wadim tood*, etc.). Both languages contained the same adjacent dependencies, which means these were not informative cues. Subjects could only distinguish the artificial languages by learning the non-adjacent dependencies. In experiment 1, adult learners were exposed to one of the two languages and told to listen to the sequences for a subsequent test. In the testing phase, subjects were then asked to identify grammatical sequences. (No measure of awareness was included.) The results showed that subjects were, in fact, able to use statistical cues to acquire non-adjacent dependencies under certain conditions. In experiment 2, the methodology was adapted in order to investigate whether infants could acquire long-distance dependencies in the same fashion. Gómez found that 18-month old infants were also able to exploit the statistical structure in the input to learn non-adjacent relationships. Maye & Gómez (2005) showed that even 15-month-old infants are able to do so.

Newport & Aslin (2004) compared the acquisition of different types of non-adjacent dependencies. In experiment 1, Newport and Aslin used an artificial language which consisted of trisyllabic words (*badite*, *gutodo*, *pidira*, etc.) and where the first and the third syllables were perfectly predictable. Adult learners had to track non-adjacent syllables in words in a continuous stream of speech. Subjects were instructed to listen to a nonsense language. They were told that the language contained words and that their task was to find out where the words began and ended. They were also told that they would be tested afterwards. In the testing phase, subjects received a 2AFC test. For each trial, subjects heard two trisyllabic strings. One of the strings was a word form from the artificial language, while the other was only a partial word. Their task was to discriminate between the two. Newport and Aslin found no learning effects under this condition: subjects were unable to learn the non-adjacent syllable dependencies. In the remaining two experiments, Newport and Aslin investigated the acquisition of non-adjacent phonemic segments, a type of long-distance dependency that occurs in natural languages (e.g., in Hebrew and Arabic). In experiment 2, the non-adjacent dependency occurred between the consonants of trisyllabic words (e.g., *p_g_t*); in experiment 3, it occurred between the vowel segments (e.g., *_a_u_e*). The task instructions were slightly changed for these experiments. Subjects were instructed to listen to a nonsense language but they were not informed about the structure of the language. They were told to listen to the language and that they would be tested afterwards. The testing procedure was the same as in experiment 1. Newport and Aslin found that, in contrast to non-adjacent syllable dependencies, non-adjacent segment regularities were readily learned. According to Newport and Aslin, the results support the hypothesis that certain types of regularities, namely those that are common in natural languages, are readily learned while those that are uncommon in natural languages are not. It remains unclear under which conditions human learners are able to use statistical cues to acquire non-adjacent dependencies (see Bonatti, Peña, Nespor, & Mehler, 2005; Onnis, Monaghan, Richmond, & Chater, 2005; Peña, Bonatti, Nespor, & Mehler, 2002; Perruchet & Pacton, 2008; Perruchet, Tyler, Galland, & Peereman, 2004; Kuhn & Dienes, 2005; Creel, Newport, & Aslin, 2004).

In sum, the research literature on language acquisition abounds with studies that employ artificial grammars in order to investigate the basic processes underlying natural language acquisition (see Gómez, 2007, and Gómez & Gerken, 2000, for overviews). The most important finding to emerge in recent years has been the observation that infants, children and adults can use statistical cues such as transitional probabilities to acquire different aspects of language, including the lexicon (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), phonology (e.g., Maye & Gerken, 2001; Maye, Werker, & Gerken, 2002), and syntax (e.g., Gómez & Gerken, 1999; Marcus et al., 1999; Newport & Aslin, 2004; Saffran, 2001, 2002; Thompson & Newport, 2007). However, these studies differ from traditional implicit learning experiments in important ways. Firstly, most of these studies do not integrate measures of awareness. This is, of course, in part explained by the fact that infants are incapable to verbally report, indicate confidence levels, or perform on fragment completion tasks. However, many of the experiments conducted, for example, within the statistical learning framework employ children or adults as subjects, which means that basic measures of awareness could be administered. Usually, lack of awareness is assumed but not empirically assessed (see Fiser & Aslin, 2002, p. 500; Aslin et al., 1999, p. 368; Gómez, 2002, p. 431, for examples). Secondly, subjects are often exposed to the grammar under conditions that are far from incidental. As mentioned, Newport & Aslin (2004) observed learning of non-adjacent segment dependencies (Experiments 2 and 3) but not of non-adjacent syllable dependencies (Experiment 1). Gómez (2007) argues that the difference might, in part, be explained by the fact that the syllable learning task was more demanding (it involved speech segmentation). However, one could also suggest that the differences in learning could be explained by the fact that experiments 2 and 3 employed more incidental tasks, while experiment 1 exposed subjects under more intentional conditions. Perhaps subjects would be able to acquire the non-adjacent syllable sequences had they had not been informed that the language contained words and that their task was to find out where the words began and ended. Finally, the studies reviewed in this section tend to employ artificial grammars that resemble the structure of natural languages more closely. In contrast to many implicit learning experiments, phrase-structure grammars, instead of finite-state systems, are used to generate the stimulus sequences. In addition, the lexicon in most of these experiments does not consist of letter or tone sequences but of artificial "words".

1.2.2. Second language acquisition

The field of second language acquisition research (SLA) has a long-standing interest in implicit and explicit learning processes (see DeKeyser, 2003; N. Ellis, 1994b, 2007; Williams, in prep, for overviews). In part, this interest was sparked by Krashen's (1981, 1982, 1985, 1994) proposal that learners possess two independent ways of developing knowledge of a second language (L2). According to Krashen, language *acquisition* is an incidental process that results in tacit linguistic knowledge, while language *learning* is an intentional process that results in conscious, metalinguistic knowledge. In online speech comprehension and production, learners are thought to rely exclusively on *acquired* (or implicit) knowledge. The role of *learnt* (or explicit) knowledge is to monitor utterances for mistakes. Importantly, Krashen claimed that there is no interface between explicit and implicit knowledge. For example, explicit knowledge of a rule does not help the implicit acquisition of the same rule. For these reasons, Krashen argued that L2 teaching should focus on creating the conditions for language *acquisition* to take place (see Krashen & Terrell's, 1983, Natural Approach as an example).

Krashen's proposals and their implications for L2 pedagogy generated considerable controversy (see Gregg, 1984; McLaughlin, 1978, for early critiques; see Gass & Selinker, 2001; R. Ellis, 1994, for discussions). But they were also partially responsible for the increased interest in implicit and explicit processes in L2 acquisition and use. Many studies have focused on the role of explicit knowledge (e.g., Fotos, 1993; Fotos & Ellis, 1991; Green & Hecht, 1992; Han & Ellis, 1998; Hulstijn & Hulstijn, 1984; Roehr, 2008; Swain, 1998). Other studies have compared implicit and explicit learning more directly, usually with the objective of determining the effectiveness of different instructional treatments (e.g., Alanen, 1995; de Graaff, 1997; DeKeyser, 1995; Doughty, 1991; see R. Ellis, 2001, and Norris & Ortega, 2000, for overviews). In these studies, subjects are first exposed to selected linguistic aspects under different treatment conditions. An incidental group, which is simply exposed to the target structures in a meaning-focused task, often serves as a control group. The experimental groups receive different types of instructional treatment, which range from more "explicit" treatments (tasks that direct subjects' attention to linguistic form) to more "implicit" treatment (tasks for which subjects' attention has to be primarily on linguistic meaning). The teaching tasks are presumed to be implicit or explicit not because these studies necessarily include measures of awareness, but simply because of the underlying assumption that "explicit" tasks (e.g., consciousness-raising activities) lead to conscious knowledge, whereas "implicit" tasks (e.g., input enhancement) lead to unconscious knowledge. Some of the implicit/explicit L2 learning studies are classroom-based (e.g., Doughty, 1991; VanPatten & Oikennon, 1996; Scott, 1989, 1990), while others are lab-based (e.g., N. Ellis, 1993; de Graaff, 1997; DeKeyser, 1994, 1995; Robinson, 1996, 1997). Some authors employ artificial or semi-artificial languages (e.g., de Graaff, 1997; DeKeyser, 1994, 1995; Yang & Givón, 1997), whereas other authors prefer natural languages instead (e.g., Doughty, 1991; N. Ellis, 1993; Robinson, 1996, 1997, 2005).

Alanen (1995), for example, investigated the acquisition of a simplified version of Finnish by native speakers of English. The focus was on the effectiveness of different teaching methods. All subjects received two individual "lessons" on the linguistic target structure (locative suffixes and the consonant gradation rule). In their lessons, the control group (mere exposure condition) was instructed to read a text which included several examples of the target structures. Three other groups received the same texts but under different conditions. One group (enhance condition) read visually enhanced versions of the texts where the target structures were made perceptually salient by printing them in italics. A second group (rule condition) was first given explicit descriptions of the grammatical rules that govern the use of the linguistic targets and then read the unenhanced texts. Finally, a third group (rule and enhance condition) received both the rule descriptions and enhanced versions of the texts. Performance was measured by means of a battery of tests, which included a sentence completion task, a comprehension test, a word translation test, a grammaticality judgment task, think-aloud protocols and a questionnaire for assessing rule knowledge. Alanen found that learners in all conditions, including the mere exposure condition, made "explicit comments" on the target structures in think-aloud protocols, rule statement questionnaires and during their grammaticality judgments, which suggests that learning was not unconscious. She also found that subjects in the more explicit treatment conditions (rule; rule and enhance) outperformed subjects in the more implicit ones (mere exposure; enhance).

In a similar study, de Graaff (1997) investigated the impact of two instructional conditions on the acquisition of four morphosyntactic rules of an artificial language ("eXperanto"). The target structures were two simple rules (plural noun ending and negation position) and two complex rules (imperative mode and object placement). All subjects received approximately 15 hours of instruction in eXperanto by means of a self-study program which included a variety of tasks, including dialogues, vocabulary activities, and production activities. Subjects in the "explicit" condition received an explanation of grammatical structures after some tasks, whereas subjects in the "implicit" condition did not. Learning was assessed three times during the course of the experiment, once halfway through the treatment and twice after the end of the experiment, by means of proficiency tests. These consisted of two types of grammaticality judgment tasks, a gap-filling task, and a vocabulary translation task. In addition, at the end of the experiment, subjects were interviewed to see whether they could report any of the four morphosyntactic rules. De Graaff found that subjects in the explicit condition outperformed subjects in the implicit condition on the proficiency tests. The results of the oral interviews were not reported in the article, so it is impossible to assess whether learning in the implicit condition was, in fact, unconscious.

An examination of the SLA literature shows that, in fact, only very few studies focus exclusively on implicit L2 learning (Williams, 2004, 2005; Leung & Williams, 2006; Williams & Kuribara, 2008; Robinson, 2005). Williams (2005), for example, examined the acquisition of form-meaning connections under incidental learning conditions. In this study, subjects were exposed to a semi-artificial language consisting of English words and four artificial determiners (*gi*, *ro*, *ul* and *ne*) which encoded both distance (near vs. far) and animacy (animate vs. inanimate). At the beginning, subjects were told that the determiners functioned like English determiners, except that they also encoded distance: *gi* and *ro* were used for near objects, while *ul* and *ne* were used for far objects. Subjects were not informed that the artificial determiners also encoded animacy: *gi* and *ul* were used with animate objects, whereas *ro* and *ne* were used with inanimate objects. In the training phase, subjects received sentences such as "I spent an hour polishing ro table before the dinner party.", "When I was out for a walk I patted gi dog and it bit me.", and "In the pub I asked my friend to get ne stool from the bar." Subjects were instructed to listen to each training sentence, to indicate whether the novel determiner used in the sentence meant "near" or "far", to repeat the sentence verbatim, and to form a mental image of the general situation described by the picture. The testing phase consisted of two parts. In the first part, subjects read part of a novel sentence such as "The lady spent many hours sewing ..." and then had to select the appropriate segment to complete it from two options, e.g. "... gi cushions / ro cushions." In the second part, subjects were given the same test sequences but this time with the instruction to discover the rules that determined the choice of determiners. Awareness was assessed by means of verbal reports after each part of the testing phase. Williams found that, after the first part, 75% of participants reported to be unaware of the relevance of animacy in determiner usage, despite performing at 61% accuracy in the sentence completion task. After the rule discovery task, 50% of subjects were still unaware of the rule, yet their accuracy was still significantly above chance (58%). The results suggest that subjects can acquire form-meaning connections without becoming aware of what those connections are.

Further evidence for the implicit learning of form-meaning connections was provided by Leung & Williams (2006). In experiment 1, for example, subjects were exposed to an artificial determiner system by means of a novel reaction time task. The study used the same system as Williams (2004, 2005), except that here *gi*, *ro*, *ul* and *ne* encoded age (children vs. adults) and agency (agent vs. patient) instead of distance and animacy. Before the experiment, subjects were told that *gi* and *ro* were used with personal names referring to adults, whereas *ul* and *ne* were used with personal names referring to children. Subjects were not told that the choice of determiner also depended on the thematic role of the noun phrase: *gi* and *ul* were used with agents, while *ro* and *ne* were used with patients. In the experiments, subjects saw a picture on the computer screen that was accompanied by an auditory description. For example, the trial might consist of a picture of a girl kissing a boy on the cheek and an auditory description such as "Kiss ul Mary a boy on the face." The subjects' task was to indicate the location of the named individual in the image (left or right) by pressing the appropriate response key. In the first 114 trials of the experiment, the artificial determiner system followed the rules outlined above, but in the final 16 trials (the violation block) the rules were reversed, i.e. *gi* and *ul* were used with patients instead of agents. Learning was assessed by measuring the time it took subjects to identify the location of the named individual. Awareness was assessed by means of post-experimental verbal reports. Leung and

Williams found that subjects became increasingly faster in indicating the location of the named individual as they progressed through the trials. However, when the rule system was changed, response times increased significantly, which suggests that subjects had learned the relationship between the determiners and the thematic roles incidentally. The verbal reports showed that, at the end of the experiment, 80% of subjects remained unaware of the simple rules that determined the choice of determiners.

More recently, Williams & Kuribara (2007) investigated the acquisition of Japanese word order rules (head-direction and scrambling) under incidental learning conditions. A semi-artificial language consisting of English words and Japanese syntax ("Japlish") was used to generate the stimulus sequences. The training set, for example, included sentences such as "Student-ga dog-ni what-o offered?", "Vet-ga injection-o gave." and "That sandwich-o John-ga ate." Experimental subjects were exposed to a wide variety of sentence types during training (declarative sentences and wh-questions, with either canonical or scrambled word order) by means a plausibility judgment task. During training, each sentence was presented both auditorily and visually, and subjects had to judge whether the statements made were semantically plausible or not. Half the training items were semantically implausible. In the testing phase, learning was assessed by means of a grammaticality judgment task: subjects were informed that the word order in the previous sentences was generated by a rule-system and asked to judge whether novel sentences followed the same rules or not. Only half the test sentences were grammatical. Awareness was measured by means of verbal reports. Controls only performed on the testing phase. Williams and Kuribara found that experimental subjects outperformed controls on the classification task, suggesting that the training phase produced a learning effect. Nonetheless, experimental subjects were unable to report the word order rules when prompted to do so at the end of the experiment, which indicates that learning was largely unaware.

In sum, a review of the SLA literature shows that there is considerable interest in the topic of implicit and explicit learning (see e.g. the contributions in N. Ellis, 1994a; Schmidt, 1995; Ellis, Loewen, Elder, Erlam, Philp & Reinders, 2009). For obvious reasons, the interest in implicit and explicit processes within SLA is often pedagogic, and understandably many studies focus on the effects of different instructional treatments on L2 acquisition (see Doughty, 2003, for an overview). The common finding here is that, generally, learners are more likely to acquire linguistic target structures when exposed to them in the context of planned or incidental instructional activities. That is, rather than simply exposing subjects to comprehensible L2 input in the language classroom, as originally proposed by Krashen (1981, 1982), form-focused instruction appears to facilitate L2 acquisition (see R. Ellis, 2001, for a comprehensive review). From a language acquisition perspective, SLA studies have several advantages. For example, even in lab-based studies great care is taken to employ stimuli that resemble natural languages more closely than the finite-state grammars commonly used in AGL experiments. Stimulus sentences are usually grammatically more complex and tend to contain semantic information. Also, many of the training tasks actually require subjects to process the meaning of the sentences, which is an important characteristic of language acquisition outside the lab. From an implicit learning perspective, however, studies in SLA are somewhat problematic. Firstly, "implicitness" is assumed but not systematically assessed, i.e. studies do not employ measures of awareness (Williams, in prep). SLA studies often refer to "implicit learning conditions" (e.g., de Graaff, 1997), when the designation "incidental learning conditions" would, perhaps, be more appropriate. Secondly, even when experimenters choose to integrate a measure of awareness, this usually means prompting subjects to report knowledge either during or at the end of the experiment (e.g., Bowles & Leow, 2005; Leow, 1997, 2000, 2001; Leow & Morgan-Short, 2004). For reasons discussed above, verbal reports might be useful to assess what elements of the input subjects notice while performing on the experimental tasks, but they are arguably not sensitive enough to serve as measures of awareness (Shanks & St. John, 1994).

1.3. The present dissertation

The aims of the present dissertation were: (i) to investigate how adult learners acquire syntactic knowledge of a novel linguistic system under incidental learning conditions, (ii) to determine whether adult learners can acquire non-native syntax without becoming aware of the knowledge they have acquired, and (iii) to establish what role awareness plays in the acquisition process. The linguistic focus was on verb placement rules. Verb phrases (VP) play a central role in determining sentence structure (Towsend & Bever, 2001), and the acquisition of verb placement has been extensively studied in both first and second language acquisition (see e.g. the contributions in Hirsh-Pasek & Golinkoff, 2006, as well as Bohnacker, 2006; Clahsen & Muysken, 1986; Jansen, 2008; Parodi, 2000). In terms of methodology, the six experiments described below are based on the AGL paradigm. An artificial system was chosen to generate the stimulus material in order to ensure that the stimuli were novel to participants, a critical feature in learning experiments. In contrast to traditional AGL studies, however, the experiments were adapted to investigate the acquisition of natural language syntax. Firstly, rather than using a finite-state system or a miniature phrase-structure language, the rule-system employed here follows the word-order rules of a natural language, namely German. Secondly, rather than using a lexicon that consists of meaningless units (letters, tones, etc.), the lexicon in the following experiments consists of English words (with the exception of experiment 4). Both features, the use of a natural language grammar and the use of meaningful lexical units, increased the similarity of the artificial system to natural languages and the generalizability of the findings to language acquisition outside the lab (see Williams, 2005; Williams & Kuribara, 2007; Williams & Leung, 2006; de Graaff, 1997; DeKeyser, 1995, 1997, for other examples of semi-artificial languages). Finally, the experiments integrate several measures of awareness. As discussed above, the research literature on language acquisition abounds with studies that employ AGs in order to investigate the basic processes underlying natural language acquisition. However, these studies usually do not integrate measures of awareness, i.e. lack of awareness is assumed but not empirically assessed. The current dissertation is thus able to assess to what extent subjects become aware of the syntactic knowledge they have acquired.

2. Experiment 1

2.1. Introduction

The aims of this experiment were twofold. The first aim was to determine whether natural language syntax, in this case verb placement rules, could be acquired implicitly, i.e. incidentally and without awareness of the learning outcome. The second aim was to explore what role awareness plays in the acquisition of syntactic structure.

2.2. Method

2.2.1. Participants

Thirty-five native speakers of English (22 women and 13 men, mean age = 24.3 years) were recruited to take part in this experiment and randomly assigned to experimental and control conditions. The majority of participants (33) were students at the University of Cambridge at the time of the experiment. As far as their language background was concerned, five participants had been brought up bilingually, i.e. they had acquired a second language in addition to English during childhood and attained native-like proficiency in it. Participants had learned, on average, 2.5 foreign languages and none had a background in German or any other verb-second (V2) language.[16] Experimental and control groups did not differ significantly across age, $t(13.04) = 1.834$, $p > .05$, gender, $\chi^2(1, N = 33) = 0.29$, $p > .05$, occupation (student vs. nonstudent), $\chi^2(1, N = 33) = 3.275$, $p > .05$, and number of languages acquired, $\chi^2(4, N = 33) = 5.705$, $p > .05$. All subjects received £5 for participating in the experiment.

2.2.2. Stimulus material

A semi-artificial grammar, consisting of English words and German syntax, was used to generate the stimulus material for this experiment (see Ellis & Schmidt, 1998, Gómez & Gerken, 1999, and Schmidt, 1994, for discussions of the use of AGs in language acquisition

[16] In V2 languages, the second constituent of simple declarative sentences is always the finite verb. V2 word order occurs in all Germanic languages (German, Dutch, Swedish, Danish, etc.), with the exception of English.

research). In creating the stimuli, English declarative sentences were rearranged in accordance with German syntactic rules as in the examples below (1-3).

(1) Simple sentence (one-clause construction; simple predicate)
   a. English: Yesterday John bought the newspaper in the supermarket.
   b. German: Gestern kaufte John die Zeitung im Supermarkt.
   c. Stimulus: Yesterday bought John the newspaper in the supermarket.

(2) Simple sentence (one-clause construction; complex predicate)
   a. English: Yesterday John has bought the newspaper in the supermarket.
   b. German: Gestern hat John die Zeitung im Supermarkt gekauft.
   c. Stimulus: Yesterday has John the newspaper in the supermarket bought.

(3) Complex sentence (two-clause construction; sequence: main-subordinate)
   a. English: Last week Susan visited Melbourne because her daughter lived in Australia.
   b. German: Letzte Woche besuchte Susan Melbourne, weil ihre Tochter in Australien lebte.
   c. Stimulus: Last week visited Susan Melbourne because her daughter in Australia lived.

(4) Complex sentence (two-clause construction; sequence: subordinate-main)
   a. English: Since his parents needed groceries, David purchased everything necessary.
   a. German: Weil seine Eltern Lebensmittel brauchten, kaufte David alles Notwendige ein.
   b. Stimulus: Since his parents groceries needed, purchased David everything necessary.

As is evident from the examples, the elements within phrase boundaries were left intact, while the specific ordering of the phrases was altered. In (1), for example, the VP was moved from third position in the phrasal sequence to second. In (2), the auxiliary was placed in second position while the participle was moved to the end of the sentence. In (3), the VP of the main clause was moved to second position while the VP of the subordinate clause was placed in final position. Finally, in (4) the VP of the subordinate clause was moved to final position, whereas the VP of the main clause was shifted to first position.

The linguistic focus in this experiment was on four rules that determine the placement of VPs in the semi-artificial language. These rules, which are largely based on German, stated that, depending on the type of predicate (simple vs. complex), the type of clause (main vs. subordinate) and the type of clause sequence (main-subordinate vs. subordinate-main), finite verbs had to be placed either in first, second or final position in terms of the phrasal sequence (see Eisenberg, 1999, Fox, 2005, and Pittner & Berman, 2004, for introductions to German word order). Table 1 illustrates the four rules in question.

Table 1

Descriptions and Examples of the Four Verb Placement Rules

Rule	Description	Examples
V2	Finite verb placed in second phrasal position of main clauses that are not preceded by a subordinate clause.	*Today bought John the newspaper in the supermarket.*
Split VP	For complex predicates in main clauses that are not preceded by a subordinate clause: Auxiliary placed in second position, participle placed in final position	*Today has John the newspaper in the supermarket bought.*
V1	Finite verb placed in first position in main clauses that are preceded by a subordinate clause	*Because his parents the newspaper in the supermarket bought, spent John the evening in his study.*
VF	Finite verb placed in final position in all subordinate clauses	*Peter repeated today that the movers his furniture scratched.*

Rules "V2" and "Split VP" applied to main clauses that were not preceded by a subordinate clause. They differed in that the former rule applied to simple predicates and the latter rule to complex predicates. In the semi-artificial language, simple predicates occured both in simple and in complex sentences; complex predicates only occured in simple sentences.[17] Rule "V1" also applied to main clauses but only to those that were preceded by a subordinate clause. In contrast, rule "VF" applied to all subordinate clauses, irrespectively of whether a main clause preceded or followed.

A total of 192 sentences were drafted for this experiment. The sentences were read out by a male native speaker of British English, digitally recorded on a Sony Mini-Disc player (MZ-

[17] Simple sentences are one-clause constructions; complex sentences are two-clause constructions. A simple predicate consists of one verbal element, a complex predicate of two verbal elements (e.g., the auxiliary and a participle).

R700) and subsequently edited with sound processing software (Audacity, version 1.2.4). The stimulus sentences were divided into a training and a testing set.

2.2.2.1. Training set

The training set, which is reproduced in the appendix, consisted of 128 sentences and was subdivided into 64 plausible and 64 implausible constructions. In other words, half the sentences in the training set were syntactically correct but expressed semantically implausible propositions. Plausible and implausible items were designed so that participants would have to process the entire auditory string before being able to judge its plausibility. In most instances, the final phrase of the sentence would reveal whether the sentence was plausible or not, and participants had to process the entire string until reaching their judgment. The sentences below are examples of plausible and implausible constructions (4-5). Following convention in linguistics, a question mark was used to indicate implausible sentences.

(5) Plausible constructions

a. Chris entertained today his colleagues with an interesting performance.
b. Brian has usually many shots during his matches defended.
c. George repeated today that his students about their classes cared.
d. Since his teacher criticism voiced, put Chris more effort into his homework.

(6) Implausible constructions

a. ? Rose abandoned in the evening her cats on planet Venus.
b. ? Sarah has usually in the afternoon a soup e-mailed.
c. ? Cate confessed today that her horse the corridor murdered.
d. ? After his wife a thief surprised, communicated George with the police banana.

In order to train the experimental group, 32 sentences were created for each verb placement rule. All sentences were in the past tense. For each rule, templates were designed to serve as models in the generation of the stimulus items. The table below outlines the template options for the four rules.

Table 2

Syntactic Templates for the Four Rules. Frequencies in Training Set Are in Parentheses. Sample Sentences Are in Italics.

Rule	Templates	
V2	$[[\mathrm{NP}]_{subj} > [\mathrm{VP}] > [\mathrm{AP}]_{temp} > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{AP}\vert\mathrm{NP}\vert\mathrm{PP}]]$	(16)
	Rose gambled today with her savings at the casino.	
	$[[\mathrm{AP}]_{temp} > [\mathrm{VP}] > [\mathrm{NP}]_{subj} > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{AP}\vert\mathrm{NP}\vert\mathrm{PP}]]$	(16)
	Today justified Brian the investments during the meeting.	
Split VP	$[[\mathrm{NP}]_{subj} > [\mathrm{VP}_{aux}] > [\mathrm{AP}]_{temp} > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{VP}_{part}]]$	(16)
	Peter has today the college statutes in his speech challenged.	
	$[[\mathrm{AP}]_{temp} > [\mathrm{VP}_{aux}] > [\mathrm{NP}]_{subj} > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{VP}_{part}]]$	(16)
	Usually were the ambassadors for important discussions summoned.	
V1	$[[[\mathrm{SUB} > [\mathrm{NP}]_{subj} > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{VP}]] > [\mathrm{VP}] > [\mathrm{NP}]_{subj} > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{NP}\vert\mathrm{PP}]]$	(16)
	After the university press his book published, sent George many copies to his friends.	
VF	$[[\mathrm{NP}]_{subj} > [\mathrm{VP}] > [\mathrm{AP}]_{temp} > [\mathrm{SUB} > [\mathrm{NP}]_{subj} > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{VP}]]]$	(16)
	Peter repeated today that the movers his furniture scratched.	
	$[[\mathrm{AP}]_{temp} > [\mathrm{VP}] > [\mathrm{NP}]_{subj} > [\mathrm{SUB} > [\mathrm{NP}]_{subj} > [\mathrm{NP}\vert\mathrm{PP}] > [\mathrm{VP}]]]$	(16)
	Today opined Sarah that the storm her parents' boat overturned.	

The V2 rule applied to main clauses that were not preceded by a subordinate clause. All V2 sentences consisted of five phrasal units. The VP always occupied second position, while the first and the third positions either featured a subject noun phrase (NP_{subj}) or an expression of time. All temporal expressions in the training set were instantiated as an adverb phrase (AP_{temp}). Half the sentences featured NP_{subj} in preverbal position (pattern 1) and the other half AP_{temp} (pattern 2). The fourth and the fifth phrasal positions could be occupied by a nontemporal AP, an NP or by a prepositional phrase (PP). Because of the position of the finite verb, the syntactic pattern associated with this rule was designated "V2."

The Split VP rule applied to complex predicates in main clauses that were not preceded by a subordinate clause. All Split VP sentences consisted of six phrasal units. The auxiliary verb (VP_{aux}) occurred in second and the past participle (VP_{part}) in final position. The first and the third positions featured either NP_{subj} or AP_{temp}: If the NP_{subj} came in first position, then AP_{temp} would come in third, and vice versa. Half the sentences featured NP_{subj} in preverbal position and the other half AP_{temp}. The fourth and the fifth phrasal positions could be occupied by NP or PP. Importantly, all subject-initial sentences (pattern 1) were in the present perfect (auxiliary *have*), while all time-initial sentences (pattern 2) were passive constructions (auxiliary *be*). The syntactic pattern for this rule, based on the position of the two verbal elements, was "V2VF."

Rule V1 applied to main clauses that were preceded by a subordinate clause. The sentences based on the V1 template consisted of a subordinate clause followed by a main clause (clause sequence: subordinate-main). Subordinate constructions were marked as such by featuring a subordinating conjunction (SUB) as an introductory word.[18] Each subordinate clause featured three slots for phrasal elements: The last slot was always occupied by the VP and the first slot by NP_{subj}. NP and PP were options for second position. The main clauses allowed the same phrase types, though with different ordering restrictions. Here, the finite verb had to be placed in first position with NP_{subj} following in second. NP and PP were possibilities for third and fourth positions. No time expressions occurred in sentences for rule V1. The syntactic pattern associated with this rule was "VF-V1."

In contrast to the previous rules, rule VF only applied to subordinate clauses. All sentences based on the VF templates consisted of a main clause followed by a subordinate clause (clause sequence: main-subordinate). All main clauses consisted of three phrasal units, namely NP_{subj}, VP and AP_{temp}. As required by the V2 rule, the finite verbs were always placed in second position of the main clause. The first position was occupied by either NP_{subj} or AP_{temp}. If a sentence started with NP_{subj}, then AP_{temp} followed in third position (pattern 1). In contrast, if a sentence began with AP_{temp}, then NP_{subj} followed in third position (pattern 2). Half the sentences for the VF rule were based on pattern 1, the other half on pattern 2. Subordinate clauses also consisted of three phrasal elements. All subordinate clauses were clearly marked by featuring a subordinator at the beginning. As required by the VF rule, the finite verb was placed in final position. The first position was occupied by NP_{subj} and the second position by either NP or PP. The syntactic pattern primarily associated with this rule was "V2-VF."

Several restrictions applied to the lexicon of the training set. Each main verb, for example, could only occur once in the training sentences. The NP_{subj} in the main clauses was realized by one of eight names, four of which were male (*Peter*, *Brian, Jack, Mike*) and four female

[18] Subordinators do not count towards phrasal positions. According to Huddleston & Pullum (2005), *since* and *after* are prepositions, not subordinators.

(*Joanna, Sarah, Rose, Jessica*). Each name occurred 16 times in the training set. The adverbs *today* and *usually* were the only temporal expressions employed, each occurring with equal frequency in the training set. In the case of the subordinate clauses, only three subordinators were employed. All V2-VF sentences featured the subordinator *that*; half the VF-V1 sentences featured the subordinator *since* and the other half the subordinator *after*. Finally, the use of short verbs was avoided in subordinate clauses since a previous study (Rebuschat & Williams, 2006), which employed an earlier version of the present experimental paradigm, found that participants were biased against grammatical stimuli with monosyllabic verbs in final position. It was assumed that this could be due to lack of perceptual salience of the VP.[19] In the current experiment, only multisyllabic verbs were therefore used in clause-final position.

A frequency analysis of the training set showed that the average stimulus length was 8.5 words per sentence in V2 constructions (8.6 for plausible items, 8.4 for implausible items), 8.7 words per sentence for V2VF constructions (8.8 plausible, 8.6 implausible), 12.3 for VF-V1 constructions (12.6 plausible, 12.1 implausible), and 9.5 words per sentence for V2-VF constructions (9.4 plausible, 9.6 implausible).

2.2.2.2. Testing set

The testing set, which is reproduced in the appendix, consisted of 64 new sentences and was subdivided into 32 grammatical and 32 ungrammatical items. Eight sentence templates were created in order to test whether participants acquired knowledge about the possible positions for VP in simple and complex sentences. Each of the templates in tables 3 and 4 was used to generate eight stimulus sentences. The grammatical templates were the same templates used to create the training sentences, with the exception that, in the case of the V2 pattern, only time-initial sequences were generated (pattern 2 in table 2). Time-initial sentences sound less like English and were therefore assumed to be better test items. Table 3 outlines the grammatical sentence templates.

[19] Participants tended to reject sentences such as "Peter said today that his father frequently apples ate."

Table 3

Sentence Templates for the Four Grammatical Patterns. Frequencies in Testing Set Are in Parentheses. Sample Sentences Are in Italics.

Pattern	Template	
V2	$[[AP]_{temp} > [VP] > [NP]_{subj} > [NP|PP] > [AP|NP|PP]]$	(8)
	Yesterday scribbled David a long letter to his family.	
V2VF	$[[NP]_{subj} > [VP_{aux}] > [AP]_{temp} > [NP|PP] > [NP|PP] > [VP_{part}]]$	(4)
	David was yesterday by a new company hired.	
	$[[AP]_{temp} > [VP_{aux}] > [NP]_{subj} > [NP|PP] > [NP|PP] > [VP_{part}]]$	(4)
	Yesterday have his students with the American researcher chatted.	
V2-VF	$[[NP]_{subj} > [VP] > [AP]_{temp} > [SUB > [NP]_{subj} > [NP|PP] > [VP]]$	(4)
	David speculated yesterday that the suspect from prison escaped.	
	$[[AP]_{temp} > [VP] > [NP]_{subj} > [SUB > [NP]_{subj} > [NP|PP] > [VP]]]$	(4)
	In the afternoon acknowledged David [that her children to England moved.	
VF-V1	$[[SUB > [NP]_{subj} > [AP]_{temp} > [NP|PP] > [VP]] > [VP] > [NP]_{subj} > [NP|PP] > [NP|PP]]$	(8)
	When his wife in the afternoon the office left, prepared Jim dinner for the entire family	

The ungrammatical templates were similar to the grammatical ones with the exception that the position of VP was incorrect.[20] *VF sentences were similar to V2 sentences (pattern 2) but featured a VP in final position instead of second. *V3VF sentences closely matched V2VF sentences but featured a VP_{aux} in third position. *VF-VF resembled VF-V1 sentences but were ungrammatical since the VP in the main clause occurred in final position. Finally, *V2-V1 sentences differed from V2-VF sentences because the subordinate clause featured a VP in first position. Table 4 outlines the ungrammatical sentence templates.

[20] Following convention in linguistics, an asterisk was used to indicate ungrammatical sentences.

Table 4

Sentence Templates for the Six Ungrammatical Patterns. Frequencies in Testing Set Are in Parentheses. Sample Sentences Are in Italics.

Pattern	Template	
*VF	$[[\mathrm{AP}]_{temp} > [\mathrm{NP}]_{subj} > [\mathrm{NP} \mid \mathrm{PP}] > [\mathrm{NP} \mid \mathrm{PP}] > [\mathrm{VP}]]$	(8)
	Recently David to his son's house in Wales drove.	
*V3VF	$[[\mathrm{AP}]_{temp} > [\mathrm{NP}]_{subj} > [\mathrm{VP}_{aux}] > [\mathrm{NP} \mid \mathrm{PP}] > [\mathrm{NP} \mid \mathrm{PP}] > [\mathrm{VP}_{part}]]$	(8)
	Yesterday the guitar was by David smashed.	
*V2-V1	$[[\mathrm{NP}]_{subj} > [\mathrm{VP}] > [\mathrm{AP}]_{temp} > [\mathrm{SUB} > [\mathrm{VP}] > [\mathrm{NP}]_{subj} > [\mathrm{NP} \mid \mathrm{PP}]]]$	(4)
	Susan explained yesterday that remained the profit below the estimate.	
	$[[\mathrm{AP}]_{temp} > [\mathrm{VP}] > [\mathrm{NP}]_{subj} > [\mathrm{SUB} > [\mathrm{VP}] > [\mathrm{NP}]_{subj} > [\mathrm{NP} \mid \mathrm{PP}]]]$	(4)
	Recently maintained David that abstained his father from unhealthy food.	
*VF-VF	$[[\mathrm{SUB} > [\mathrm{NP}]_{subj} > [\mathrm{AP}]_{temp} > [\mathrm{NP} \mid \mathrm{PP}] > [\mathrm{VP}]] > [\mathrm{NP}]_{subj} > [\mathrm{NP} \mid \mathrm{PP}] > [\mathrm{NP} \mid \mathrm{PP}] > [\mathrm{VP}]]$	(8)
	Because his son an instrument wanted, David with the music teacher talked.	

Several restrictions applied to the lexicon of the testing set. With the exception of a few words (determiners, prepositions, etc.), no items were repeated from the training set, making the test analogous to the letter-set transfer paradigm in Artificial Grammar Learning research. For example, no verb from the training set was repeated in the testing phase. The NP_{subj} in the main clauses was again realized by eight names, four of which were male and four female. However, the testing set featured the names *David, John, Jim, Paul, Emma, Susan, Chloe,* and *Jennifer*, each occurring 16 times. Temporal expressions were instantiated by two different adverbs, in this case *recently* and *yesterday*, each of which occurred with equal frequency. In the case of the subordinate clauses, three subordinators were employed throughout the testing set. All V2-VF and *V2-V1 sentences featured the subordinator *that*. VF-V1 and *VF-VF sentences featured two new subordinators, namely *because* and *when*. As in the training set, monosyllabic verb forms were avoided in clause-final position.

A frequency analysis of the testing set indicated that the average stimulus length was 8.8 words per sentence for grammatical items and 9.1 for ungrammatical items. There was no significant difference between training and testing sets with regard to sentence length, $t(31) = .911$, $p > .05$, which suggests that length could not serve as a reliable cue to grammaticality in the testing phase. In addition, the bigram statistics of the training set were used to compute

Associative Chunk Strength (ACS), a standard measure of association.[21] This measure served to establish how similar the testing sentences were in relation to the training sentences in terms of surface similarity. Ideally, grammatical and ungrammatical items should not differ significantly on standard similarity measures. Since very few lexical items occurred in both sets, the bigram statistics were calculated over phrase types (NP, VP, NP, etc.) and not actual words (*John*, *bought*, *apples*, etc.). The analysis indicated that grammatical sentences had an average ACS of 60.9 and ungrammatical sentences an average ACS of 57.3. That is, grammatical items resembled the training set more closely than ungrammatical items, though not significantly so, $t(31) = 2.015$, $p = .053$.

2.2.3. Procedure

The experiment consisted of (i) a training phase, (ii) a testing phase, and (iii) a debriefing session. Experimental participants were trained on the semi-artificial grammar under incidental learning conditions by means of plausibility judgments. In the unexpected testing phase, they were then asked to classify novel instances of the semi-artificial grammar and to report confidence levels for each classification decision. The experiment concluded with a debriefing session during which subjects completed a questionnaire. All tasks were run on a Dell PC (Windows XP) and delivered via Cedrus SuperLab Pro (version 2.0). The precise instructions are reproduced in the appendix.

2.2.3.1. Training phase

At the beginning, participants were told they were taking part in an experiment on sentence comprehension during which they would be exposed to 128 “scrambled” sentences. The training task required participants to listen to the training set on an item-by-item basis and to judge whether the statement made was semantically plausible or not. The figure below illustrates the training procedure.

[21] The chunk strength of a test item refers to the frequency with which its constituent bigram or trigram chunks occurred in the training set. For each letter sequence, the overall chunk strength can be calculated by averaging the chunk strength of the letter string’s component bigrams and trigrams (see Knowlton & Squire, 1994, p. 85).

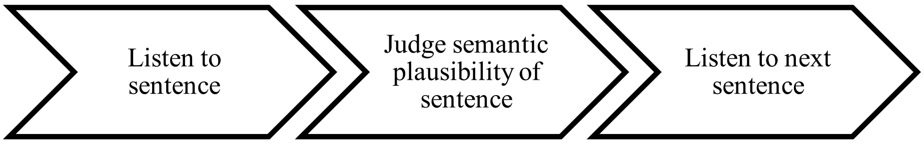

Figure 4. Training procedure in experiment 1.

Subjects were informed that the 128 scrambled sentences would be presented twice, i.e. they would listen to 256 training sentences in total. The presentation order of the training items was determined randomly for each participant. After the presentation of the first block, subjects could take a brief break before being exposed to the same stimulus set again, though in a different random order. While the auditory stimuli were played, the computer screen was blank. A prompt question ("Do you think this is plausible?") appeared immediately after the end of stimulus presentation. The two response keys for making the plausibility judgments (*/* for plausible, *z* for implausible constructions) were displayed on the screen below the prompt. Subjects were told before the experiment that they would receive feedback on their plausibility judgments. When the plausibility of a given sentence was misjudged, the sentence was played again and participants were required to repeat the plausibility judgment. There was a 500 ms delay between the end of the plausibility judgment and the presentation of the following sentence. The experiment began with a short practice session to familiarize participants with the training task. This consisted of four trials which were not repeated in the actual training phase. No mention was made that the scrambling followed the word order rules of a natural language. The entire training phase took, on average, 40 minutes to complete.

2.2.3.2. Testing phase

After training, experimental participants were informed that the word order in the previous sentences was not arbitrary but that it followed a "complex system" instead. They were then instructed to listen to 64 new scrambled sentences, only half of which would follow the same rule-system as the sequences they had just been exposed to. Those sentences that did obey the rules should be endorsed as "grammatical" and those that did not rejected as "ungrammatical." For each test sentence, participants were required to decide on its grammaticality ("Was the sentence grammatical?") and to report how confident they were in their judgment ("How confident are you in your decision?"). In the case of the control group, participants were merely told that they would listen to 64 scrambled sentences and asked to judge whether or not a sentence was grammatical. There were no time intervals between the three prompts, i.e. subjects would receive the next prompt immediately after making a response. There was, however, a 500 ms delay between the confidence rating and the presentation of a new test sentence. Participants were given no feedback as to the accuracy of their grammaticality decisions. Figure 5 illustrates the testing procedure.

Figure 5. Testing procedure in experiment 1.

Participants could indicate their levels of confidence by selecting two response options: low confidence or high confidence. A "low confidence" response indicated that participants had only little confidence in their classification decisions, while a "high confidence" response indicated that participants were very confident in their classification decisions. Participants were told to rely on their intuition when judging the grammaticality of the sentences. All keyboard options (*/* for grammatical, *z* for ungrammatical, *1* for low confidence, *9* for high confidence) were displayed on the screen below the relevant prompts. Accuracy in the grammaticality judgment task was used as a measure of learning. The binary confidence ratings were collected in order to assess the conscious or unconscious status of participants' metaknowledge. The 64 test sentences were presented to each participant in random order. The testing phase began with a short practice session to familiarize the participants with the new task. This consisted of four trials which were not repeated in the actual testing set. The entire testing phase took approximately 15 minutes to complete.

2.2.3.3. Debriefing

At the end of the experiment, all participants completed the debriefing questionnaire which is reproduced in the appendix. The questionnaire first required participants to estimate their accuracy in the classification task. The accuracy estimates served as an additional measure of awareness. If subjects were aware of the knowledge that determined their classification performance, there should be no difference between estimated and observed accuracy; if they were unaware, there should be a difference. The task is similar to Cheesman & Merikle's (1984, Experiment 2) block estimation procedure, except that the latter asked subjects to rate their accuracy repeatedly throughout the experiment (after each block of 48 threshold discriminations).[22] For this experiment, it was decided to ask subjects to estimate their accuracy just once because the testing phase was relatively short (68 judgments in total) and because subjects were asked to submit confidence ratings after each grammaticality judgment. It was felt that prompting subjects to estimate their accuracy repeatedly might affect their performance in the testing phase.

After the accuracy estimates, participants were then reminded that the word order of the sentences was determined by a "complex system" and prompted to verbalize any rule or regularity they might have noticed during the course of the experiment. Specifically, participants were asked:

(1) "While performing on the tasks of the experiment, did you notice any particular rule or regularity? If yes, please indicate what you believe you have noticed."

(2) "As mentioned in the experiment, the scrambling of the sentences was not arbitrary. Instead, the word order of the sentences was based on a complex system. Reflecting now specifically on the placement of words within the sentences, can you recall any specific rule or regularity?"

Like the accuracy estimates, the verbal reports served as measures of awareness. Finally, the questionnaire also asked participants to supply their name, age, gender, nationality, occupation and linguistic background.

[22] The use of the block estimation procedure as a measure of awareness has been criticized. Several studies have found that subjects tend to be underconfident when asked to estimate their performance (Björkman, Juslin, & Winman, 1993; Kunimoto, Miller, & Pashler, 2001).

2.3. Results

2.3.1. Training phase

Performance during the training phase was assessed by analyzing the plausibility judgments of the experimental group. In terms of frequency, participants committed, on average, 11.2 errors in their plausibility judgments. Because misjudged sentences were repeated, this suggests that the average experimental participant was exposed to 267.2 sentences during training. The analysis also showed that the experimental group judged 96.2% (*SD* = 2.3%) of sentences correctly. Experimental participants were significantly more accurate when judging plausible sentences (*M* = 97.1%, *SD* = 2.4%) than implausible ones (*M* = 94.8%, *SD* = 3.6%), $t(19) = 2.777$, $p < .05$, which indicates that implausibility affected accuracy in the plausibility judgment task negatively.

2.3.2. Testing phase

2.3.2.1. Grammaticality judgments

Overall performance

The analysis of the grammaticality judgments showed that the experimental group classified 54.6% (*SD* = 12.2%) of the test items correctly and the control group 51.9% (*SD* = 5.6%). The difference between the two groups was not significant, $t(33) = .805$, $p > .05$. Further analysis showed that neither the experimental group, $t(19) = 1.685$, $p > .05$, nor the control group, $t(14) = 1.302$, $p > .05$, performed significantly above chance. The results thus indicate that the training phase did not produce a learning effect in the experimental participants.

In order to establish whether any improvements in accuracy occurred during the testing phase, the grammaticality judgment data was equally divided into eight chronological blocks. The mean accuracy for each block of eight judgments was then computed for both groups. A repeated measures ANOVA showed no significant differences across the stages in the experimental group, $F(7, 133) = 1.214$, $p > .05$, or in the control group, $F(7, 77) = 1.650$, $p > .05$. That is, no online learning effect was observed during the testing phase. Figure 6 displays the mean accuracy across the eight stages for both groups.

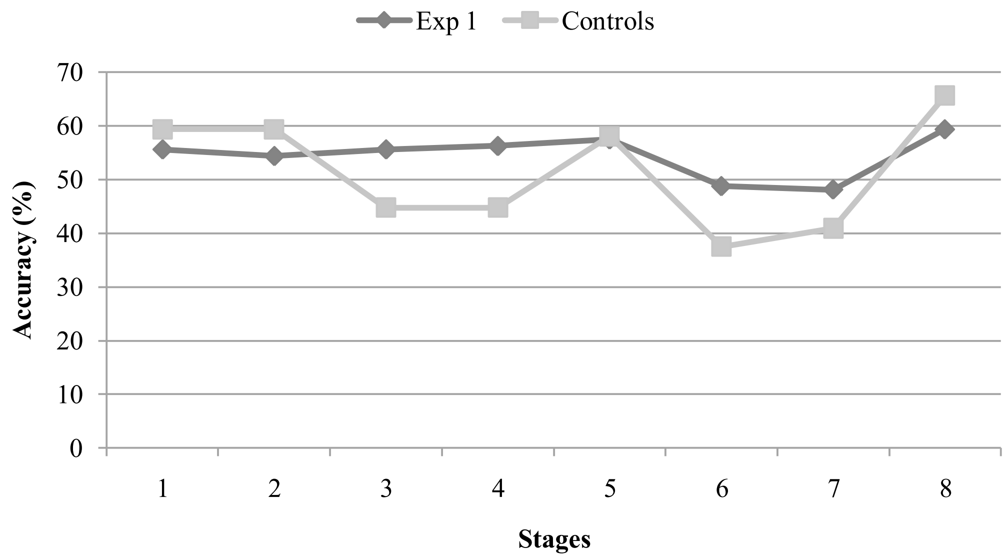

Figure 6. Classification accuracy of experimental and control subjects across eight chronological stages of the testing set.

Classification performance on grammatical and ungrammatical items

The following analyses report endorsement rates, rather than accuracy scores. The results showed that the experimental group endorsed 71.6% (*SD* = 17%) of all grammatical sentences and the control group 59% (*SD* = 14.5%). The difference between the two groups on grammatical sentences was significant, $F(1, 33) = 5.285$, $p < .05$, i.e. experimental subjects were significantly more likely to correctly endorse grammatical strings. The experimental group, $t(19) = 5.662$, $p < .001$, and the control group, $t(14) = 2.412$, $p < .05$, performed significantly above chance on grammatical items. The analysis also showed that the experimental group erroneously endorsed 62.3% (*SD* = 17.4%) of all ungrammatical sentences and the control group 55.5% (*SD* = 16.1%). The difference between the two groups on ungrammatical strings was not significant, $F(1, 33) = 1.376$, $p > .05$. Figure 7 illustrates the performance of the two groups on grammatical and ungrammatical items.

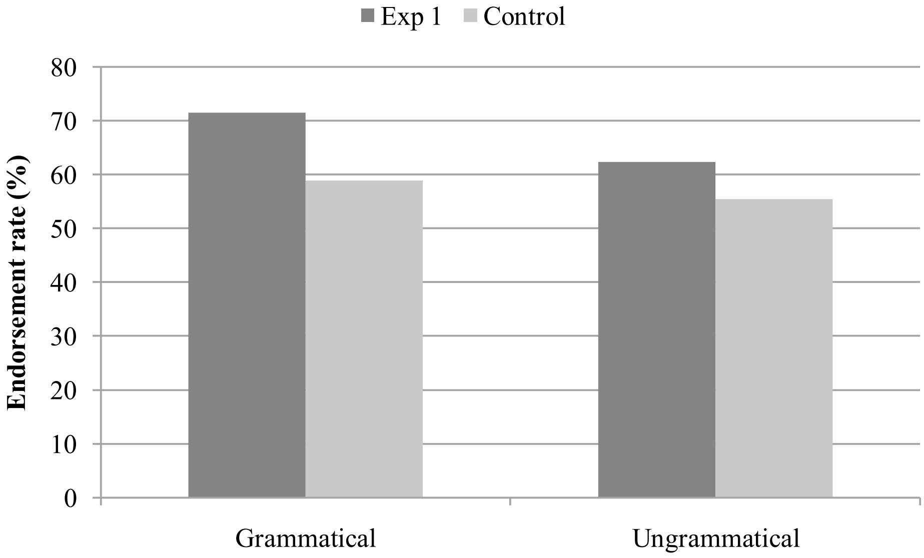

Figure 7. Classification performance of experimental and control subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

Classification performance across sentence types

Figure 8 illustrates participants' endorsement rates across the eight sentence types employed in this experiment. The mean values for both groups are summarized in table 5. As far as grammatical patterns are concerned, the analysis indicated that the experimental group was significantly more likely than the control group to correctly endorse V2-VF sentences, $F(1, 33) = 11.734$, $p < .05$, and VF-V1 sentences, $F(1, 33) = 6.146$, $p < .05$. The difference between the two groups on V2 and V2VF sentences was not significant. In the case of the ungrammatical patterns, there were no significant differences between the two groups. However, the difference on *VF-VF sentences approached significance, $F(1, 33) = 3.920$, $p = .056$.

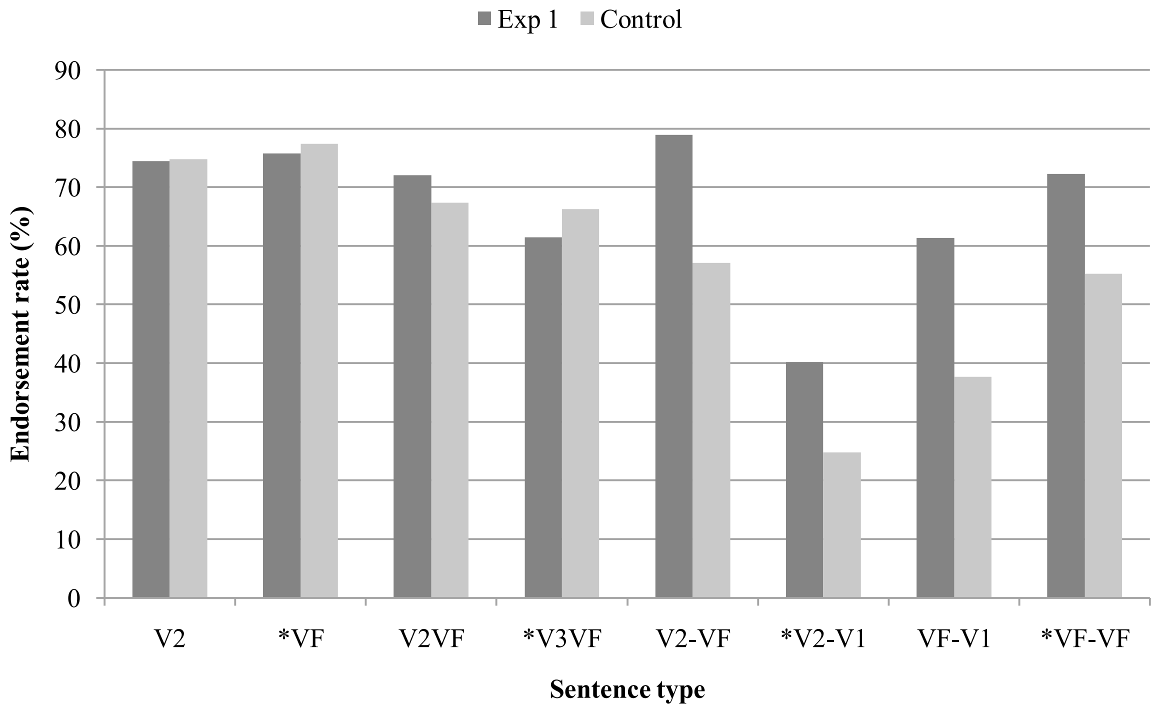

Figure 8. Classification performance of experimental and control subjects across sentence types. Bars represent endorsement rates.

Table 5
Mean Endorsement Rates (%), Standard Deviations, and Standard Error Values for Experimental and Control Participants across Sentence Types

		Sentence types							
Groups		V2	VF	V2VF	*V3VF	V2-VF	*V2-V1	VF-V1	*VF-VF
Exp 1	*M*	74.5*	75.8*	72.1**	61.5$^+$	79**	40.2	61.4$^+$	72.3**
	SD	35.2	27.8	21.3	22.6	17.7	30	29.2	20.6
	SE	7.9	6.2	4.8	5.1	4	6.7	6.5	4.6
Control	*M*	74.8**	77.4**	67.3*	66.3*	57.1	24.7*	37.7$^+$	55.2
	SD	19.9	27.6	16.7	21.4	20	22.5	26.3	30.5
	SE	5.1	7.1	4.3	5.5	5.2	5.8	6.8	7.9

Significance from chance: $^+p < 0.1$. * $p < .05$. ** $p < .001$.

2.3.2.2. Confidence ratings

Overall performance

The following analyses only focus on the results of the experimental group. The average confidence level in the experimental group was 5.9 (*SD* = 1.1), which suggests more high-confidence decisions than low-confidence decisions. In order to determine whether participants became more confident while performing on the testing phase, the confidence ratings were equally divided into eight chronological stages, and the average confidence levels were calculated for each stage. A repeated measures ANOVA showed no significant differences in mean confidence across the stages, $F(7, 133) = 1.019$, $p > .05$, i.e. the experimental group did not display significant changes in confidence levels while performing on the testing phase. In terms of proportion, experimental participants reported high confidence levels in 66.7% of their grammaticality decisions and low confidence levels in 33.3%. Experimental participants were significantly more accurate in high confidence decisions ($M = 56.4\%$) than in low confidence decisions ($M = 51.2\%$), $p < .05$.

Zero correlation criterion

The Chan difference score was computed in order to establish whether learning in the experimental group was implicit by the zero correlation criterion (Chan, 1992; Dienes et al., 1995). The average confidence for correct grammaticality decisions was 6.1 and the average confidence for incorrect decisions was 5.5, i.e. experimental participants were more confident in correct decisions than in incorrect ones. The difference (0.6) was significant, $t(19) = 2.141$, $p < .05$.

In order to apply the binary confidence technique developed by Kunimoto et al. (2001), the grammaticality judgments and the confidence ratings of the experimental group were converted into *d'* scores. In the case of the grammaticality judgments, a positive d'_{acc} value indicated a higher proportion of "yes" responses to grammatical stimuli and a negative d'_{acc} a higher proportion of "yes" responses to ungrammatical stimuli. Good discrimination between grammatical and ungrammatical stimuli is reflected in a positive d'_{acc} value and poor discrimination in a d'_{acc} value of zero or below. In the case of the confidence ratings, a positive d'_{conf} indicated more correct high-confidence decisions than incorrect ones. A negative d'_{conf}, on the other hand, suggested more incorrect high-confidence decisions than correct ones. Conscious knowledge is reflected in d'_{conf} values greater than zero; unconscious knowledge is reflected in d'_{conf} values close to or below zero. The average d'_{conf} score for the experimental group was .02 ($SD = .79$), which was not significantly different from zero, $t(19) = .116$, $p > .05$. Figure 9 plots d'_{conf} against d'_{acc} for all experimental subjects.

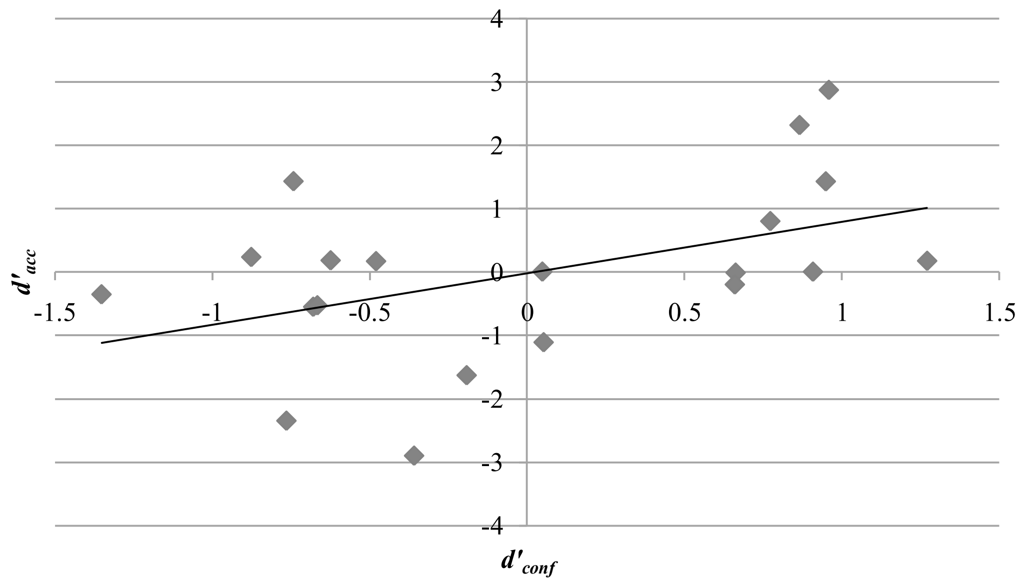

Figure 9. d'_{conf} and d'_{acc} scores. Each dot represents one experimental subject.

Dissociating aware and unaware participants

As figure 10 indicates, 10 of the experimental subjects had a negative d'_{conf} value (M = -0.7, SD = 0.3) while the other 10 had a positive d'_{conf} value (M = 0.7, SD = 0.4), $t(18) = 8.752$, $p < .0005$. Because a negative d'_{conf} characterizes subjects in which high confidence levels are not correlated with high accuracy scores, the group with a negative d'_{conf} score will be referred to as "unaware group" in subsequent analyses. In contrast, the group with the positive d'_{conf} score will be referred to as "aware group."

2.3.2.3. Classification performance of aware, unaware and control participants

Overall performance

The reanalysis of the grammaticality judgments showed that the aware group classified 60.3% (SD = 10.7%) of the test items correctly and the unaware group 49% (SD = 11.4%). A one-way ANOVA showed that there were significant differences between the aware, unaware and control groups, $F(2, 32) = 4.371$, $p < .05$. Post-hoc analysis with Tukey's Honestly Significant Differences (HSD) test showed that the aware group significantly outperformed both the unaware group, $p < .05$, and the control group, $p < .05$, which suggests that the training phase did produce a learning effect but only in those participants that developed conscious knowledge. The aware group also performed significantly above chance, $t(9) = 3.027$, $p < .05$.

Classification performance on grammatical and ungrammatical items

The performance of the three groups on grammatical and ungrammatical items is represented in figure 10. The results showed that the aware group endorsed 72.1% (*SD* = 16.3%) of grammatical sentences and 51.6% (*SD* = 14.9%) of ungrammatical sentences, $t(9) = 3.078$, $p < .05$. The unaware group endorsed 70.9% (*SD* = 18.5%) of grammatical sentences and 73% (*SD* = 12.9%) of ungrammatical sentences, $t(9) = .298$, $p > .05$. A one-way ANOVA further showed that there were significant differences among the three groups on both grammatical items, $F(2, 32) = 2.581$, $p < .05$, and ungrammatical items, $F(2, 32) == 6.024$, $p < .01$. Results of the post-hoc analysis using Tukey's HSD test showed that aware participants significantly outperformed controls on grammatical items, $p < .05$, but not on ungrammatical ones, $p > .05$, which suggests that learning in the aware group was largely associated with the endorsement of previously encountered syntactic patterns. There were no differences between aware and unaware subjects on grammatical items, but aware subjects were significantly less likely to reject ungrammatical items, $p < .05$. Finally, there were no significant differences between unaware and control subjects on grammatical strings, but unaware subjects were significantly more likely to endorse ungrammatical strings than the controls, $p < .05$. Figure 10 illustrates the performance of aware, unaware and control subjects across grammaticality.

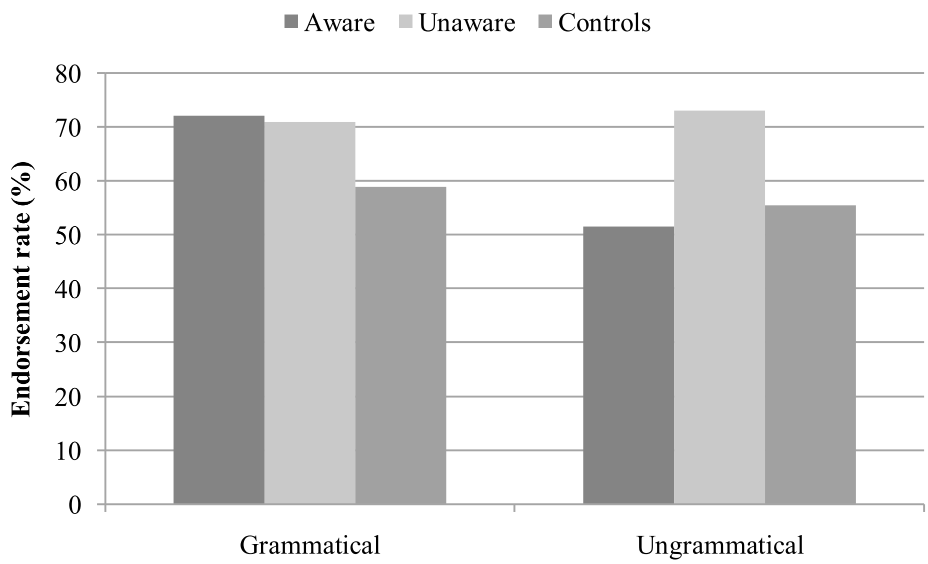

Figure 10. Classification performance of aware, unaware and control subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

Classification performance across sentence types

Figure 11 illustrates participants' endorsement rates across the eight syntactic patterns employed in this experiment. The scores are summarized in table 6. As has been mentioned, the superior performance of the aware group was associated with the ability to reject ungrammatical items. Further analysis showed that there were significant differences between the three groups on V2-VF sentences, $F(2, 32) = 0.789$, $p < .01$, and *VF-VF sentences, $F(2, 32) = 4.568$, $p < .05$. The effects approach significance on VF-V1 sentences, $F(2, 32) = 3.046$, $p = .06$, *VF sentences, $F(2, 32) = 2.704$, $p = .08$, and *V2-V1 sentences, $F(2, 32) = 2.588$, $p = .09$. The assumption of homogeneity of variance was violated for *VF-VF sentences, but the Brown-Forsythe F-ratio was nonetheless significant, $F(2, 29.9) = 5.488$, $p < .05$. Results of the post-hoc analyses, using Tukey's HSD test for V2-VF sentences and the Games-Howell procedure for *VF-VF sentences, showed that aware and unaware subjects were significantly more likely to endorse V2-VF sentences than the control, both $p < .05$, and that unaware subjects were significantly more likely to erroneously endorse *VF-VF sentences than either aware or control subjects, both $p < .05$. There was no difference between aware and unaware subjects on V2-VF sentences and no difference between aware subjects and controls on *VF-VF sentences.

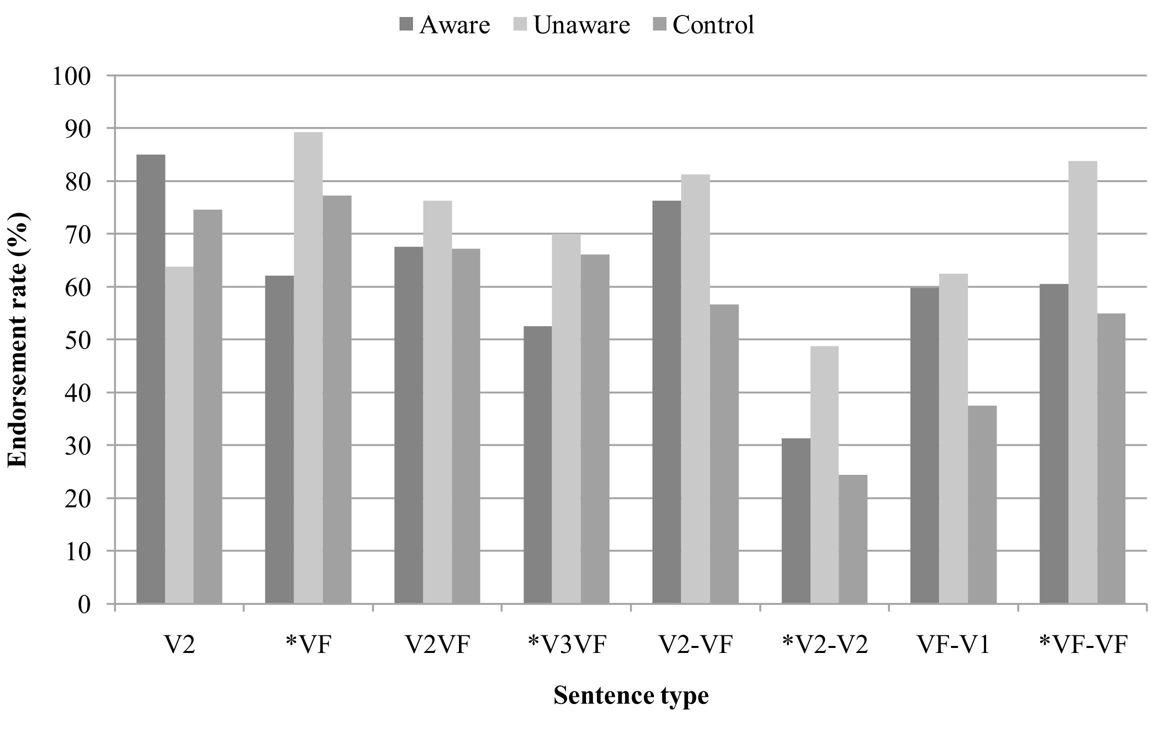

Figure 11. Classification performance of aware, unaware and control subjects across sentence types. Bars represent endorsement rates.

Table 6
Mean Endorsement Rates (%), Standard Deviations, and Standard Error Values for Aware, Unaware and Control Participants across Sentence Types.

		Sentence type							
Groups		V2	*VF	V2VF	*V3VF	V2-VF	*V2-V1	VF-V1	*VF-VF
Aware	*M*	85*	62.1	67.5^{+}	52.5	76.3*	31.3*	59.8	60.5
	SD	31.6	29.3	25.1	16.5	18.1	23.8	30.3	18.7
	SE	10	9.3	7.9	5.2	5.7	7.5	9.6	5.9
Unaware	*M*	63.8	89.3**	76.3*	70*	81.3**	48.8	62.5	83.8**
	SD	37	19.3	17.1	25.1	17.9	34.1	29.5	15.6
	SE	11.7	6.1	5.4	7.9	5.7	10.8	9.3	4.9
Controls	*M*	74.6**	77.3*	67.1*	66.1*	56.7	24.4*	37.5^{+}	55
	SD	19.8	27.5	16.7	21.4	20	22.4	26.3	30.5
	SE	5.1	7.1		5.5	5.2	5.8		7.9

Significance from chance: $^{+}p < 0.1$. * $p < .05$. ** $p < .001$.

2.3.3. Debriefing

2.3.3.1. Accuracy estimates

Since there was no evidence of learning in either the control or the unaware group, the following analysis only focuses on the aware group. The aware group's assessment of its performance on the grammaticality judgment task (*estimated accuracy*) was compared to its actual performance (*observed accuracy*). The results are summarized in table 7 below. The comparison showed that, within the experimental group, aware subjects overestimated their performance by 2.2% and unaware subjects by 5%. Again, these accuracy estimates were not significantly different from the observed accuracy levels. The analysis thus supports the view that the experimental subjects were at least partially aware of their knowledge of the grammaticality of strings.

Table 7
Estimated Accuracy and Observed Accuracy in the Aware Group (n = 10)

	Accuracy		
	Estimated	Observed	Difference
M	62.5	60.3	+2.2
SD	12.5	10.7	
SE	4.0	3.4	

2.3.3.2. Verbal reports

The transcription of all verbal reports can be found in the appendix. The analysis showed that all control subjects and 19 out of 20 experimental subjects verbalized knowledge when prompted, at the end of the experiment, to describe any rules or regularities they might have noticed. Participants did not rely on phrasal categories (NP, VP, etc.) in their reports but employed functional categories (subject, object, etc.) and semantic categories (time, action, etc.) instead. Importantly, the analysis showed that no subject was able to provide rule descriptions such as the ones outlined in table 1 above.

As far as the experimental group is concerned, the most frequently remarked regularity concerned the placement of verbs at the end of the sentence: Subjects 3, 5, 6, 7, 9, 10, 11, 18, and 19 mentioned that verbs could feature in sentence-final position but did not link this position to subordinate clauses. One subject (1) mentioned that verbs were sentence-final in "embedded clauses," despite the fact that the stimulus material did not contain any. However, since embedded clauses are subordinate clauses, this could suggest that the subject 1 correctly associated verb final position with subordinate clauses and simply used the wrong terminology. Four subjects (1, 6, 7, 16) commented on the fact that the verb could be placed in second position but did not state that this option was uniquely associated with main clauses that were not preceded by a subordinate clause. Three different subjects (3, 4, 11) noticed a "reversal" of subject-verb order. A different subject (4) commented on the fact that sentences were implausible when a fruit or a planet occurred. In sum, experimental subjects seemed aware that verbs could be placed in final or in second position. The fact that a verb could occupy first position was not mentioned. There was no evidence in the verbal reports suggesting that subjects were aware of the role of clause type and clause sequence for the placement of VPs in the semi-artificial language.

The reports of the control group provide an indication of what is salient to native speakers of English when exposed to sentences with a wide variety of word orderings. Like the experimental subjects, controls noticed the "reversal" of English word order: Subjects 5 and 8, for example, mentioned that subject and object frequently switched places, and subject 13 commented on the fact that "most sentences contained one major inversion; some contained two." Subjects 7, 8, 9 and 11 reported that the verb often came at the end. Several control subjects (2, 3, 5) mentioned that "shorter sentences", i.e. one-clause constructions, were easier to understand. Subjects 2 and 5 were more likely to endorse a short sentence as grammatical because they were easier to understand than "long jumbled sentences" (2).

2.4. Discussion

The results of experiment 1 indicate that adult learners are able to acquire natural language syntax under incidental learning conditions, while processing sentences for meaning, without the benefit of corrective feedback and after a relatively brief exposure period. The results also show that learners are able to transfer knowledge to stimuli with the same underlying structure but new surface features, which suggests that an abstract representation might have been derived from the original surface form. However, the findings provide no evidence for the unaware learning of syntax. The analysis of participants' confidence ratings suggests that learning was restricted to those subjects that became aware of the existence of knowledge that was guiding their grammaticality decisions. While learning in the experiments was incidental, i.e. a by-product of auditory exposure to the training sentences, there was no learning without awareness. Considering that language acquisition is frequently cited as a prime example of implicit learning (Berry & Dienes, 1993; Winter & Reber, 1994), these results are somewhat surprising.

One possible explanation is that all learning involves some degree of awareness. According to this view, there is no learning without awareness and previous experimental evidence is questioned because many researchers relied on verbal reports as measures of awareness (see Shanks & St. John, 1994; Shanks, 2005). Since it is possible to be aware of something while remaining unable to put it into words, it is argued that lack of verbalization provides no evidence for lack of awareness. Another possibility is that this experiment is more adequately described as a study of second, rather than first, language acquisition. While first language acquisition might proceed implicitly, i.e. incidentally and without awareness of the learning outcome, L2 acquisition might not. Several researchers (e.g., Bley-Vroman, 1988, 1989; Schachter, 1988) have argued that learning a first language is fundamentally different from learning a second one in many important aspects (see Gass & Selinker, 2001, for an overview). Like the participants in experiment 1, many L2 learners are adults, i.e. cognitively mature individuals who already possess knowledge of a full linguistic system, namely their native language. Adult learning styles are also different from those of children, in that adults are more likely to look for rules and to construct conscious hypotheses about the nature of the stimulus material. In addition, adults have a superior attention span and working memory capacity, which might facilitate the establishment of explicit knowledge. For these reasons, it would not be unexpected to find that learning in this experiment was associated with awareness. It would be interesting to determine whether a replication of the experiment with young children, ideally pre-schoolers with little metalinguistic knowledge, would provide similar results.

Assuming that unaware learning does indeed exist, it could also be argued that implicit learning is restricted to less complex learning situations. Gómez (1997), for example, investigated the role of complexity and abstractness in a series of AGL experiments. She found that implicit learning was primarily associated with the learning of first-order dependencies. In contrast, more complex types of learning (e.g. learning of second-order dependencies and transfer) were closely linked to explicit knowledge.[23] Since above-chance performance in experiment 1 required transfer to novel stimulus sentences, the fact that learning was limited to aware participants could be explained by the complexity of the learning task. Unaware learning might have been observed had the testing phase not involved transfer to novel sentences. Other aspects, which are related to the nature of the stimulus domain, might have encouraged the development of explicit knowledge. From a psycholinguistic perspective, for example, the verb can be considered the most prominent element of the sentence (Townsend & Bever, 2001). Since this experiment concentrated on the acquisition of a very noticeable syntactic element, it is perhaps not surprising that learning involved awareness. In addition, the fact that verbs often occurred close to the beginning or at the end of a sentence, further encouraged the establishment of conscious representations. Chan (1992, cited in Allwood et al., 2000) has observed, for example, that subjects in AGL experiments are more likely to develop conscious knowledge of those elements of a stimulus sequence that occur in anchor positions, i.e. at the beginning or at the end of a string. Many subjects in the present experiment reported noticing the occurrence of verbs at the beginning and, especially, at the end of a stimulus. The occurrence of verbs in the middle of sentences, as happened in the case of VF-V1 sentences, was not reported, which seems to support Chan's observation. Focusing on a less noticeable syntactic element in less prominent sentence positions might have resulted in unaware learning.

[23] Interestingly, the notion that implicit learning is limited to relatively simple learning processes is opposed to Reber's (1989, 1993) views. According to Reber, a complex stimulus domain is a prerequisite for the occurrence of implicit learning. If the rule-system is too simple, or perceived as being simple, subjects often engage in conscious learning strategies and implicit learning processes will not be observed.

Although a learning effect was observed in the aware group, the results also showed that subjects did not acquire the relevant rule-system. The acquisition of a rule-system is best reflected in the capacity of dissociating grammatical and ungrammatical strings categorically. In this experiment, for example, subjects should have been able to endorse or reject a sentence purely on the basis of the placement of the verb: if the VP occupied the appropriate position, a sentence should have been endorsed; if it did not, the sentence should have been rejected. This classification performance would have resulted in high endorsement rates for grammatical stimuli and low endorsement rates for ungrammatical stimuli. The analysis of the grammaticality judgments in this experiment showed that this was not the case. The aware group performed significantly above-chance on grammatical items but performed only at chance level on ungrammatical ones. The data suggests that, rather than basing decisions on rule knowledge, aware subjects relied heavily on memory for previously encountered patterns when making their grammaticality judgments. If a test stimulus matched a training pattern, as was the case with the grammatical test sequences, subjects were likely to endorse it. However, when a test stimulus did not resemble a training pattern, subjects had to rely on guessing.

Closer analysis of the judgment data also showed that learning in this experiment was somewhat limited. The aware group displayed a preference for V2 and V2VF sentences over their ungrammatical counterparts, *VF and *V3VF. However, control subjects displayed a similar preference, which makes it difficult to assess whether learning took place. In addition, aware subjects tended to overendorse stand-alone *VF, which suggests they did not know that VPs could only occur in final position in subordinate clauses. In the case of the complex sentences, aware subjects displayed a clear preference for V2-VF sentences over *V2-V1 but unfortunately the controls displayed the same pattern. Finally, aware subjects were just as likely to endorse VF-V1 sentences as *VF-VF sentences, which indicates that subjects had no knowledge of the V1 rule. In all, aware subjects seem to have learned two possible positions for VP (second and final) but not that clause type and clause sequence were also relevant cues for verb placement.

Several reasons might explain why there was limited learning in this experiment. Firstly, subjects might simply have received an insufficient amount of exposure for learning to take place. The training phase consisted of 128 instances of the semi-artificial grammar, which were repeated twice, and lasted approximately 40 minutes. Although some AGL experiments report learning effects after considerably briefer exposure periods, it is likely that subjects in natural language experiments require additional exposure. In contrast to AGL subjects, participants in this experiment possessed a rule-system that might have interfered with the grammar to be acquired. Considering the resilience of native language transfer errors in L2 acquisition, it seems natural that prolonged stimulus exposure might be required. Secondly, participants' metalinguistic knowledge might have distracted their attention from relevant verb placement cues. As the verbal reports show, subjects used metalinguistic terminology to describe the rules they might have noticed.[24] The availability of categories such as subject, verb and object might aid acquisition by increasing the likelihood that subjects will notice these elements in the input. On the other hand, it might also distract them from paying attention to linguistic notions which they do not have as readily available (such as clause type). It could also be that categories such as clause type or clause sequence are simply not perceived to be relevant elements of grammar and hence not noticed. Directing subjects' attention to suprasentential cues might well increase learning, and experiment 5 investigated to what extent this is this case. Finally, the limited learning might also be explained by the choice of training task. The plausibility judgment task is useful for training because it requires that participants focus on sentence meaning, which mirrors language acquisition outside the lab. However, there is a growing literature in SLA research that suggests that adult language acquisition benefits from training tasks that encourage both the processing of meaning and of "linguistic form", in this case of word order. Experiment 3 introduces a change to the training task in order to establish whether this would, in fact, facilitate learning.

One explanation for the lack of rule learning is that the training sentences did not fully represent the possible configurations of the semi-artificial grammar. As Reber (1993) pointed out, the formation of rule knowledge is encouraged by exposing subjects to the widest possible variety of grammatical configurations. For example, if subjects are to learn that the verb has to be placed in second position independently of the location of other phrases, they should be exposed to sentences that reflect this rule. Arguably, this was not the case in the present experiment. All V2, V2VF and V2-VF sentences, for instance, began with the sequence [NP_{subj}] > [VP] > [AP_{temp}] or with the sequence [AP_{temp}] > [VP] > [NP_{subj}], which could have suggested to participants that NP_{subj} and AP_{temp} played a relevant role in the grammar. (Subjects might think, for example, that the VP had to be surrounded by an expression of time and a subject and consequently reject all sentences in which this is not the

[24] Compare this, for example, with AGL experiments on musical structure, where the lack of available terminology makes it exceedingly hard for musically untrained (and often trained) subjects to report knowledge.

case.) In order to make it more evident that the other phrasal categories are both completely irrelevant to the placement of VP and not limited to first and third positions, it might have been more appropriate if the training sentences had fixed positions for the VPs (first, second and final) but a relatively free word order for the remaining phrasal elements. This would also have the added benefit of increasing the complexity of the stimulus sentence, which in turn might hinder the development of explicit knowledge (Reber, 1989, 1993). Experiment 2 investigated whether this modification to the training set would improve learning of the four verb placement rules.

On a methodological note, the results of the experiment confirm that relying on verbal reports as a measure of awareness is not sufficient. The verbal reports collected at the end of the experiment were helpful in determining what aspects of the semi-artificial grammar subjects had noticed. At the same time, however, reliance on verbal reports would not have permitted the separation of the experimental group into aware and unaware subgroups. The learning effect in the aware group would thus have gone unnoticed. The binary confidence ratings, on the other hand, appeared sufficiently sensitive to detect low levels of awareness but they provided little information about what subjects noticed and were able to verbalize. The findings of experiment 1 suggest that a battery of awareness measures, e.g. confidence ratings and verbal reports, might be more adequate for the study of implicit language learning.

3. Experiment 2

3.1. Introduction

Experiment 2, like the previous experiment, investigates the acquisition of verb placement rules of a semi-artificial language under incidental learning conditions. Participants in this experiment were first exposed to the grammar by listening to the training sentences and by judging their semantic plausibility. In the testing phase, participants judged the grammaticality of new sentences and indicated their confidence level for each judgment. In contrast to the previous experiment, however, the structure of the training stimuli was substantially diversified. In order to highlight the fact that only the VP had to be placed in a certain position, the training sentences in this experiment featured fixed positions for the VPs (first, second and final) but a relatively free word order for the remaining phrasal elements. It was predicted that this alteration to the training set might facilitate learning by making the link between the VP and its position more prominent. It was also hoped that the increased syntactic complexity of the training sentences would make it more difficult to learn the system consciously and therefore encourage implicit learning processes. The aim of experiment 2 was to establish whether the modifications to the training set outlined below had an effect on the acquisition of the four verb placement rules.

3.2. Method

3.2.1. Participants

Fifteen native speakers of English (11 women and 4 men, mean age = 25.4 years) were recruited for this experiment ("experimental group 2"). The data of the 15 untrained subjects from experiment 1 served as control data. All participants were students at the University of Cambridge at the time of the experiment. As far as their language background was concerned, all subjects were native speakers of English with no background in German or any other V2 language. Participants spoke, on average, three foreign languages (M = 2.7). Experimental group 2 and control group 1 did not differ significantly across age, $t(15.84) = .854$, $p > .05$, gender, $\chi^2(1, N = 28) = 0.57$, $p > .05$, occupation (student vs. nonstudent), $\chi^2(1, N = 28) = 2.485$, $p > .05$, and number of languages acquired, $\chi^2(5, N = 28) = 5.369$, $p > .05$. All subjects received £5 for participating in the experiment.

3.2.2. Stimulus material

The semi-artificial language described in the previous chapter was used in this experiment. The linguistic focus was again on four rules that determine the placement of verbs in the semi-artificial grammar (see table 1 above). New training sentences were drafted for this experiment and recorded in the same fashion as the stimuli in the previous experiment. The testing set was the same as the one employed in experiment 1.

3.2.2.1. Training set

The training set, which is reproduced in the appendix, consisted of 128 sentences and was subdivided into 64 plausible constructions and 64 implausible ones. In order to train the experimental group, 32 sentences were created for each verb placement rule. All sentences were in the past tense. In contrast to the previous training items, the stimuli in experiment 2 displayed more word order variation. Verb forms always occurred in first, second or final position, depending on the VP rule, but the other phrase types (NP, AP, PP) could occupy any of the remaining positions. Whereas in experiment 1, for example, most sentences began with NP$_{subj}$ or AP$_{temp}$, the present training stimuli also featured complements and nontemporal adjuncts in first phrasal position. Also, temporal expressions were not limited to adverbs in experiment 2. In addition to the adverbs used in the previous experiment (*today* and *usually*), experiment 2 also employed the NP *some time ago* and the PP *in the evening* to express temporality. These new lexical items further diversified the training sentences. For each rule, four templates were designed to serve as models in the generation of the stimulus items. The table below outlines the 16 template options.

Table 8

Syntactic Templates for the Four Rules. Frequencies in Training Set Are in Parentheses. Sample Sentences Are in Italics.

<table>
<tr><th>Rule</th><th>Templates</th><th></th></tr>
<tr><td>V2</td><td>[[NP|PP] > [VP] > [NP]$_{subj}$ > [AP|NP|PP]$_{temp}$ > [AP|NP|PP]]</td><td>(8)</td></tr>
<tr><td></td><td>A thousand pounds won Peter today in the lottery.</td><td></td></tr>
<tr><td></td><td>[[AP|NP|PP] > [VP] > [AP|NP|PP]$_{temp}$ > [NP]$_{subj}$ [NP|PP]]</td><td>(8)</td></tr>
<tr><td></td><td>In the casino jeopardized some time ago Jessica her savings.</td><td></td></tr>
<tr><td></td><td>[[AP|NP|PP]$_{temp}$ > [VP] > [AP|NP|PP] > [NP]$_{subj}$ [NP|PP]]</td><td>(8)</td></tr>
</table>

Today executed with efficiency Peter the plan.

[[NP]$_{subj}$ > [VP] > [NP|PP]] [AP|NP|PP]$_{temp}$ > [AP|NP|PP] (8)

Jack entertained his friends today with an interesting performance.

Split VP [[NP|PP] > [VP$_{aux}$] > [NP]$_{subj}$ > [AP|NP|PP]$_{temp}$ > [NP|PP] [VP$_{part}$]] (8)

Many shots has Brian usually during his matches defended.

[[NP|PP] > [VP$_{aux}$] > [AP|NP|PP]$_{temp}$ > [NP]$_{subj}$ > [NP|PP] > [VP$_{part}$]] (8)

At the club has usually Mike with complete strangers danced.

[[AP|NP|PP]$_{temp}$ > [VP$_{aux}$] > [NP|PP] > [NP]$_{subj}$ > [VP$_{part}$]] (8)

In the evening were for important discussions the ambassadors summoned.

[[NP]$_{subj}$ > [VP$_{aux}$] > [NP|PP] > [AP|NP|PP]$_{temp}$ > [VP$_{part}$]] (8)

Her collections were to a foundation some time ago donated.

V1 [[SUB > [NP]$_{subj}$ > [NP|PP] > [VP]] > [VP] > [NP]$_{subj}$ > [NP|PP] > [NP|PP]] (8)

Since the factory the river polluted, avoided Joanna contact with tap water.

[[SUB > [NP|PP] > [NP]$_{subj}$ > [VP]] > [VP] > [NP|PP] > [NP]$_{subj}$ > [NP|PP]] (8)

After the university press his book published, sent many copies Mike to his friends.

[[SUB > [NP|PP] > [NP]$_{subj}$ > [VP]] > [VP] > [NP]$_{subj}$ > [NP|PP] > [NP|PP]] (8)

After against the weak opponents her team lost, fired Sarah the manager for the first time.

[[SUB > [NP]$_{subj}$ > [NP|PP] > [VP]] > [VP] > [NP|PP] > [NP]$_{subj}$ > [NP|PP]] (8)

After the vending machine fifty pence withheld, complained about this Rose to the company.

VF [[NP]$_{subj}$ > [VP] > [AP|NP|PP]$_{temp}$ > [SUB > [NP]$_{subj}$ >[NP|PP] > [VP]]] (8)

Peter repeated today that the movers his furniture scratched.

[[NP]$_{sub}$ > [VP] > [AP|NP|PP]$_{temp}$ > [SUB > [NP|PP] > [NP]$_{sub}$ > [VP]]] (8)

Mike reckoned usually that about their classes his students cared.

[[AP|NP|PP]$_{temp}$ > [VP] > [NP]$_{subj}$ > [SUB > [NP]$_{subj}$ >[NP|PP] > [VP]]] (8)

Usually advised Brian that his students the vocabulary memorized.

[[AP|NP|PP]$_{temp}$ > [VP] > [NP]$_{subj}$ > [SUB > [NP|PP] > [NP]$_{subj}$ > [VP]]] (8)

Today reflected Jack that too few hours the fire department trained.

As in the previous experiment, all V2 sentences consisted of five phrasal units, and the VP always occupied second position. In contrast, the first, third and fourth positions could be occupied by NP$_{subj}$, AP, NP or PP. The final position featured either NP, AP or PP.

All Split VP sentences consisted of six phrasal units. The VP_{aux} always occurred in second and the VP_{part} in final position. The first, third and fourth positions could be occupied by NP_{subj}, NP, AP, or PP. All sentences based on patterns 1 and 2 were in the present perfect (auxiliary *have*), while all sentences based on pattern 2 were passive constructions (auxiliary *be*).

The sentences based on the V1 template consisted of a subordinate clause followed by a main clause. Subordinate constructions were marked as such by featuring a subordinating conjunction as an introductory word. Each subordinate clause featured three slots for phrasal elements: The last slot was always occupied by the VP whereas the first and the second slots could feature either NP_{subj}, NP or PP. The main clauses allowed the same phrase types, though with different ordering restrictions. The finite verb had to be placed in first position. The second and third positions could feature NP_{subj}, NP or PP. The final position could have an NP or a PP. No time expressions occurred in sentences for rule V1.

All sentences based on the VF templates consisted of a main clause followed by a subordinate clause. All main clauses comprised three phrasal units, namely NP_{subj}, VP and an expression of time (AP, NP, PP). As required by the V2 rule, the finite verbs were always placed in second position of the main clause. The first position was occupied by either NP_{subj} or an expression of time. If a sentence started with NP_{subj}, then a time expression followed in third position (patterns 1 and 2). In contrast, if a sentence began with an expression of time, then NP_{subj} followed in third position (patterns 3 and 4). Subordinate clauses also consisted of three phrasal elements. All subordinate clauses were clearly marked by featuring a subordinator at the beginning. As required by the VF rule, the finite verb was placed in final position. The first and second positions could be occupied by NP_{subj}, NP or PP.

The restrictions to the lexicon of the training set were identical to those in experiment 1. Each main verb occurred once in the training set and only multisyllabic verbs featured in clause-final position. The NP_{subj} in the main clauses was again realized by one of eight names (four male, four female), which were balanced for frequency. The same subordinators (*that*, *since* and *after*) were employed throughout the training set. Unlike experiment 1, however, there were four expressions of time in the new training set. As mentioned above, the NP *some time ago* and the PP *in the evening* were used in addition to the adverbs *today* and *usually* to express temporality.

A frequency analysis of the training set showed that the average stimulus length was 9.6 words per sentence in V2 constructions (10 for plausible items, 9.2 for implausible ones), 9.6 words per sentence for V2VF constructions (9.9 plausible, 9.3 implausible), 14 words per sentence for VF-V1 constructions (14.2 plausible, 13.8 implausible), and 10.5 for V2-VF constructions (10.4 plausible, 10.7 implausible).

3.2.2.2. Testing set

The test sentences were identical to those used in experiment 1. The sentences are reproduced in the appendix.

3.2.3. Procedure

The training and the testing procedures were the same as the ones in experiment 1. The instructions are reproduced in the appendix.

3.3. Results

In the section below and in subsequent chapters, the numbers after each experimental group refer to different experiments and not to different subject groups within an experiment. That is, "experimental group 1" refers to the experimental subjects in experiment 1, "experimental group 2" to experimental subjects in experiment 1, and so forth.

3.3.1. Training phase

Performance during the training phase was assessed by analyzing participants' plausibility judgments. In terms of frequency, participants committed, on average, 13.5 errors in their plausibility judgments. Because misjudged sentences were repeated, this suggests that the average experimental participant was exposed to 269.5 sentences during training. The analysis of the plausibility judgments showed that experimental group 2 judged 95% (SD = 2%) of sentences correctly, which was not significantly different from experimental group 1, $t(33) = 1.502$, $p > .05$. The additional scrambling in the training sentences did not affect accuracy on the plausibility judgment task. Participants in experiment 2 were slightly more accurate when judging plausible sentences (M = 95.7%, SD = 2.6%) rather than implausible ones (M = 94.4%, SD = 2.6%). The difference was not significant, however, $t(14) = 1.607$, $p > .05$.

3.3.2. Testing phase

3.3.2.1. Grammaticality judgments

Overall performance

The analysis of the grammaticality judgments showed that participants in this experiment classified 51.9% (SD = 7.1%) of the test items correctly. The overall performance of experimental group 2 was not significantly different from chance, $t(14) = 1.015$, $p > .05$. A one-way ANOVA showed no effects between groups, $F(2, 47) = .530$, $p > .05$. The results thus provide no evidence of learning.

In order to establish whether any improvements in accuracy occurred during the testing phase, the grammaticality judgment data was equally divided into eight chronological blocks. The mean accuracy for each block of eight judgments was then computed for experimental group 2. As figure 12 illustrates, the data did not show significant gains in the testing phase. A repeated measures ANOVA showed no significant differences across stages, $F(7, 98) = 1.151$, $p > .05$, i.e. no online learning effect was observed.

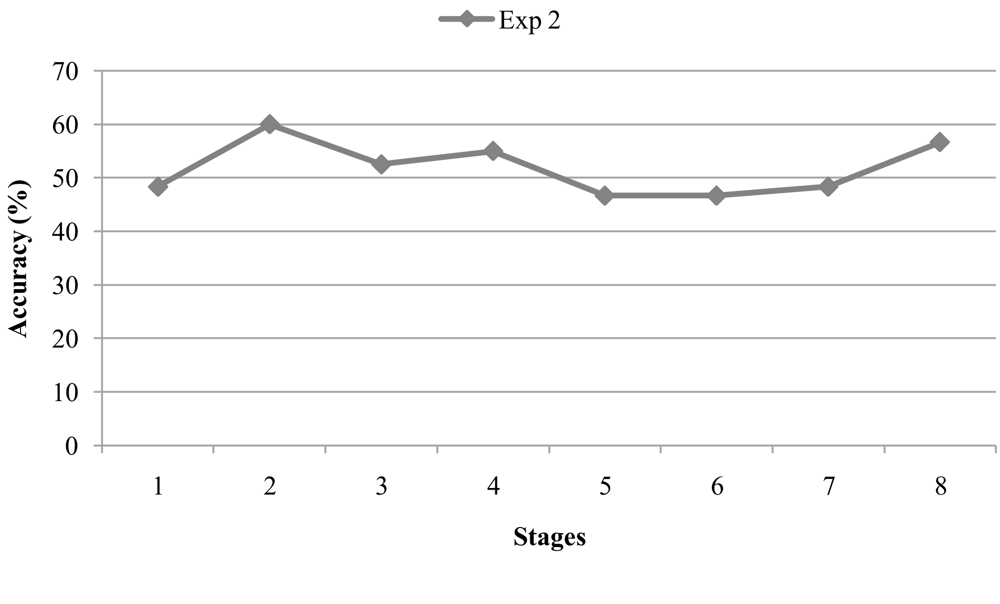

Figure 12. Classification accuracy of experimental group 2 across eight chronological stages of the testing set.

Classification performance on grammatical and ungrammatical items

The following analyses report endorsement rates. The results showed that experimental group 2 endorsed 69.8% of the grammatical stimuli, which was significantly above chance, $t(14) = 4.176$, $p < .001$. The difference between groups was not significant, but approaches significance, $F(2, 47) = 2.66$, $p = .08$. The results also indicated that experimental group 2 wrongly endorsed 66.1% of ungrammatical stimuli. Performance on ungrammatical items was significantly above chance, $t(14) = 4.050$, $p < .05$, but there were no significant effects between groups, $F(2, 47) = 1.587$, $p > .05$. Figure 13 illustrates the performance of the three groups on grammatical and ungrammatical items. Table 9 summarizes their mean endorsement rates.

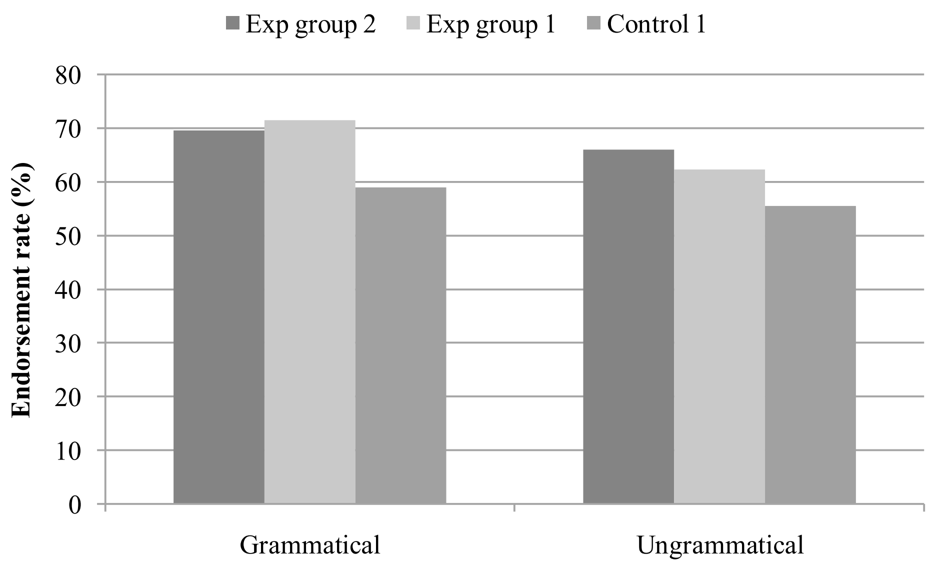

Figure 13. Classification performance of experimental and control subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

Table 9
Endorsements of Grammatical and Ungrammatical Sentences (%)

Group		Sentences	
		Grammatical	Ungrammatical
Exp 2	*M*	69.8**	66.1**
	SD	18.4	15.4
	SE	4.7	4
Exp 1	*M*	71.6**	62.3*
	SD	17	17.4
	SE	3.8	3.9
Control	*M*	59*	55.5
	SD	14.5	16.1
	SE	3.7	4.2

Significance from chance: * $p < .05$. ** $p < .001$.

Classification performance across sentence types

Figure 14 illustrates participants' endorsement rates across the eight sentence types employed in this experiment. The mean endorsement rates of the two experimental groups and the control group are summarized in table 10. As far as grammatical patterns are concerned, the analysis indicated that experimental group 2 scored significantly above chance on V2VF sentences, $t(14) = 6.361$, $p < .001$, V2-VF sentences, $t(14) = 2.543$, $p < .05$, and VF-V1 sentences, $t(14) = 3.790$, $p < .01$. In the case of ungrammatical patterns, experimental group 2 scored significantly above chance on *VF sentences, $t(14) = 3.338$, $p < .01$, *V3VF sentences, $t(14)$ 2.601, $p < .05$, and *VF-VF sentences, $t(14) = 4.103$, $p < .01$. A one-way ANOVA showed significant effects between groups on V2-VF sentences, $F(2, 47) = 3.956$, $p < .05$, VF-V1 sentences, $F(2, 47) = 6.706$, $p < .005$, and *V2-V1 sentences, $F(2, 47) = 4.128$, $p < .05$. The effect on *VF-VF sentences approaches significance, $F(2, 47) = 2.589$, $p = .086$. Post-hoc analysis with Tukey's HSD test indicated that experimental group 2 was significantly more likely than the control gorup to endorse VF-V1 sentences, $p < .005$, and significantly less likely to endorse *V2-V1 sentences, $p < .05$. There were no significant differences between experimental group 2 and experimental group 1 across sentence types.

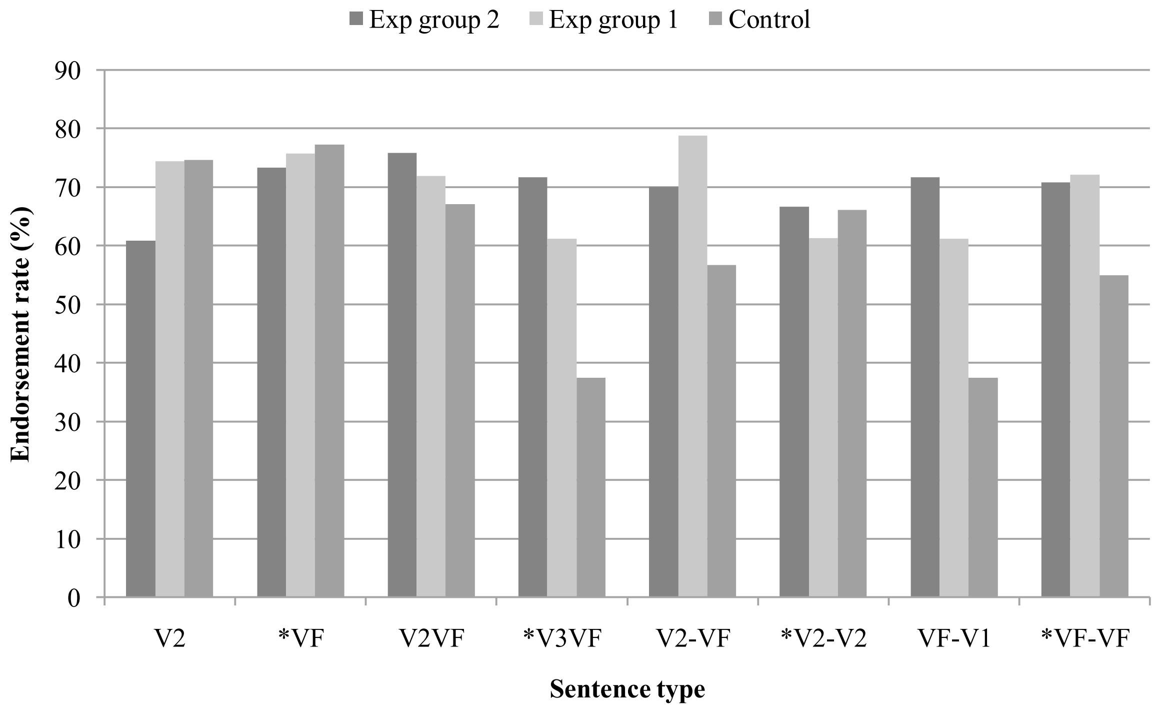

Figure 14. Classification performance of experimental and control subjects across sentence types. Bars represent endorsement rates.

Table 10
Mean Endorsement Rates (%), Standard Deviations, and Standard Error Values for Experimental and Control Participants across Sentence Types.

		Sentence type							
Group		V2	*VF	V2VF	*V3VF	V2-VF	*V2-V1	VF-V1	*VF-VF
Exp group 2	*M*	61	73.6*	76.2**	66.9*	70.3*	53.6	71.9*	71.1*
	SD	33.7	27.4	16	25.2	30.9	29.7	22.3	19.9
	SE	8.7	7.1	4.1	6.5	8	7.7	5.8	5.2
Exp group 1	*M*	74.5*	75.8*	72.1**	61.5^{+}	79**	40.2	61.4^{+}	72.3**
	SD	35.2	27.8	21.3	22.6	17.7	30	29.2	20.6
	SE	7.9	6.2	4.8	5.1		6.7	6.5	4.6
Control group	*M*	74.8**	77.4**	67.3*	66.3*	57.1	24.7*	37.7^{+}	55.2
	SD	19.9	27.6	16.7	21.4	20.0	22.5	26.3	30.5
	SE	5.1	7.1	4.3	5.5	5.2	5.8	6.8	7.9

Significance from chance: $^{+}p < 0.1$. * $p < .05$. ** $p < .001$.

3.3.2.2. Confidence ratings

Overall performance

The average confidence level in the experimental group was 5.3 (*SD* = 2.3), which indicates slightly more high-confidence decisions than low-confidence decisions. In order to determine whether participants became more confident while performing on the testing phase, the confidence ratings were equally divided into eight chronological stages, and the average confidence levels were calculated for each stage. A repeated measures ANOVA showed no significant differences in mean confidence across the stages, $F(7, 98) = 1.889$, $p > .05$, i.e. experimental group 2 did not display significant changes in confidence levels while performing on the testing phase. In terms of proportion, participants in this experiment reported high confidence levels in 56.7% of their grammaticality decisions and low confidence in 43.3%. Participants were more accurate in low confidence decisions (M = 52.1%) than in high confidence decisions (M = 51.1%), but this difference was not significant, $p > .05$. Participants did not score significantly above chance in either case, $p > .05$.

Zero correlation criterion

The Chan difference score was computed in order to establish whether learning in the experimental group was implicit by the zero correlation criterion. The average confidence for correct grammaticality decisions was 5.2 and the average confidence for incorrect decisions was 4.7, i.e. experimental participants tended to be more confident in correct decisions than in incorrect ones. The difference (0.5) was not significant, $t(14) = .762$, $p > .05$.

As in the previous experiment, Kunimoto et al.'s (2001) binary confidence technique was employed to further analyze the confidence rating data. Grammaticality judgments and confidence ratings were converted into *d'* scores. Good discrimination between grammatical and ungrammatical stimuli is reflected in a positive d'_{acc}, poor discrimination in a d'_{acc} close to or below zero. Conscious knowledge is reflected in d'_{conf} values greater than zero; unconscious knowledge is reflected in d'_{conf} values close to or below zero. The average d'_{conf} score for the experimental group was 2.81E-16 (SD = 0.4), which was not significantly different from zero, $t(14) = .000$, $p > .05$. Figure 15 plots d'_{conf} values against d'_{acc} scores for all experimental subjects.

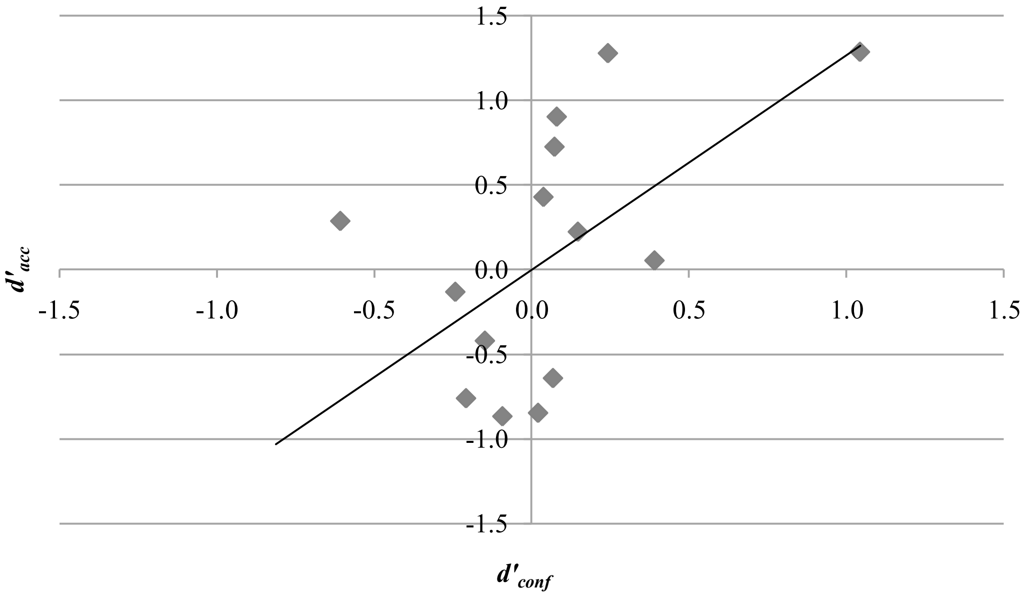

Figure 15. d'_{conf} and d'_{acc} scores. Each dot represents one experimental subject.

Dissociating aware and unaware participants

As figure 15 indicates, six of the experimental subjects had a negative d'_{conf} value (M = -0.4, SD = 0.3) while the other nine had a positive d'_{conf} value (M = 0.2, SD = 0.3). Because a negative d'_{conf} characterizes subjects in which high confidence levels are not correlated with high accuracy scores, the group with a negative d'_{conf} score will be referred to as “unaware group 2” in subsequent analyses. In contrast, the group with the positive d'_{conf} score will be referred to as “aware group 2.”

3.3.2.3. Classification performance of aware, unaware and control participants

Overall performance

The reanalysis of the grammaticality judgments showed that aware group 2 classified 55.2% (*SD* = 5.6%) of the test items correctly and unaware group 2 46.6% (*SD* = 6.4%). Aware group 2 scored significantly above chance, $t(8) = 2.774$, $p < .05$, in contrast to unaware group 2 and controls. A one-way ANOVA showed that there was a significant effect between groups in terms of overall performance, $F(2, 27) = 4.021$, $p < .03$. Post-hoc analysis using Tukey's HSD test indicated that aware group 2 significantly outperformed unaware group 2, $p < .05$, but not the controls. There were no differences between unaware group 2 and control subjects. The results suggest that the training phase only produced a learning effect in those participants that developed conscious knowledge. The analyses below will therefore concentrate on aware group 2. In comparison to the experimental subjects from the previous experiment, a one-way ANOVA showed a significant effect in terms of overall performance, $F(3, 31) = 3.85$, $p < .05$. Post-hoc analysis with Tukey's HSD test indicated that the only significant difference was between unaware group 2 and aware group 1, $p < .05$. Table 11 summarizes the overall performance of the different groups.

Table 11
Classification Performance (%) of Experimental and Control Subjects

Exp 2	Aware (n = 9)	*M*	55.2
		SD	5.6
		SE	1.9
	Unaware (n = 6)	*M*	46.6
		SD	6.4
		SE	2.6
Exp 1	Aware (n = 10)	*M*	60.3
		SD	10.7
		SE	3.4
	Unaware (n = 10)	*M*	49.0
		SD	11.4
		SE	3.6
Controls	(n = 15)	*M*	51.8
		SD	5.6
		SE	1.4

Significance from chance: * $p < .05$. ** $p < .001$.

Classification performance on grammatical and ungrammatical items

The results showed that aware group 2 endorsed 76% of grammatical sentences and 65.6% of ungrammatical sentences. Unaware group 2 endorsed 59.9% of grammatical sentences and 66.7% of ungrammatical sentences. A one-way ANOVA indicated that there was a significant effect between groups on grammatical items, $F(2, 27) = 3.572$, $p < .05$, but not on ungrammatical ones, $F(2, 27) = 1.631$, $p > .05$. A post-hoc analysis using Tukey's HSD test further showed that aware group 2 was significantly more likely to endorse grammatical items than the control group, $p < .05$. There were no differences between unaware and aware subjects. The results thus suggest that, in contrast to the previous experiment, the learning effect in the aware group was due to better endorsement of grammatical items.
In comparison to previous experimental subjects, a one-way ANOVA showed that there was a significant effect between groups on ungrammatical items, $F(3, 31) = 3.717$, $p < .05$, but not on grammatical ones, $F(3, 31) = 1.09$, $p > .05$. Post-hoc analysis by means of Tukey's HSD test indicated, however, that this effect related to a significant difference between aware group 1 and unaware group 1. There were no other significant differences across grammaticality between aware and unaware subjects in the two experiments. The performance of the groups on grammatical and ungrammatical items is represented in figure 16. Table 12 summarizes the endorsement rates of aware, unaware and control subjects.

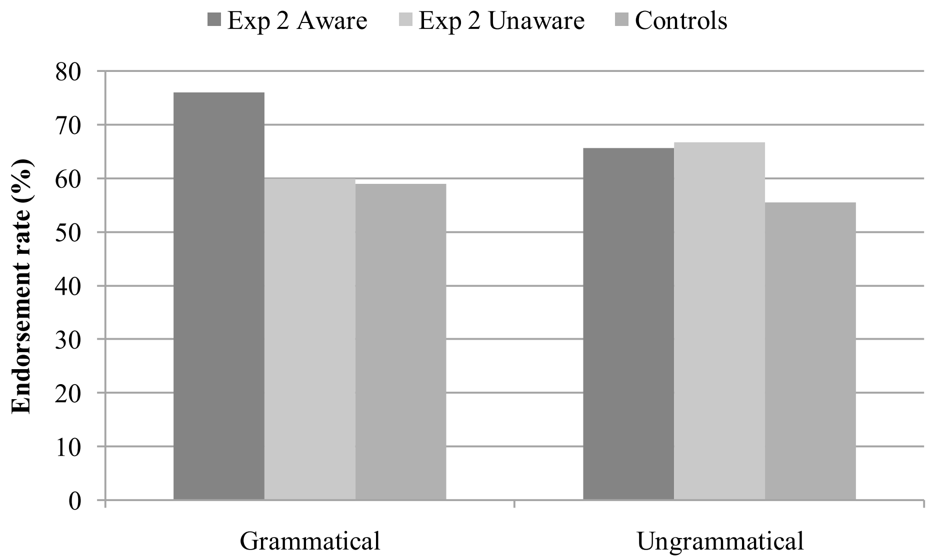

Figure 16. Classification performance of aware, unaware and control subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

Table 12
Endorsement Rates (%) for Aware, Unaware and Control Participants across Grammaticality

Group		Sentences	
		Grammatical	Ungrammatical
Exp 2 Aware	*M*	76**	65.6*
	SD	12.2	17.4
	SE	4.1	5.8
Exp 2 Unaware	*M*	59.9	66.7*
	SD	22.8	12.8
	SE	9.3	5.2
Exp 1 Aware	*M*	72.1*	51.6
	SD	16.3	14.9
	SE	5.1	4.7
Exp 1 Unaware	*M*	70.9*	73**
	SD	18.5	12.9
	SE	5.9	4.1
Controls	*M*	58.9*	55.5
	SD	14.4	16.1
	SE	3.7	4.2

Significance from chance: * $p < .05$. ** $p < .001$.

Classification performance across sentence types

Figure 17 illustrates the endorsement rates across the eight syntactic patterns employed in this experiment. The scores are summarized in table 13. A one-way ANOVA showed a significant effect on V2 sentences, $F(2, 27) = 4.763$, $p < .05$, VF-V1 sentences, $F(2, 27) = 7.387$, and *V2-V1 sentences, $F(2, 27) = 4.38$, $p < .05$. In the case of V2 sentences, however, the assumption of homogeneity of variance was violated, and the Brown-Forsythe *F*-ratio failed to reach significance, $F(3, 9.683) = 3.409$, $p = .076$. A post-hoc analysis with Tukey's HSD test indicated that aware group 2 was significantly more likely than the controls to endorse VF-V1 sentences, $p < .05$, and significantly more likely to erroneously endorse *V2-V1 sentences, $p < .05$. Furthermore, a one-way ANOVA showed that there were no significant differences aware and unaware groups from experiments 1 and 2, though the effects on V2 sentences, $F(3, 31) = 2.636$, $p = .067$, and *VF-VF sentences, $F(3, 31) = 2.627$, $p = .068$, approach significance.

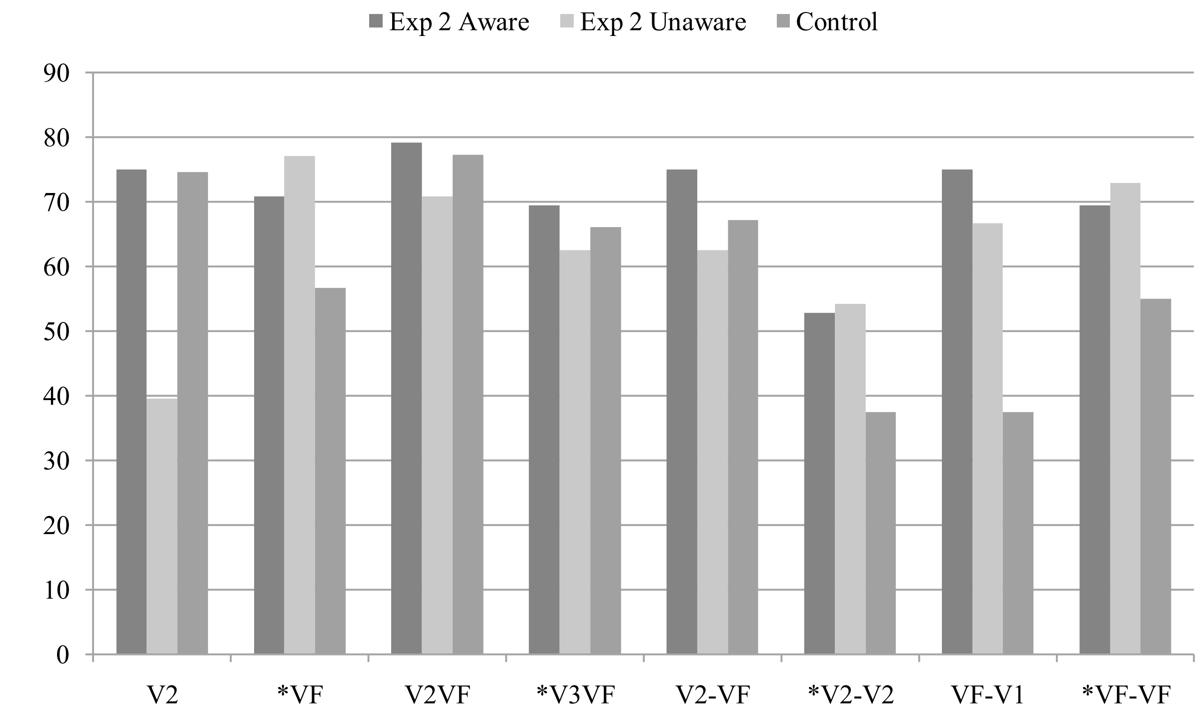

Figure 17. Classification performance of aware, unaware and control subjects across sentence types. Bars represent endorsement rates.

Table 13

Endorsement Rates (%) for Aware, Unaware and Control Participants across Sentence Types

		Sentence type							
		V2	*VF	V2VF	*V3VF	V2-VF	*V2-V1	VF-V1	*VF-VF
Exp 2 Aware	*M*	75*	70.8^{+}	79.2*	69.4^{+}	75^{+}	52.8	75*	69.4*
	SD	21.7	31.3	17.7	30	33.1	27.1	19.8	22.6
	SE	7.2	10.4	5.9	10	11	9	6.6	7.5
Exp 2 Unaware	*M*	39.6	77.1*	70.8*	62.5	62.5	54.2	66.7	72.9*
	SD	39.1	22.9	12.9	17.7	28.5	35.9	27	16.6
	SE	15.9	9.4	5.3	7.2	11.6	14.7	11	6.8
Exp 1 Aware	*M*	85*	62.1	67.5^{+}	52.5	76.3*	31.3*	59.8	60.5
	SD	31.6	29.3	25.1	16.5	18.1	23.8	30.3	18.7
	SE	10	9.3	7.9	5.2	5.7	7.5	9.6	5.9
Exp 1 Unaware	*M*	63.8	89.3**	76.3*	70*	81.3**	48.8	62.5	83.8**
	SD	37.0	19.3	17.1	25.1	17.9	34.1	29.5	15.6
	SE	11.7	6.1	5.4	7.9	5.7	10.8	9.3	4.9
Controls	*M*	74.6**	77.3*	67.1*	66.1*	56.7	24.4*	37.5^{+}	55
	SD	19.8	27.5	16.7	21.4	20.0	22.4	26.3	30.5
	SE	5.1	7.1	4.3	5.5	5.2	5.8	6.8	7.9

Significance from chance: $^{+}p < 0.1$. * $p < .05$. ** $p < .001$.

3.3.3. Debriefing

3.3.3.1. Accuracy estimates

Since there was no evidence of learning in either the control group or in unaware group 2, the following analysis only focuses on the aware group 2. The latter group's assessment of its performance on the grammaticality judgment task (*estimated accuracy*) was compared to its actual performance (*observed accuracy*). For aware group 2, the comparison showed that the difference between estimated and observed accuracy (+ 0.8%) failed to reach significance, $t(14) = .142$, $p > .05$, which suggests that subjects were aware of the knowledge they had acquired.

3.3.3.2. Verbal reports

The transcription of all verbal reports can be found in the appendix. All subjects reported knowledge when prompted, at the end of the experiment, to describe any rules or regularities they might have noticed. The analysis showed that most participants did not rely on phrasal categories such as NP, PP or VP in their reports, employing functional categories (such as subject, object, etc.) and semantic categories (such as time or action) instead. It also showed that no subject was able to provide rule descriptions such as the ones outlined in table 1 above. Subjects were unaware of the knowledge that determined word order in the semi-artificial language and aware of the fact that they did not know the relevant system.

The most frequently remarked regularity concerned the placement of verbs at the end of the sentence: Subjects 5, 7, 8, 10, 11 and 15 mentioned that verbs were placed at the end of the sentence but did not associate this position with subordinate clauses. Subjects 3 and 9 commented on the fact that verb phrases were often split up into auxiliary and participle forms. Subject 3 stated that he used the V2VF template as a model to guide his grammaticality decisions. Subjects 13 and 15 mentioned that the auxiliary form (*have, be*) could not occur at the end of a sentence. Subject 2 erroneously reported that verbs must not end sentences and that only NPs or APs could be placed in sentence-final position. Subject 15 reported that he focused on the position of the verb to make grammaticality decisions.

Five subjects (3, 5, 7, 9, 12) mentioned that time expressions could come at the beginning. Subjects 5 and 10 mentioned that subjects could occur at the beginning. Only one subject (11) stated that verbs could also be placed "close to the beginning", which could suggest that she was aware of V2 placement. Subject 8 noticed that subordinate clauses came first. No other reference to clause types (main vs. subordinates) was made. Subject 6 mentioned that "sections", as opposed to individual words, were moved, which suggests that she was aware that phrases, and not just lexical items, were moved. Subject 1 commented on the fact that the word order made the sentences sound like a foreign language.

In sum, experimental subjects seemed aware that verbs could be placed in final position. There is no evidence to suggest that subjects were aware of the fact that a verb could occupy first position. Just one subject commented on the fact that verbs could sometimes be placed in second position. In terms of possible verb locations, the experimental group only seemed aware of VF placement but not of V2 and V1. As in the previous experiment, there was no evidence in the verbal reports suggesting that subjects were aware of the role of clause type and clause sequence for the placement of VP. Despite the fact that the training items in experiment 2 were more scrambled, participants in this experiment made similar observations as participants in experiment 1.

3.4. Discussion

The results of experiment 2 confirm that adult learners are able to acquire the syntactic structure of a natural language without intending to, while processing sentences for meaning, without the benefit of feedback and after a brief exposure period. The results also show that learners are able to transfer knowledge to stimuli with the same underlying grammar but new surface features. However, the findings provide no evidence for the unaware learning of natural language syntax. The analysis of confidence ratings suggests that learning was restricted to those subjects that became aware of the knowledge that was guiding their grammaticality decisions. In addition, the findings also show that the alterations to the training set had a negative effect on learning. Although the use of more complex sentences did not affect participants' ability to perform on the plausibility judgment task in the training phase, the analysis of the grammaticality judgments clearly showed that it reduced learning. Contrary to expectations, highlighting the association between the VPs and their respective positions by presenting subjects with sentences with a greater syntactic variety was not beneficial.

There are several possibilities for explaining why learning was affected negatively by the changes made to the training items. On the one hand, the results could be explained by the fact that subjects in experiment 2 received less exposure to those grammatical patterns that reoccurred in the testing phase. As mentioned, the sentences used in the testing phase were the same for trained subjects in experiments 1 and 2. Participants in experiment 2, however, received less exposure to those grammatical sentence patterns that would reoccur during the testing phase. While in experiment 1 the seven grammatical test patterns (see table 2) had been presented 32 times during training, in the case of experiment 2 the same test patterns only occurred 16 times during training. (The patterns, not the actual sentences, were repeated.**)** This is because experiment 2 featured twice as many training patterns as experiment 1, which implied that each pattern had to be repeated less frequently in order to maintain the same number of training sentences for both experiments (128). It could be argued that more learning would have been observed if all 16 training patterns of experiment 2 had reoccurred during testing. On the other hand, the increased variation in the word order of the training stimuli could have hindered the formation of chunk knowledge and thus affected performance on the grammaticality judgment task. As outlined above, the training set of the present experiment had fewer repetitions of the same patterns, which made it less likely that subjects would register specific chunks or phrases. Whereas in experiment 1, for example, subjects appeared to have noticed chunks such as [NP_{subj} > VP > AP_{temp}] or [AP_{temp} > VP > NP_{subj}], the verbal reports suggest that this was not the case in experiment 2. If, as argued by Perruchet & Pacton (1990), knowledge of chunks plays an important role in the classification of new test strings, then it is perhaps not surprising that subjects in experiment 2 underperformed in the testing phase. In experiment 3, the training set features less word order variation but the lexical diversity is maintained.

In terms of methodology, the combination of confidence ratings and verbal reports was again found to be useful for measuring awareness. At the same time, however, it was felt that the confidence ratings would benefit from the introduction of a ‚guess' category, which would allow the observation of the guessing criterion (Dienes et al., 1995). The experiment would also have benefited from featuring source attributions in addition to confidence ratings. As discussed above, confidence ratings assess the conscious/unconscious state of judgment knowledge, whereas source attributions assess the conscious/unconscious state of structural knowledge. Both methodological changes – the introduction of a guess category to the confidence ratings and the introduction of source attributions – were made for experiment 3. The following experiment will also introduce a change to the way subjects are trained on the semi-artificial grammar. In addition to judging semantic plausibility, subjects in experiment 3 will be asked to repeat each sentence after a delayed prompt (*elicited imitations*). As mentioned in the previous chapter, there is a growing literature in SLA research that suggests that adult language acquisition benefits from training tasks that encourage both the processing of meaning and of linguistic form (see Doughty, 2003; R. Ellis, 2001). It is assumed that the combination of elicited imitations, which require the retention of word order, and plausibility judgments, which require processing of sentence meaning, could facilitate learning of the verb placement rules of the semi-artificial system.

4. Experiment 3

4.1. Introduction

Like the previous two experiments, experiment 3 investigates the acquisition of verb placement rules of a semi-artificial language under incidental learning conditions. Several changes were made to the methods of the previous experiments. Firstly, only three verb placement rules determined the placement of VP in this experiment. Secondly, experimental subjects were trained on the grammar by means of elicited imitations and plausibility judgments. Finally, the testing phase was modified by adding a guess category to the confidence ratings and by adding source attributions as additional measures of awareness. The aims of this experiment were (i) to establish whether the addition of a training task that required subjects to process word order more directly (through elicited imitations) had a positive effect on learning and (ii) to evaluate the usefulness of source attributions for measuring awareness.

4.2. Method

4.2.1. Participants

Thirty native speakers of English (22 women and 8 men, mean age = 24.3 years) were recruited to take part in this experiment and evenly distributed into experimental and control conditions. The majority of participants (28) were students at the University of Cambridge at the time of the experiment. As far as their language background was concerned, five participants had been brought up bilingually, i.e. they had learned a second language in addition to English during childhood and attained native-like proficiency in it. Participants spoke, on average, 1.6 foreign languages and none had a background in German or any other V2-language. Experimental and control groups did not differ significantly across age, $t(13.95) = 1.629$, $p > .05$, gender, $\chi^2(1, N = 30) = 0.682$, $p > .05$, occupation (student vs. nonstudent), $\chi^2(1, N = 30) = 2.143$, $p > .05$, and number of languages acquired, $\chi^2(3, N = 30) = 5.369$, $p > .05$. All subjects received £5 for participating in the experiment.

4.2.2. Stimulus material

As in experiments 1 and 2, a semi-artificial grammar, consisting of English words and German syntax, was used to generate the stimulus material. In contrast, however, the linguistic focus was on only three verb placement rules, namely V2, VF, and V1. The Split VP rule of the previous experiments was removed because it was felt that the presence of two rules associated with VPs in final positions – Split VP and VF – might hinder learning. The rule-system in this experiment was somewhat simpler in that the placement of VP was only dependent on the type of clause (main vs. subordinate) and on the type of clause sequence (main-subordinate vs. subordinate-main). The type of predicate was no longer relevant for verb placement. Table 14 illustrates the three rules in question.

Table 14

Descriptions and Examples of the Three Verb Placement Rules

Rule	Description	Examples
V2	Finite verb placed in second phrasal position of main clauses that are not preceded by a subordinate clause.	*In the evening ate Rose excellent dessert at a restaurant.*
V1	Finite verb placed in first position in main clauses that are preceded by a subordinate clause	*Since his teacher criticism voiced, put Chris more effort into his homework.*
VF	Finite verb placed in final position in all subordinate clauses	*George repeated today that the movers his furniture scratched.*

Rules V2 and V1 both applied to main clauses. They differed in that the former rule applied to main clauses that were not preceded by a subordinate clause and the latter to main clauses that were preceded by a subordinate clause. In contrast, rule VF only applied to subordinate clauses, irrespectively of whether a main clause preceded or followed. A total of 180 sentences were drafted for this experiment. The sentences were read out by a male native speaker of British English, digitally recorded on a Sony Mini-Disc player (MZ-R700) and subsequently edited with sound processing software (Audacity, version 1.2.4). The stimulus sentences were divided into a training and a testing set.

4.2.2.1. Training set

The training set, which is reproduced in the appendix, consisted of 120 sentences and was subdivided into 60 plausible constructions and 60 implausible ones. That is, half of the sentences were syntactically correct but expressed semantically implausible propositions. Plausible and implausible items were again designed so that participants would have to process the entire auditory string before being able to judge plausibility. In most instances, the final phrase of the sentence would reveal whether the sentence was plausible or not.

In order to train the experimental group, 40 sentences were created for each verb placement rule. All sentences were in the past tense. For each rule, templates were designed to serve as models in the generation of the stimulus items. The table below outlines the template options for the three rules.

Table 15

Syntactic Templates for the Three Rules. Frequencies in Training Set Are in Parentheses. Sample Sentences Are in Italics.

Rule	Templates	
V2	[[NP]$_{subj}$ > [VP] > [AP|NP|PP]$_{temp}$ > [AP|NP|PP] > [AP|NP|PP]]	(20)
	Brian played usually an important part in the school plays.	
	[[AP|NP|PP]$_{temp}$ > [VP] > [NP]$_{subj}$ > [AP|NP|PP] > [AP|NP|PP]]	(20)
	In the evening ate Rose excellent dessert at a restaurant.	
V1	[[[SUB > [NP]$_{subj}$ > [NP|PP] > [VP]] > [VP] > [NP]$_{subj}$ > [NP|PP] > [NP|PP]]	(40)
	After the university press his book published, sent George many copies to his friends.	
VF	[[NP]$_{subj}$ > [VP] > [AP|NP|PP]$_{temp}$ > [SUB > [NP]$_{subj}$ > [NP|PP] > [VP]]]	(20)
	Jack asserted in the evening that his father interesting appliances created.	
	[[AP|NP|PP]$_{temp}$ > [VP] > [NP]$_{subj}$ > [SUB > [NP]$_{subj}$ > [NP|PP] > [VP]]]	(20)
	In the evening explained Rose that the profit below the estimate remained.	

In the case of the V2 rule, all sentences consisted of five phrasal units. The VP was placed in second position, while the first and the third positions of the clause featured an NP$_{subj}$ and a temporal expression. The latter was instantiated either as an adverb phrase (AP$_{temp}$), a noun phrase (NP$_{temp}$) or a prepositional phrase (PP$_{temp}$). Half of the sentences featured NP$_{subj}$ in preverbal position and the other half an expression of time. The fourth and fifth phrasal positions could be occupied by either AP, NP or PP. The syntactic pattern for this rule, based on the position of the verb, was "V2."

The sentences based on the V1 template consisted of a main and a subordinate clause, with the subordinate always preceding the main clause. Subordinate constructions featured three slots for phrasal elements: The first and the last slot were always occupied by NP_{subj} and VP, respectively, while NP and PP were options for second position. The main clauses allowed the same type of elements, though with different ordering restrictions. Here, the finite verb had to be placed in first position with NP_{subj} following in second. NP and PP were both possibilities for third and fourth positions. No time expressions occurred in sentences for the V1 rule. The subordinators *since* and *after* were used throughout the training set. The syntactic pattern for this rule was "VF-V1."

All sentences based on the VF templates also consisted of a main and a subordinate clause. In contrast to the previous sentences, however, the main clause always preceded the subordinate clause. The main clauses were invariably composed of three phrasal units, namely NP_{subj}, VP and an expression of time (AP, NP or PP). Again, half of the sentences started with NP_{subj}, while the other half featured a temporal expression in first position. The subordinate clauses allowed for three additional phrasal elements. The VF rule stated that the finite verb in subordinate clauses had to be placed in final position, which left NP_{subj} to occupy the clause-initial position and either NP or PP to occupy the second position. The subordinator *that* was used in all cases. The syntactic pattern, based on the positioning of the VP in the two clauses, was "V2-VF."

Several restrictions applied to the lexicon of the training set. Each verb occurred once in the training set and only multisyllabic verbs featured in clause-final position. The NP_{subj} of the main clauses was realized as one of eight names, five of which were male (*Chris, Brian, Jack, Mike, George*) and five female (*Joanna, Sarah, Rose, Jessica, Cate*). Each name occurred 12 times in the training set. Time expressions were limited to five options, namely the APs *today* and *usually*, the NPs *a few months ago* and *last June*, and the PP *in the evening*, each occurring with equal frequency in the training set. In the case of the subordinate clauses, three subordinators were employed. All V2-VF sentences featured the subordinator *that*; half of the VF-V1 sentences featured the subordinator *since* and the other half the subordinator *after*.

A frequency analysis of the training set showed that the average sentence length was 9.7 words per sentence in V2-constructions (9.9 for plausible items, 9.5 for implausible ones), 12.9 for VF-V1 constructions (13.2 plausible, 12.7 implausible), and 10.8 words per sentence for V2-VF constructions (10.6 plausible, 11 implausible).

4.2.2.2. Testing set

The testing set, which is reproduced in the appendix, consisted of 60 new sentences and was subdivided into 30 grammatical and 30 ungrammatical items. Twelve sentence templates were created in order to test whether participants had acquired knowledge about the possible positions for VP in simple and complex sentences. Forty sentences were created for each syntactic pattern. Table 16 outlines the grammatical sentence templates.

Table 16

Sentence Templates for the Three Grammatical Patterns. Frequencies in the Testing Set Are in Parentheses. Sample Sentences Are in Italics.

<table>
<tr><th>Pattern</th><th>Template</th><th></th></tr>
<tr><td>V2</td><td>[[AP|NP|PP]$_{temp}$ > [VP] > [NP]$_{subj}$ > [AP|NP|PP] > [AP|NP|PP]]</td><td>(5)</td></tr>
<tr><td></td><td>Yesterday scribbled David a long letter to his family.</td><td></td></tr>
<tr><td></td><td>[[NP]$_{subj}$ > [VP] > [AP|NP|PP]$_{temp}$ > [AP|NP|PP] > [AP|NP|PP]]</td><td>(5)</td></tr>
<tr><td></td><td>Jessica gambled last June with her savings at the casino.</td><td></td></tr>
<tr><td>V2-VF</td><td>[[NP]$_{subj}$ > [VP] > [AP|NP|PP]$_{temp}$ > [SUB > [NP]$_{subj}$ > [NP|PP] > [VP]]]</td><td>(5)</td></tr>
<tr><td></td><td>David speculated yesterday that the suspect from prison escaped.</td><td></td></tr>
<tr><td></td><td>[[AP|NP|PP]$_{temp}$ > [VP] > [NP]$_{subj}$ > [SUB > [NP]$_{subj}$ > [NP|PP] > [VP]]]</td><td>(5)</td></tr>
<tr><td></td><td>In the afternoon acknowledged David that her children to England moved.</td><td></td></tr>
<tr><td>VF-V1</td><td>[[SUB > [NP]$_{subj}$ > [AP|NP|PP]$_{temp}$ > [NP|PP] > [VP]] > [VP] > [NP]$_{subj}$ > [NP|PP] > [NP|PP]]</td><td>(10)</td></tr>
<tr><td></td><td>When his wife in the afternoon the office left, prepared Jim dinner for the entire family.</td><td></td></tr>
</table>

The ungrammatical templates were similar to the grammatical ones with the exception that the position of the VP was incorrect. In order to test whether participants had acquired knowledge about the relevance of clause type in the placement of VP, their performance on V2 and *VF constructions was observed. While participants should accept V2 as grammatical, they should reject *VF sentences as this option was limited to subordinate clauses. In order to establish whether participants had learned about the relevance of clause sequence, their performance on complex constructions was observed. Here, the acquisition of relevant grammatical knowledge would be reflected in the endorsement of V2-VF and VF-V1 constructions and the rejection *VF-V2 and *V1-VF constructions. Finally, in order to determine whether learning was restricted to the clause-level, participants' performance on five types of simple sentences was compared, namely on *V1, V2, *V3, *V4, and *VF. If, as experiments 1 and 2 suggested, learning was restricted to the clause level, then experimental participants should display a tendency to endorse V2, *VF and *V1 but not *V3 or *V4. This performance would suggest that subjects knew where the VP could be placed within a clause, but that they were unaware of the relevance of other cues to verb placement (e.g., clause type).

Table 17

Sentence Templates for the Six Ungrammatical Patterns. Frequencies in Testing Set Are in Parentheses. Sample Sentences Are in Italics.

<table>
<tr><th>Pattern</th><th>Template</th><th></th></tr>
<tr><td>*V1</td><td>[[VP] > [NP]$_{subj}$ > [AP|NP|PP]$_{temp}$ > [AP|NP|PP] > [NP|PP]]</td><td>(5)</td></tr>
<tr><td></td><td>Invited Emma after dinner some colleagues to her birthday party.</td><td></td></tr>
<tr><td>*V3</td><td>[[AP|NP|PP]$_{temp}$ > [NP]$_{subj}$ > [VP] > [NP|PP] > [NP|PP]]</td><td>(5)</td></tr>
<tr><td></td><td>Some time ago John filled the bucket with apples.</td><td></td></tr>
<tr><td>*V4</td><td>[[AP|NP|PP]$_{temp}$ > [NP]$_{subj}$ > [NP|PP] > [VP] > [NP|PP]]</td><td>(5)</td></tr>
<tr><td></td><td>Recently Susan much furniture imported for her new weekend retreat.</td><td></td></tr>
<tr><td>*VF</td><td>[[AP|NP|PP]$_{temp}$ > [NP]$_{subj}$ > [NP|PP] > [NP|PP] > [VP]]</td><td>(5)</td></tr>
<tr><td></td><td>After dinner Susan an old car with her savings bought</td><td></td></tr>
<tr><td>*VF-V2</td><td>[[SUB > [NP]$_{subj}$ > [AP|NP|PP]$_{temp}$ > [NP|PP] > [VP]] > [NP]$_{subj}$ > [VP] > [NP|PP]</td><td>(5)</td></tr>
<tr><td></td><td>Because his children recently a calculator required Jim called the electronics store.</td><td></td></tr>
<tr><td>*V1-VF</td><td>[[VP] > [NP]$_{subj}$ > [NP|PP] > [SUB > [NP]$_{subj}$ > [AP|NP|PP]$_{temp}$ > [NP|PP] > [VP]]]</td><td>(5)</td></tr>
<tr><td></td><td>Stayed Jennifer at the hotel because her husband yesterday a boring conference attended.</td><td></td></tr>
</table>

As in the previous experiments, several restrictions applied to the lexicon of the testing set. With the exception of a limited number of function words, no words were repeated from the training set. No verb was repeated from the training set and only multisyllabic verbs were used in clause-final position. The NP$_{subj}$ was again realized as one of 10 names, five of which were male and five female. This time, however, the names *David, Jim, John, Paul, Peter, Chloe, Emma, Janet, Jennifer* and *Susan* were employed. Temporal expressions were instantiated by new APs (*recently* and *yesterday*), NP (*some time ago*) and PPs (*after dinner*, *in the afternoon*). In the case of the subordinate clauses, two new subordinators were employed in the testing set: *because* and *when* were used as subordinators in VF-V1, *V1-VF and *VF-V2 sentences.

A frequency analysis of the testing set indicated that the average sentence length was 11.1 words per sentence for grammatical items and 11.6 for ungrammatical items. There was no significant difference between training and testing sets with regard to sentence length, $F(1, 193) = .922$, $p > .05$, i.e. sentence length was not a reliable cue to grammaticality in the testing phase. In addition, the bigram statistics of the training set were used to compute ACS to assess the similarity of the testing set to the training set. Since very few lexical items occurred in both sets, the bigram statistics were calculated over phrase types (NP, VP, NP, etc.) and not actual words (*John*, *bought*, *apples*, etc.). The analysis indicated that grammatical sentences had an average ACS of 44.9 and ungrammatical sentences an average ACS of 38.7, which suggests that grammatical items resembled the training set more closely than ungrammatical items. Statistical analyses indicate a significant difference between grammatical and ungrammatical items of the testing set if ACS is taken as a measure of similarity, $t(29) = 2.516$, $p < .05$.

4.2.3. Procedure

The experiment consisted of (i) a training phase, (ii) a testing phase, and (iii) a debriefing session. Experimental participants were trained on the semi-artificial grammar by means of elicited imitations and plausibility judgments. Controls moved on directly to the testing phase, which required classifying novel instances of the semi-artificial grammar as well as reporting confidence levels and the basis for each classification decision. The experiment concluded with a debriefing session during which subjects completed a questionnaire. All tasks were run on a Dell PC (Windows XP) and delivered via Cedrus SuperLab Pro (version 2.0). The instructions are reproduced in the appendix.

4.2.3.1. Training phase

At the beginning of the training phase, participants were told they were taking part in an experiment on sentence comprehension during which they would be exposed to 120 “scrambled” sentences. The training task required participants to listen to the training set on an item-by-item basis, to repeat each sentence after a delayed prompt (1,500 ms), and to judge whether the statement made was semantically plausible or not. The figure below illustrates the training procedure.

Figure 18. Training procedure in experiment 3.

A 1,500 ms delay after stimulus presentation was incorporated in order to avoid elicited reconstructions based exclusively on verbatim memory. The elicited imitations were captured on a directional microphone and recorded on tape for subsequent analysis. Immediately after each repetition, participants were asked to decide on the semantic plausibility of the sentence. The two response keys (*/* for plausible, *z* for implausible constructions) were displayed on the screen below the prompt question ("Do you think this is plausible?"). When the plausibility of a given sentence was misjudged, the sentence was played again and participants were required to repeat the plausibility judgment. The presentation order of training items was determined randomly for each participant. There was a 500 ms delay between the end of the plausibility judgment and the presentation of the following sentence. The screen was blank while the audio files were played. The experiment began with a short practice session to familiarize participants with the training task. This consisted of four practice sentences which were not repeated in the actual training phase. No mention was made that the scrambling followed the word order rules of a natural language. The entire training phase took, on average, 40 minutes to complete.

4.2.3.2. Testing phase

After training, experimental participants were informed that the word order in the previous sentences was not arbitrary but that it followed a "complex system" instead. They were then instructed to listen to 60 new scrambled sentences, only half of which would follow the same rule-system as the sequences they had just been exposed to. Those sentences that did obey the rules should be endorsed as "grammatical" and those that did not rejected as "ungrammatical." For each test sentence, participants were required to decide on its grammaticality ("Was the sentence grammatical?"), to report how confident they were in their judgment ("How confident are you in your decision?"), and to indicate what the basis of their judgment was ("What was the basis of your grammaticality judgment?"). In the case of the control group, participants were merely told that they would listen to 60 scrambled sentences and asked to judge whether or not a sentence was grammatical. There were no time intervals between the three prompts, i.e. subjects would receive the next prompt immediately after making a response. There was, however, a 500 ms delay between the source judgment and the presentation of a new test sentence. Participants were given no feedback regarding the accuracy of their grammaticality decisions. Figure 19 illustrates the testing procedure.

Figure 19. Testing procedure in experiment 3.

Participants could indicate their levels of confidence by selecting one of three response options: guess, somewhat confident or very confident. The „guess' category indicated that the participant had no confidence in her classification decision and believed to be guessing, while the „somewhat confident' category and the „very confident' category indicated that the participant was either somewhat or very confident in her decision. In the case of the source attributions, there were four response options which participants could use to report the basis of their grammaticality judgments: guess, intuition, memory and rule. The „guess' category indicated that the participant believed the classification decision to be based on a guess. The "intuition" category indicated that the participant was somewhat confident in her decision but did not know why it was right. The "memory" category indicated that the judgment was based on the recollection of parts or entire sentences from the training phase. Finally, the "rule" category indicated that the participant based the decision on a rule that was acquired during the training phase and would be able to report the rule at the end of the experiment. All participants were provided with these definitions before starting the testing phase. The keyboard options (*/* for grammatical, *z* for ungrammatical; *0* for guess, *5* for somewhat confident, *9* for very confident; *g* for guess, *i* for intuition, *m* for memory, and *r* for rule) were displayed on the screen below the relevant prompts. The test sentences were presented to each participant in random order. The testing phase began with a short practice session to familiarize the participants with the new task. This consisted of four practice trials which were not repeated in the actual testing set. The entire testing phase took approximately 15 minutes to complete.

4.2.4. Debriefing

At the end of the experiment, participants were given the same debriefing questionnaire as participants in the previous two experiments (see appendix). The questionnaire required subjects to estimate their accuracy in the classification task and prompted them verbalize any rule or regularity they might have noticed during the course of the experiment. Finally, the questionnaire also asked participants to supply their name, age, gender, nationality, occupation and linguistic background.

4.3. Results

4.3.1. Training phase

Performance during the training phase was assessed by analyzing the elicited imitations and the plausibility judgments. In the case of the elicited imitations, an obligatory occasion analysis was conducted on all repetitions produced by the experimental subjects. Participants received a point for every verb that was placed in the correct location, i.e. the correct repetition of a one-clause sentence added one point to participants' score and the correct repetition of a two-clause sentence two points. The proportion of correct responses over the training sentences was computed for each participant. In the case of the plausibility judgments, the proportion of correct judgments was calculated for each subject.

4.3.1.1. Elicited imitations

Due to a technical failure, the audio files of two participants could not be used for this analysis. After scoring the performance of 13 subjects, the analysis showed that the mean accuracy score on this task was 96% (*SD* = 7%). One subject performed two standard deviations below the group mean, namely at 74% accuracy. Given the importance of the training phase in any learning experiment, it was decided to remove this subject from all subsequent analyses. The mean accuracy score for the remaining 12 experimental subjects was 97.7% (*SD* = 2.3%), i.e. almost 98% of all repetitions were correct with regards to the placement of VPs.

A closer analysis of the errors showed that only 33 out of 1,443 repetitions were erroneous. Of the wrongly repeated sentences, 48.5% were plausible and 51.5% implausible. Analyzing the errors by sentence type showed that the majority of mistakes (97%) were committed on complex sentences. This is likely to be explained by working memory constraints since complex sentences tended to be longer and thus placed higher demands on retention. Errors on VF-V1 constructions accounted for 75.8% of errors (*n* = 25, of which 13 plausible, 12 implausible), errors on V2-VF constructions for 21.2% (*n* = 7, of which 3 plausible, 4 implausible) and errors on V2 constructions for only 3% (*n* = 1 implausible).

Breaking down participants' performance into quartiles showed that the mean accuracy for the first quartile, i.e. repetitions 1-30, was 97%, for the second quartile (reps 31-60) 96%, for the third quartile (reps 61-90) 99%, and for the fourth quartile (reps 91-120) 98%. A repeated measures ANOVA showed no significant differences in mean accuracy across the four quartiles, $F(3, 33) = 1.661$, $p > .05$, i.e. experimental subjects performed equally well across the different stages of the elicited imitation task.

4.3.1.2. Plausibility judgments

In terms of frequency, participants committed, on average, 14.4 errors in their plausibility judgments. Because misjudged sentences were repeated, this suggests that the average experimental participant was exposed to 134.4 sentences during training. The analysis also showed that the experimental group judged 89.5% (*SD* = 3.6%) of sentences correctly. Experimental participants were significantly more accurate when judging plausible sentences (*M* = 94.8%, *SD* = 4.1%) rather than implausible ones (*M* = 84.9%, *SD* = 5.6%), $t(13) =$ 5,514, $p < 0.001$, which suggests that implausibility affected accuracy in the plausibility judgment task negatively.

4.3.2. Testing phase

4.3.2.1. Grammaticality judgments

Overall performance

The analysis of the grammaticality judgments showed that the experimental group classified 61.6% (*SD* = 8.3%) of the test items correctly and the control group 42.9% (*SD* = 5.1%). The difference between the two groups is significant, $t(27) = 7.289$, $p < .001$. Further analysis showed that the experimental group performed significantly above chance on this task, $t(13) = 5.150$, $p < .001$, while the controls scored significantly below chance, $t(14) = -5.361$, $p < .001$. The training phase produced a clear learning effect in the experimental participants.

In order to establish whether any improvements in accuracy occurred during the testing phase, the grammaticality judgment data was equally divided into 10 chronological stages. The mean accuracy for each block of six judgments was then computed for both groups. A repeated measures ANOVA showed no significant differences across the stages in the experimental group, $F(9, 117) = .760$, $p > .05$, or in the control group, $F(5.1, 71.9) =.722$, $p > .05$, with Greenhouse-Geisser correction. That is, no online learning effect was observed during the testing phase. Figure 20 displays the mean accuracy across the 10 stages for both groups.

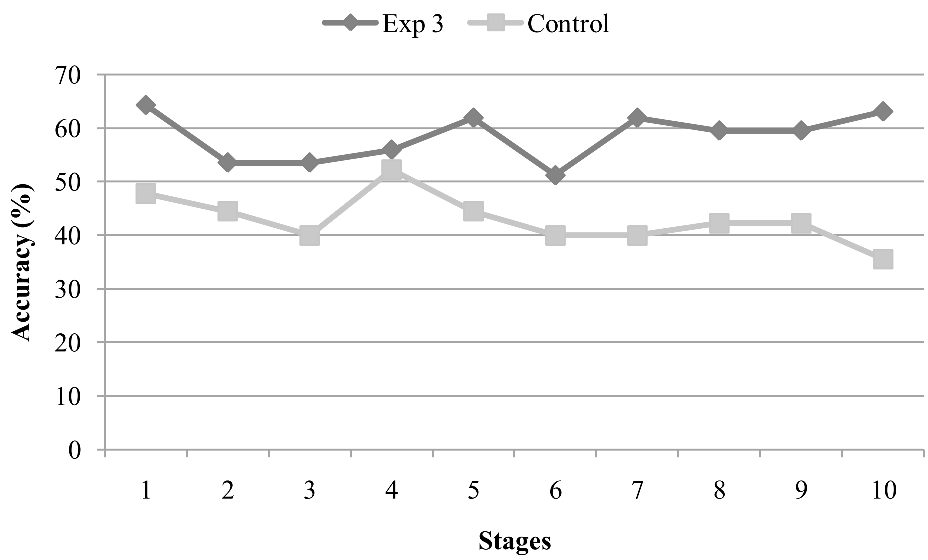

Figure 20. Classification accuracy of experimental and control subjects across 10 chronological stages of the testing set.

Classification performance on grammatical and ungrammatical items

The following analyses report endorsement rates, rather than accuracy scores. The results showed that the experimental group endorsed 71% (*SD* = 13.5%) of all grammatical sentences and 47% (*SD* = 19.7%) of all ungrammatical ones. No experimental subject endorsed more than 90% of grammatical sequences and less than 10% of ungrammatical ones, i.e. there was no evidence of categorical classification performance. The control group only endorsed 36.4% (*SD* = 30.3%) of grammatical sentences and 51% (*SD* = 28.4%) of ungrammatical ones, $t(14) = -5.268$, $p < .0005$. The difference between experimental and control subjects on grammatical items was significant, $F(1, 27) = 14.824$, $p < .001$, i.e. the experimental group was significantly more likely to correctly endorse grammatical strings. The difference between groups on ungrammatical strings was not significant, however, $F(1, 27) = .125$, $p > .05$. As in experiment 1, learning appeared to be largely restricted to grammatical items.

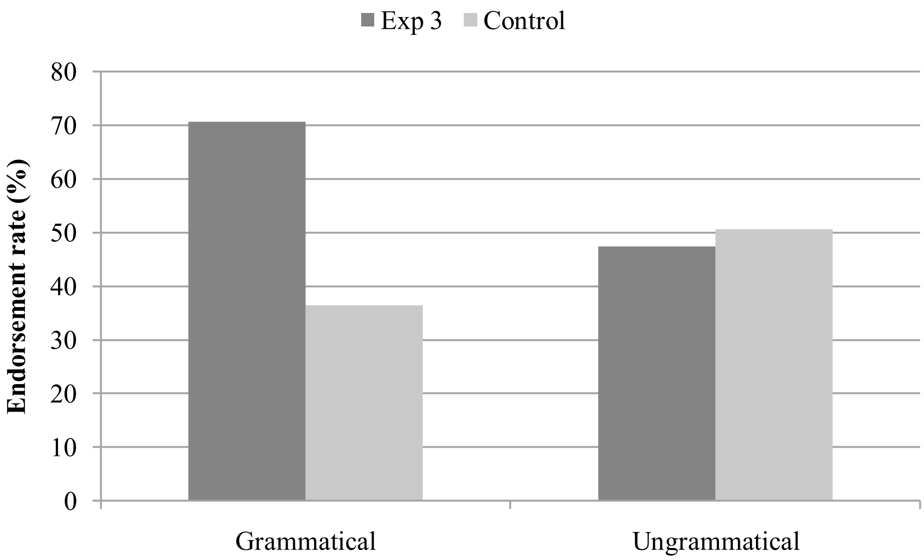

Figure 21. Classification performance of experimental and control subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

Classification performance across sentence types

Figure 22 illustrates participants' endorsement rates across the nine syntactic templates employed in this experiment. The scores are summarized in table 18. The analysis indicated that the experimental group was significantly more likely than the control group to correctly endorse all grammatical sentence types, i.e. V2 sentences, $F(1, 27) = 6.426$, $p < .05$, V2-VF sentences, $F(1, 27) = 15.778$, $p < .001$, and VF-V1 sentences, $F(1, 27) = 8.831$, $p < .05$. The experimental group thus displayed a clear preference for endorsing previously encountered sentence patterns. The experimental group was also significantly more likely to correctly reject *V3 sentences, $F(1, 27) = 62.401$, $p < .001$. The fact that experimental subjects only endorsed 22.9% of *V3 sentences, compared to the 97.3% endorsement rate in the control group, is interesting because the latter are the most English-like structures in the experiment. It suggests that the experimental group had learned to reject a sentence type that would otherwise be acceptable in English. Experimental subjects were also significantly more likely to erroneously endorse stand-alone *VF sentences, $F(1, 27) = 5.18$, $p < .05$, which suggests that they had not learned that clause-final VPs were restricted to subordinate clauses. In the case of *V1-VF sentences, the difference between experimental and control subjects approaches significance, $F(1, 27) = 4.083$, $p = .053$. There were no significant differences between the groups on *V1, *V4, and *VF-V2.

The performance on one-clause constructions indicated that experimental group 3 preferred V2 sentences over *V1 sentences, $t(13) = -3.702$, $p < .05$, *V3 sentences, $t(13) = 4.979$, $p < .0005$, and *V4 sentences, $t(14) = 3.342$, $p < .05$. However, experimental group 3 was just as likely to endorse V2 as *VF sentences, $t(13) = -.450$, $p > .05$, which suggests that subjects did not know that verb-final is limited to subordinate clauses. The performance on complex constructions showed that no significant differences between V2-VF sentences and *V1-VF sentences, $t(13) = 1.351$, $p > .05$, but experimental group 3 was significantly more likely to endorse VF-V1 sentences than *VF-V2 sentences, $t(13) = 2.874$, $p < .05$.

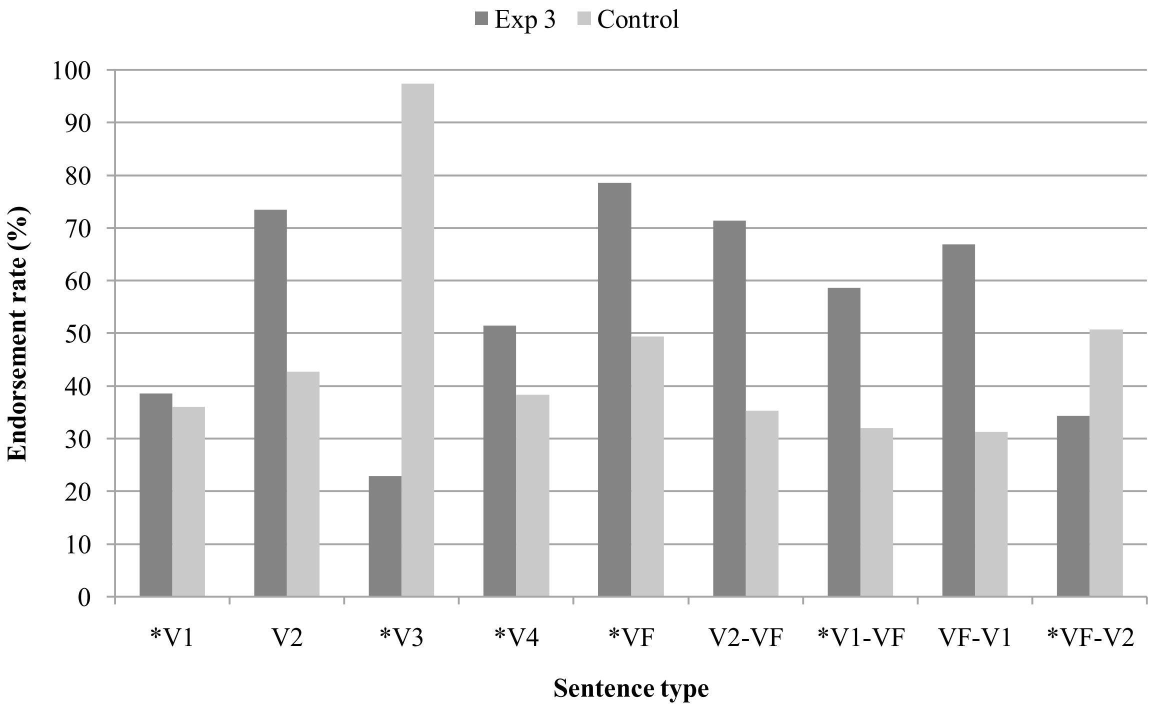

Figure 22. Classification performance of experimental and control subjects across sentence types. Bars represent endorsement rates.

Table 18
Classification Performance (%) of Experimental and Control Subjects across Sentence Types

		Sentence types								
Group		*V1	V2	*V3	*V4	*VF	V2-VF	*V1-VF	VF-V1	*VF-V2
Exp 3	*M*	38.6	73.4*	22.9*	51.4	78.6**	71.4**	58.6	66.9*	34.3
	SD	36.3	27.6	35.8	33	22.8	18.8	35.5	29.2	30.8
	SE	9.7	7.4	9.6	8.8	6.1	5	9.5	7.8	8.2
Controls	*M*	36	42.7	97.3**	38.3	49.3	35.3^{+}	32^{+}	31.3	50.7
	SD	34	36.7	7	37.0	42.7	28.8	35.3	34.8	38.4
	SE	8.8	9.5	1.8	9.6	11.0	7.4	9.1	9	9.9

Significance from chance: $^{+} = p < .07$. * $p < .05$. ** $p < .001$.

4.3.2.2. Confidence ratings

Overall performance

The following analyses only report the results of the experimental group. The average confidence level in the experimental group was 5.9 (*SD* = 1.8). In terms of proportion, experimental participants tended to select the option somewhat confident most frequently and the guess option least frequently. In terms of accuracy, the analysis indicated that experimental participants were most accurate when reporting to be very confident in their decisions and slightly less accurate when reporting to be somewhat confident. Accuracy was lowest for those grammaticality decisions in which subjects had no confidence whatsoever. Experimental participants scored significantly above chance when reporting to be somewhat confident and very confident. When participants reported to be guessing, performance was indistinguishable from chance. Table 19 summarizes the findings.

Table 19
Accuracy and Proportions (%) across Confidence Ratings

	Guess	Somewhat confident	Very confident
Accuracy	53	60*	65*
Proportion	12	54	34

Significance from chance: * $p < .05$.

In order to determine whether participants became more confident while performing on the testing phase, the confidence ratings were equally divided into 10 chronological stages, and the average confidence levels were calculated for each stage. A repeated measures ANOVA showed no significant differences in mean confidence across the stages, $F(3.44, 44.68) = .810$, $p > .05$, with Greenhouse-Geisser correction. That is, the experimental group did not display significant changes in confidence levels while performing on the testing phase. In addition, the proportions of guess, somewhat confident and very confident responses were calculated for each of the 10 stages. As figure 23 suggests, there were relatively few changes in the proportion of somewhat confident responses across the testing phase. Interestingly, however, the proportion of guesses was higher towards the end of the testing phase, while the proportion of very confident responses was higher at the beginning. The analysis confirms that participants did not become more confident while performing on the grammaticality judgment task.

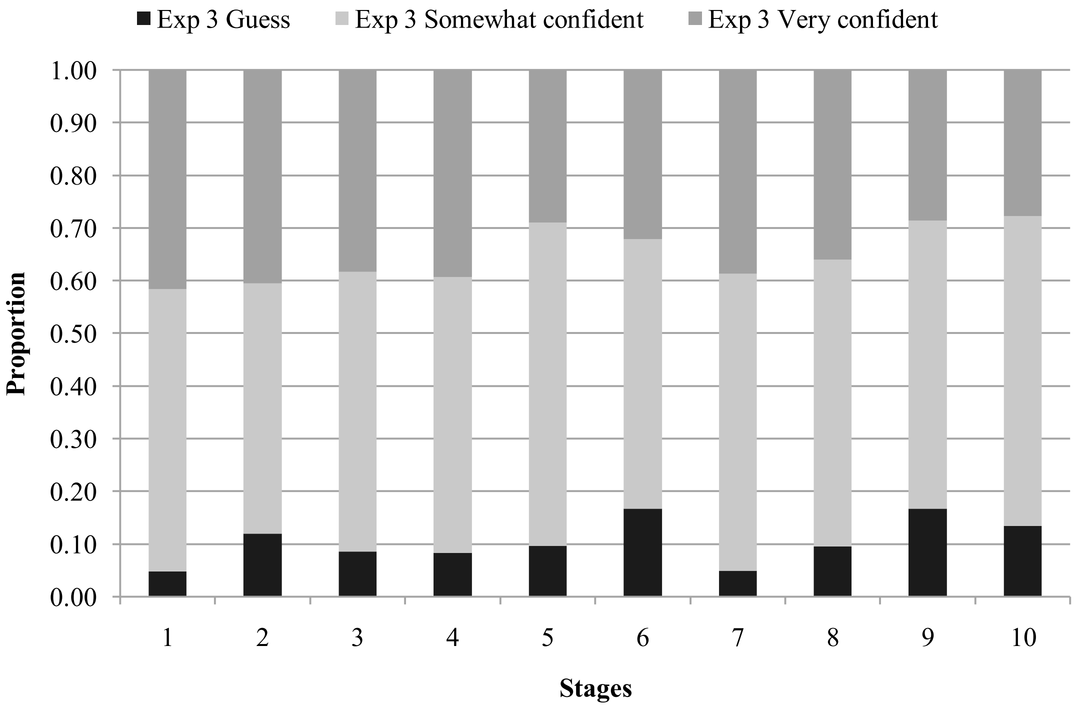

Figure 23. Proportions of guess, somewhat confident and very confident responses across 10 stages of the testing set.

Zero correlation criterion

The Chan difference score was computed in order to establish whether learning in the experimental group was implicit by the zero correlation criterion. The average confidence for correct grammaticality decisions was 6.1 (*SD* = 1.6) and the average confidence for incorrect decisions was 5.6 (*SD* = 1.7), i.e. experimental participants were more confident in correct decisions than in incorrect ones. The difference (0.5) was significant, $t(13) = 2{,}310$, $p < .05$, which indicates that participants were partially aware of the knowledge they acquired.

4.3.2.3. Source attributions

Overall performance

In terms of proportion, experimental participants most frequently believed their classification decisions to be based on rule knowledge, followed by intuition and memory. The guess category was selected least frequently. In terms of accuracy, experimental participants scored highest when reporting to use rule knowledge to guide their decisions, followed by the intuition and memory categories. Subjects were least accurate when basing decisions on guesses. Participants only performed significantly above chance when basing their decisions on intuition or on rule knowledge.

Table 20

Accuracy and Proportions (%) across Source Attributions

	Guess	Intuition	Memory	Rule
Accuracy	56	59*	57	65**
Proportion	10	32	15	43

Significance from chance: * $p < .05$. ** $p < .001$.

Relationship between confidence ratings and source attributions

Mean confidence ratings were calculated for each source attribution in order to assess the relationship between confidence ratings and source attributions. Participants reported the highest confidence levels for decisions based on rule knowledge (M = 7.67), followed by decisions based on memory (M = 5.65), intuition (M = 4.76) and guesses (M = 0.99). Participants also gave significantly higher confidence levels for attributions associated with conscious structural knowledge (rules and memory, combined M = 6.66) rather than unconscious structural knowledge (guess and intuition, combined M = 2.88), $p < .05$.

Table 21 summarizes the mean accuracy score for pooled confidence ratings and source attributions. In terms of proportion, the analysis indicates that the majority of grammaticality judgments were either somewhat confident and based on intuition or very confident and based on rule knowledge. In terms of accuracy, participants performed significantly above chance when making somewhat confident decisions based on intuition, which further suggests that participants acquired unconscious knowledge. Participants also performed significantly above chance on somewhat confident and very confident decisions based on rule knowledge. In the remaining cases, performance was not significantly different from chance.

Table 21
Accuracy and Proportions (%) for Combined Confidence Ratings and Source Attributions

		Source attributions			
Confidence ratings		Guess	Intuition	Memory	Rule
Guess	Accuracy	55.1	46.9		
	Proportion	7.9	3.7	0	0
Somewhat confident	Accuracy	61.5	60.5*	55.8	62.6*
	Proportion	1.5	25.6	12.9	14.1
Very confident	Accuracy	50	59.1	68.2	66.1**
	Proportion	0.2	2.5	2.5	29.1

Significance from chance: * $p < .05$. ** $p < .001$.

4.3.3. Debriefing

4.3.3.1. Accuracy estimates

Participants' assessment of their performance on the grammaticality judgment task (*estimated accuracy*) was compared to their actual performance (*observed accuracy*). The comparison showed that the experimental group was impressively accurate in estimating their performance. The difference between estimated accuracy (M = 61.57%, SD = 7.7%) and observed accuracy (M = 61.56%, SD = 8.3%) was 0.01. This difference was not significant, $t(13) = .004$, $p > .05$, which suggests that subjects were aware of the fact that they had acquired knowledge. The control participants, on the other hand, significantly overestimated their performance on the grammaticality judgment task by 17.1%, $t(14) = 3.049$, $p < .05$. Control subjects believed to be accurate in 59.9% (SD = 19.5%) of classification decisions but were, in fact, only accurate in 42.8% (SD = 5.8%). The analysis thus supports the view that the experimental subjects, but not the controls, were at least partially aware of the knowledge that guided their grammaticality decisions.

4.3.3.2. Verbal reports

The transcription of all verbal reports can be found in the appendix. The analysis showed that subjects in both groups reported knowledge when prompted, at the end of the experiment, to describe any rules or regularities they might have noticed. In the experimental group, all subjects verbalized knowledge, whereas in the control condition only 11 did. The analysis showed again that participants did not rely on phrasal categories such as NP, PP or VP in their reports, employing functional categories (such as subject, object, etc.) and semantic categories (such as time or action) instead. It also showed that no subject was able to provide rule descriptions such as the ones outlined in table 1 above.

As far as the experimental group is concerned, the most frequently remarked regularity concerned the “reversed” order of subject and verb (see e.g. reports of subjects 2, 8, 9, 13, 14, and 15). One subject (13) was able to provide a correct example of a V2 sentence and of a VF-V1 sentence. Several subjects reported that sentences could end with a VP in final position (8, 11, 12, 14) but did not associate this position with subordinate clauses. One of the experimental subjects (8) reported a bias against sentences that begin with a verb. A few subjects reported wrong rules, endorsing for example *V4 constructions (1) or making their grammaticality decision dependent on the presence of a time expression (4, 5). Finally, several other verbalizations were largely irrelevant to the grammar used in this task, with subjects correctly reporting, for example, that the inclusion of astronomical references or fruits made sentences implausible (7, 11).

Like the experimental subjects, controls most frequently commented on the “reversed” order of subjects and verb (e.g. control subjects 1, 2, 3, 4, 6, 10, and 11). One of the controls (5) reported a bias against “verb-subject constructions”, i.e. for sentences that began with a VP. Other verbalizations were largely irrelevant to the grammar used in this task, with subjects reporting, for example, the presence of many different expressions of time in the testing set (11) or erroneously assuming an absence of definite articles (10).

In sum, the analysis of the verbal reports showed that, while most participants in this experiment were able to verbalize knowledge, there were very few verbalizations that were relevant to the rules of the semi-artificial grammar. No subject was able to provide descriptions of the verb placement rules employed to generate the stimulus material, and only one subject provided two correct examples (of V2 and VF-V1 sentences). The analysis of the reports suggests that participants focused exclusively on the ordering of words within clauses, disregarding important cues such as clause type and clause sequence. This might be taken as an indication that participants did not pay attention to phrasal arrangements above the clause-level. Despite the fact that subjects were conscious of some knowledge, the analysis of the verbal reports suggests that subjects were largely unaware of the rules that determine the placement of VPs.

4.4. Discussion

Experiment 3 introduced several modifications to the design of the previous experiments. Three, rather than four, verb placement rules determined the placement of VP, which meant that the position of the verb was only dependent on clause type and on clause sequence. Elicited imitations were introduced as an additional training task, and source attributions were added to the testing phase as an additional measure of awareness. The number of training stimuli was reduced from 240 to 120 sentences, i.e. subjects were exposed to the training set only once. As mentioned above, experiment 3 was designed in order to establish whether the addition of a training task that required subjects to process word order more directly had a positive effect on learning and to evaluate the usefulness of source attributions for measuring awareness.

The results showed that the modifications produced a greater learning effect in the experimental group (61.6%, compared to 54.6% in experiment 1 and 51.9% in experiment 2). At the same time, however, learning was again somewhat limited. As in experiment 1, performance was largely driven by the above-chance endorsement of grammatical sentences. Experimental participants endorsed 71% of all grammatical items but also 47% of ungrammatical items, which indicates that classification performance might have been partially based on memory for previously encountered patterns. The fact that classification performance was not categorical, but probabilistic, in nature also suggests that subjects did not acquire the verb placement rules. When participants based their classifications on rule knowledge, these rules did not correspond to the word order rules of the semi-artificial grammar. The analysis of the verbal reports supports this view. Furthermore, the performance across sentence types suggests that subjects did not learn that clause type and clause sequence determined the placement of VPs. *VF sentences, for example, were overendorsed, which means that subjects did not know that a verb could only occur at the end of a specific clause type. Subjects were more likely to endorse VF-V1 sentences than their ungrammatical counterparts, *VF-V2, but the preference for VF-V1 could be explained by the acquisition of a micro-rule like “a verb can follow a verb in the middle of a sentence”. There would be no need to assume that subjects had acquired the V1 rule. The fact that subjects were just as likely to endorse V2-VF sentences as *V1-VF sentences supports the notion that subjects did not learn that clause sequence was a relevant cue to verb placement.

Since several changes were made to the previous design, it is difficult to judge what particular modification has improved performance in this experiment. One possibility is that the addition of the elicited imitations improved overall performance. In contrast to experiments 1 and 2, subjects in the present experiment had to focus both on the arrangement of words and on sentence meaning. The fact that subjects had to recall word order in order to perform on the training task could have increased learning by making them process serial order more directly. A different possibility is that the removal of the Split VP rule simplified the grammar and made the structure of the stimulus sentences easier to acquire. While subjects did not learn to associate verb positions with specific types of clauses or clause arrangements, this simplification could have led them to learn what positions were licensed by the grammar. This seems unlikely, however, given that trained subjects in experiment 1 appear to have learned similar things. Both groups knew, for example, that a verb could come in second or in final position but not that clause type or clause sequence played a role in verb placement. A third possibility is that the ungrammatical templates used in experiment 3 were easier to reject than those employed in the previous experiments. Experimental subjects in experiments 1 and 3 were equally likely to endorse grammatical sentences, but subjects in the former experiment were more likely to wrongly endorse ungrammatical ones. Perhaps subjects in experiment 3 would have performed equally poorly had they been asked to judge *V3VF, *V2-V1 and *VF-VF sentences instead of *V1, *V3, *V4, *VF, *V1-VF, and *VF-V2.

Why did subjects not learn that clause type and clause sequence determined the position of the verb? One explanation is that subjects were not paying attention to the appropriate cues in the input. Nonlinguists are usually unaware of notions such as clause type or clause sequence and are therefore less likely to notice them. In addition, performance could also have been affected by subjects' pre-existing knowledge. As argued in chapter 1, metalinguistic knowledge (e.g., knowledge of functional categories such as subject, verb, and object) could have distracted attention from linguistic notions which are not as readily available. Experiment 4 investigates whether metalinguistic knowledge prevented subjects from acquiring the verb placement rules by replacing the English lexicon, employed in the previous experiments, with an artificial lexicon consisting of nonsense syllables. If pre-existing knowledge affected learning negatively, then exposure to a rule-system to which this knowledge does not apply should result in increased learning. A different explanation is that the training task only required subjects to recall the order of words but did not encourage them to pay attention to the type of clause they were processing. Assuming that attention to relevant aspects of the grammar is necessary for learning to take place, the results could be explained by the fact that the training tasks did not direct subjects' attention to elements such as clause type and clause sequence. Experiment 5 introduces a new training task that requires subjects to process clause types more directly in order to determine whether this increases learning.

In terms of awareness, experiment 3 provided evidence of implicit learning. The analysis of confidence ratings and source attributions showed that, while subjects were aware of the fact that they had acquired knowledge, they were at least partially unaware of what knowledge they had acquired. When attributing grammaticality judgments to intuition, subjects performed significantly above chance, i.e. they had acquired unconscious structural knowledge. At the same time, it is important to highlight that subjects in experiment 3 were significantly more accurate when reporting higher levels of confidence and when basing their decisions on explicit categories (memory and rule). Conscious (but unverbalizable!) knowledge was clearly linked to improved performance in the grammaticality judgment task. As far as methodology is concerned, experiment 3 further confirms that the sole reliance on verbal reports is clearly inadequate for assessing awareness. The analysis of the verbal reports showed that participants were unable to verbally describe the rules of the semi-artificial system, which would have supported the erroneous assumption that learning in experiment 3 occurred entirely without awareness. The fact that subjects developed conscious judgment knowledge would have gone unnoticed, and it would have been difficult to explain why estimated and observed accuracy were exceptionally well-calibrated. The combined use of confidence ratings and source attributions appears to be a promising method for assessing awareness. Experiments 4, 5 and 6 therefore employ the same awareness measures as experiment 3.

5. Experiment 4

5.1. Introduction

The present experiment investigates the acquisition of syntactic rules of a semi-artificial language under incidental learning conditions. The rules were still based on German syntax, but the English lexicon of previous experiments was replaced by a lexicon that consisted of nonsense syllables. Participants were trained on the new language by means of elicited imitations. The testing phase consisted of the same components as experiment 3, i.e. grammaticality judgments, confidence ratings and source attributions. At the end of the experiment, participants were required to estimate their accuracy and to report any rules they might have noticed. The aim of this experiment was to determine whether linguistic knowledge prevented subjects from acquiring the verb placement rules.

5.2. Method

5.2.1. Participants

Thirty native speakers of English (17 women and 13 men, mean age = 23.7 years) were recruited to take part in this experiment. All participants were students at the University of Cambridge at the time of the experiment. In terms of linguistic background, all participants were native speakers of English with no background in German or any other V2 language. Participants spoke, on average, 2.2 foreign languages. Experimental and control groups did not differ significantly across age, $t(26) = .946$, $p > .05$, gender, $\chi^2(1, N = 30) = 3.589$, $p > .05$, occupation (student vs. nonstudent), $\chi^2(1, N = 30) = 2.143$, $p > .05$, and number of languages acquired, $\chi^2(4, N = 30) = 3.470$, $p > .05$. All subjects received £5 for participating in the experiment.

5.2.2. Stimulus material

Table 22 summarizes the lexicon used in this experiment. A semi-artificial grammar, consisting of nonsense syllables and German syntax, was used to generate the stimulus material for this experiment. The lexicon consisted of six categories, each of which was uniquely associated with distinct surface instantiations. The subject category, for example, was realized either as “sa” or “se”, the verb category as “ko” or “ki”, the object category as “pu” or pa”, and so forth. In creating the stimuli, V2, V2-VF and VF-V1 sentences from experiment 3 were broken down into their functional categories and then translated into syllable sequences by following the lexicon outlined in the table below. Functional categories were employed for the conversion process because the verbal reports of the previous experiments indicated that subjects were more likely to perceive sentences in terms of functional, rather than syntactic, categories.[25] A division was made between temporal and nontemporal adjuncts because subjects in experiments 1-3 distinguished between these components. For example, sentences that consisted of temporal adjunct, predicate, subject, complement, and nontemporal adjunct (e.g. “Yesterday bought John the newspaper in the supermarket.”) were converted into sequences such as “Fa Ko Se Pu Te.” Sentences such as “Last week purchased Susan many apples because her daughter fruit loved.” (temporal adjunct, predicate, subject, complement, subordinator, subject, complement, predicate) corresponded to “Fa Ki Sa Pu Ri Se Pu Ki”, and so forth. The different surface items were balanced for frequency. The linguistic focus in this experiment was on the placement of K-syllables, the analogue of VPs, within a given syllable sequence.

Table 22
Lexicon Employed in Experiment 4

Category	Syllable type	Stimulus syllables
Subject	S	Sa
		Se
Predicate	K	Ko
		Ki
Complement	P	Pu
		Pa
Adjunct (nontemporal)	T	Te
		Tu
Adjunct (temporal)	F	Fa
		Fi
Subordinator	R	Ra (= *that*)
		Ru (= *since*)
		Ri (= *because*)

A total of 105 syllable sequences were drafted for this experiment. The sequences were read out by a male native speaker of British English who was instructed to maintain the intonation patterns of the original sentences. In other words, a sequence like “Ru Sa Pa Fi Ko, Ki Se Pa Te” was supposed to have a similar intonation as a sentence like “Since his parents houses in

[25] A complement is an obligatory syntactic component, an adjunct is a nonobligatory component.

Scotland sold, purchased Jim an expensive watch for his wife." The idea was to remove the semantic layer from the stimulus structures but to keep the remaining information as similar as possible to the stimuli in experiment 3. The syllable sequences were digitally recorded on a Sony Mini-Disc player (MZ-R700) and subsequently edited with sound processing software (Audacity, version 1.2.4).

5.2.2.1. Training set

The training set, which is reproduced in the appendix, consisted of 45 syllable sequences. Three templates (V2, V2-VF and VF-V1) were used in the experiment. Fifteen sequences were generated for each template.

Table 23

Syntactic Templates for the Three Rules. Frequencies in the Training Set Are in Parentheses. Sample Sequences Are in Italics.

Rule	Templates	
V2	[Subj] > [Pred] > [Adj$_{temp}$] > [Comp|Adj] > [Comp|Adj]	(8)
	Se Ko Fi Tu Pu	
	[Adj$_{temp}$] > [Pred] >[Subj] > [Comp|Adj] > [Comp|Adj]	(7)
	Fa Ki Se Pu Tu	
V1	[[Sub] > [Subj] > [Comp|Adj] > [Pred]] > [Pred] > [Subj] > [Comp|Adj] > [Comp|Adj]]	(15)
	Ru Se Pa Ki Ki Sa Pa Te	
VF	[[Subj] > [Pred] > [Adj$_{temp}$] > [[Sub] > [Subj] > [Comp|Adj] > [Comp|Adj] > [Pred]]]	(8)
	Sa Ko Fa Ra Se Pa Te Ki	
	[[Adj$_{temp}$] > [Pred] > [Subj] > [[Sub] > [Subj] > [Comp|Adj] > [Comp|Adj] > [Pred]]]	(7)
	Fi Ko Sa Ra Sa Pu Te Ki	

In the case of the V2 rule, all sequences consisted of five syllables. The predicate (k-syllables) was placed in second position, while the first and the third positions of the sequence featured a subject (s-syllables) and a temporal adjunct (f-syllables). Eight sequences featured the subject in first position and seven a temporal adjunct. The fourth and fifth phrasal positions could be occupied either by a complement (p-syllables) or a nontemporal adjunct (t-syllables). The syntactic pattern for this rule, based on the position of the predicate, was "V2."

The sequences based on the V1 rule consisted of eight syllables. The sequence always began with a subordinator (r-syllables) in first position, followed by a subject in second and a complement or a nontemporal adjunct in third. The fourth and fifth positions featured two predicates. The sixth position was occupied by a subject, while the seventh and eight positions could either feature a complement or a nontemporal adjunct. The syntactic pattern for this rule was “VF-V1.”

Sequences based on the VF rule also consisted of eight syllables. A predicate had to occur in second and in final position. The first and the third position featured a temporal adjunct or a subject. If a subject occurred in the first position, the temporal adjunct had to occur in third position, and vice versa. Eight sequences featured the subject in first position and seven a temporal adjunct. The fourth position was always occupied by the subordinator “ra”, the equivalent to *that*, and the fifth position by the subject. The sixth and the seventh position could either be occupied by a complement or a nontemporal adjunct. The syntactic pattern, based on the positioning of the predicate in the sequence, was “V2-VF.”

5.2.2.2. Testing set

The testing set, which is reproduced in the appendix, consisted of 60 syllable sequences and was subdivided into 30 grammatical and 30 ungrammatical items. The same lexicon was used for generating the training and the test stimuli, but no actual training sequences were repeated in the testing phase. Subjects thus judged different sequences in the second part of the experiment. Nine patterns (three grammatical, six ungrammatical) were used to test whether subjects had learned where predicates could be placed. Table 24 outlines the templates used to generate grammatical and ungrammatical stimuli for the testing phase.

Table 24

Syllable Templates for Grammatical and Ungrammatical Patterns. Frequencies in the Testing Set Are in Parentheses. Sample Sequences Are in Italics.

Pattern	Grammatical	
V2	[Subj] > [Pred] > [Adj$_{temp}$] > [Comp|Adj] > [Comp|Adj]	(5)
	Sa Ko Fu Te Pa	
	[Adj$_{temp}$] > [Pred] >[Subj] > [Comp|Adj] > [Comp|Adj]	(5)
	Fi Ko Sa Pa Te	
V2-VF	[[Subj] > [Pred] > [Adj$_{temp}$] > [[Sub] > [Subj] > [Comp|Adj] > [Comp|Adj] > [Pred]]]	(5)
	Se Ki Fa Ra Se Pu Pa Ko	
	[[Adj$_{temp}$] > [Pred] > [Subj] > [[Sub] > [Subj] > [Comp|Adj] > [Comp|Adj] > [Pred]]]	(5)
	Fi Ki Sa Ra Se Tu Pa Ko	
VF-V1	[[Sub] > [Subj] > [Comp|Adj] > [Pred]] > [Pred] > [Subj] > [Comp|Adj] > [Comp|Adj]]	(10)
	Ri Sa Te Ki Ko Se Pu Pa	
	Ungrammatical	
*V1	[Pred] > [Subj] > [Adj$_{temp}$] > [Comp|Adj] > [Comp|Adj]	(5)
	Ko Se Fi Te Pa	
*V3	[Subj] > [Adj$_{temp}$] > [Pred] > [Comp|Adj] > [Comp|Adj]	(5)
	Sa Fa Ki Pa Te	
*V4	[Subj] > [Adj$_{temp}$] > [Comp|Adj] > [Pred] > [Comp|Adj]	(5)
	Se Fa Pa Ko Te	
*VF	[Subj] > [Adj$_{temp}$] > [Comp|Adj] > [Comp|Adj] > [Pred]	(5)
	Se Fa Pu Tu Ko	
*V1-VF	[[Pred] > [Subj] > [Adj$_{temp}$] > [[Sub] > [Subj] > [Comp|Adj] > [Comp|Adj] > [Pred]]]	(5)
	Ki Sa Fi Ra Se Pa Te Ki	
*VF-V2	[[Sub] > [Subj] > [Comp|Adj] > [Pred]] > [Subj] > [Pred] > [Comp|Adj] > [Comp|Adj]]	(5)
	Ri Sa Pu Ko Se Ko Pa Te	

A frequency analysis of the testing set indicated that the average sequence length was 7 syllables per sequence for grammatical items and 6 for ungrammatical items. There was no significant difference between training and testing sets with regard to sequence length, $t(2) = 2.000$, $p > .05$, i.e. sequence length was not a reliable cue to grammaticality in the testing phase.

5.2.3. Procedure

The experiment consisted of (i) a training phase, (ii) a testing phase, and (iii) a debriefing session. Experimental participants were trained on the semi-artificial system by means of elicited imitations. Controls moved on directly to the testing phase, which required classifying novel instances of the semi-artificial grammar as well as reporting confidence levels and the basis for each classification decision. The experiment concluded with a debriefing session during which subjects completed a questionnaire. All task were run on a Dell PC (Windows XP) and delivered via Cedrus SuperLab Pro (version 4.0). The instructions are reproduced in the appendix.

5.2.3.1. Training phase

Experimental participants were trained on the semi-artificial grammar under incidental learning conditions by means of elicited imitations. At the beginning of the experiment, participants were told that they were taking part in a memory study. They were then instructed to listen attentively to 90 syllable sequences and to repeat each sequence after a delayed prompt appeared on the screen ("Please repeat now."). A 1,500 ms delay after stimulus presentation was used to avoid imitations based exclusively on verbatim memory. In order to ensure that subjects repeated the sequences accurately, they were given the option of repeating stimuli if they were unsure about the correct syllable order. For the same reason, subjects' elicited imitations were also monitored by the experimenter via headphones connected to the computer in the testing booth. If subjects were satisfied with the accuracy of their repetitions, they could press a different key to proceed to the next stimulus. Both response keys (*r* for stimulus repetition, / for next stimulus) were displayed below the repetition prompt. Subjects were told that their performance would be evaluated by the accuracy of their repetitions. The figure below illustrates the training procedure.

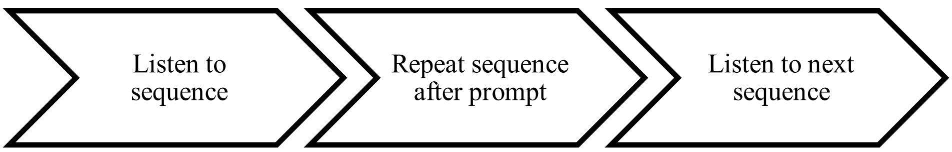

Figure 24. Training procedure in experiment 4.

The presentation order of training items was determined randomly for each participant. The experiment began with a short practice session to familiarize participants with the training task. This consisted of four practice sequences which were not repeated in the actual training phase. The experimenter was present in the testing booth during the practice session to check that subjects were repeating the syllable sequences accurately. The entire training phase took, on average, 30 minutes to complete. The instructions given to participants are reproduced in the appendix.

5.2.3.2. Testing phase

After training, experimental subjects were informed that the syllable arrangement in the previous sequences was determined by a complex rule-system, or "grammar", which might state, for example, "that syllable X must follow syllable Y but not be preceded by syllable Z." Participants were instructed to listen to 60 new syllable sequences, only half of which followed the same rule-system as the previous sequences. Those sequences that followed the rules should be endorsed as "grammatical" and those that did not rejected as "ungrammatical." For each test sequence, participants were required to differentiate grammatical and ungrammatical sequences ("Was the sequence grammatical?"), to report how confident they were in their judgment ("How confident are you in your decision?"), and to indicate the basis of their judgment ("What was the basis of your grammaticality judgment?"). In the case of the control group, participants were merely told that they would listen to 60 syllable sequences, half of which were assembled in accordance with a grammar, and asked to distinguish grammatical sequences from ungrammatical ones. There were no time intervals between the three prompts, i.e. subjects received the next prompt immediately after making a response. There was, however, a 500 ms delay between the source judgment and the presentation of a new test sequence. Participants were given no feedback regarding the accuracy of their classification decisions.

As in experiment 3, experimental and control participants could indicate their confidence levels by selecting one of three response options, namely guess, somewhat confident, and very confident. The four response options for the source attributions were again guess, intuition, memory, and rule knowledge. Participants were provided with the same definitions to these categories as participants in experiment 3. The keyboard options were displayed on the screen below the relevant prompts. The test sequences were presented to each participant in random order. The testing phase began with a short practice session to familiarize the participants with the new task. This consisted of four practice trials which were not repeated in the actual testing set. The entire testing phase took approximately 15 minutes to complete. The instructions given to participants are reproduced in the appendix.

5.2.3.3. Debriefing

At the end of the experiment, all participants were given the same debriefing questionnaire as participants in the previous three experiments (see appendix).

5.3. Results

5.3.1. Testing phase

5.3.1.1. Grammaticality judgments

Overall performance

The analysis of the grammaticality judgments showed that the experimental group classified 65.5% (*SD* = 5.2%) of the test items correctly and the control group 48.3% (*SD* = 6.3%). Further analysis showed that the experimental group performed significantly above chance on this task, $t(14) = 11.277$, $p < .001$, while controls performed at chance, $t(14) = -1.030$, $p > .05$. A one-way ANOVA showed that there were significant differences between the experimental groups 4 and 3 and the control group, $F(2, 41) = 52.847$, $p < .0005$. Post-hoc analysis with Tukey's HSD test showed that experimental group 4 outperformed the control group significantly, $p < .0005$, but not experimental group 3. That is, the training phase produced a clear learning effect in experimental group 4, but the changes made to experiment 3 did not result in increased learning.

In order to establish whether any improvements in accuracy occurred during the testing phase, the grammaticality judgment data was equally divided into 10 chronological stages. The mean accuracy for each block of six judgments was then computed for both groups. A repeated measures ANOVA showed no significant differences across the stages in the experimental group, $F(9, 126) = .650$, $p > .05$, or in the control group, $F(9, 126) = 1.405$, $p > .05$, with

Greenhouse-Geisser correction. That is, no online learning effect was observed during the testing phase. Figure 25 displays the mean accuracy across the 10 stages for both groups.

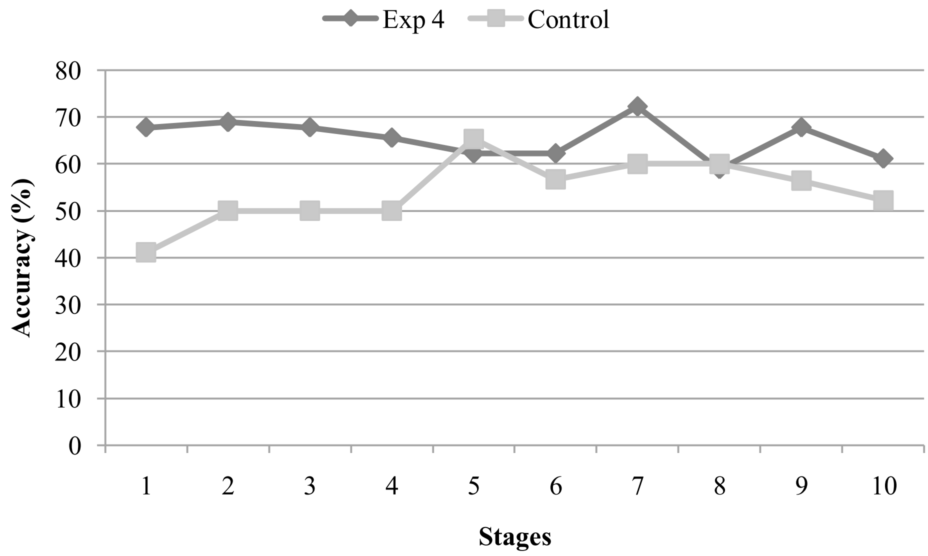

Figure 25. Classification accuracy of experimental and control subjects across 10 chronological stages of the testing set.

Classification performance on grammatical and ungrammatical items

The following analyses report endorsement rates. The results showed that experimental group 4 endorsed 70.4% (*SD* = 9.4%) of all grammatical sequences and 39.3% (*SD* = 10.5%) of all ungrammatical sequences. No experimental subject endorsed more than 90% of grammatical sequences and less than 10% of ungrammatical ones, i.e. there was no evidence of categorical classification performance. The control group endorsed 50.7% (*SD* = 13.2%) of all grammatical items and 54% (*SD* = 14.7%) of all ungrammatical ones, $t(14) = -1.030$, $p > .05$. A one-way ANOVA further showed that there were significant differences between experimental groups 4 and 3 and the control group on grammatical items, $F(2, 41) = 14.112$, $p < .0005$. The assumption of homogeneity of variance was violated. However, the Brown-Forsythe *F*-ratio was nonetheless significant, $F(2, 2.48) = 14.386$, $p < .0005$. Results of the Games-Howell post-hoc analysis indicate that experimental group 4 was significantly more likely to endorse grammatical items than the control group, $p < .005$. With regard to performance on ungrammatical items, a one-way ANOVA confirmed that there were no significant differences among the experimental groups and the control group, $F(2, 41) = 1.185$, $p > .05$. Figure 26 illustrates the performance of experimental and control subjects.

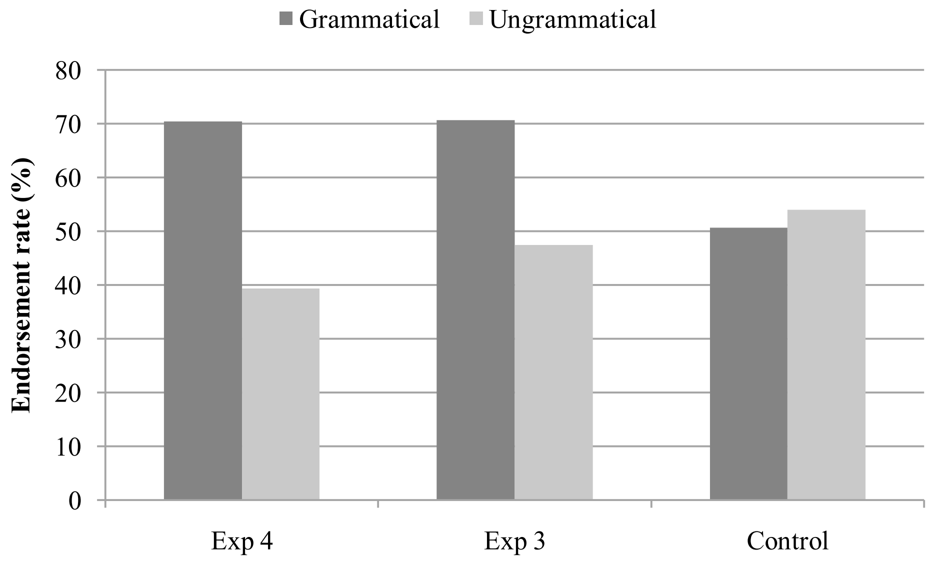

Figure 26. Classification performance of experimental and control subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

Classification performance across sentence types

Figure 27 represents participants' endorsement rates across the nine syntactic patterns employed in this experiment. The scores are summarized in table 25. The analysis indicated that experimental group 4 scored significantly above chance on V2 sequences, $t(14) = 3.162$, $p < .01$ and VF-V1 sequences, $t(14) = 10.425$, $p < .001$. The difference on V2-VF sequences approaches significance, $t(14) = 2.044$, $p < .06$. Experimental group 4 scored significantly below chance on *V1 sequences, $t(14) = 10.248$, $p < .001$, and *V1-VF sequences, $t(14) = 5.434$, $p < .001$. Performance on *V3, *V4, *VF and *VF-V2 sequences was not significantly different from chance. In comparison to experimental group 3 and the controls, a one-way ANOVA showed significant effects between groups for the three grammatical sentence types – V2 items, $F(2, 41) = 4.53$, $p <. 005$, V2-VF items, $F(2, 41) = 9.574$, $p < .0005$, VF-V1 items, $F(2, 41) = 16.05$, $p < .0005$ – and for several ungrammatical sentence types, namely *V1 items, $F(2, 41) = 3.956$, $p < .05$, *V3 items, $F(2, 41) = 25.341$, $p < .005$, *VF items, $F(2, 41) = 3.858$, $p < .05$, and *VF-V1 items, $F(2, 41) = 5.713$, $p < .01$. There were no effects on *V4 and *VF-V2 sentences. The assumption of homogeneity of variance was violated for V2, *V3, *VF and VF-V1 sentences. However, the Brown-Forsythe F-ratios were nonetheless significant on V2 constructions, $F(2, 31.6) = 4.536$, $p < .05$, *V3 constructions, $F(2, 27.87) = 24.886$, $p < .0005$, *VF constructions, $F(2, 42.682) = 3.928$, $p < .05$, and VF-V1 sentences, $F(2, 31.23) = 15.974$, $p < .0005$.

Results of the post-hoc analyses, using Tukey's HSD test for *V1, V2-VF and *V1-VF items and Games-Howell procedure for the remaining items, showed that experimental group 4 was significantly more likely than the controls to endorse V2-VF sentences, $p < .0005$ and VF-V1 sentences, $p < .0005$, and significantly less likely to endorse *V1 sentences, $p < .05$. In comparison to experimental group 3, experimental group 4 was significantly more likely to endorse VF-V1 sentences, $p < .0005$, and *VF-V2 sentences, $p < .05$, and significantly less likely to endorse *V1 sentences, $p < .05$, *VF sentences, $p < .05$, and *V1-VF sentences, $p < .01$.

The performance on five-syllable sequences indicated that experimental group 4 preferred V2 over *V1 sequences, $t(14) = -10.694$, $p < .0005$, and *VF sequences, $t(14) = 2.432$, $p < .05$. However, experimental group 4 was just as likely to endorse V2 as *V3 sequences, $t(14) = 1.781$, $p > .05$, and *V4 sequences, $t(14) = .816$, $p > .05$. The performance on complex constructions showed that experimental group 4 preferred V2-VF over *V1-VF sequences, $t(14) = 6.025$, $p > .0005$, and VF-V1 sequences over *VF-V2 sequences, $t(14) = 5.093$, $p < .05$.

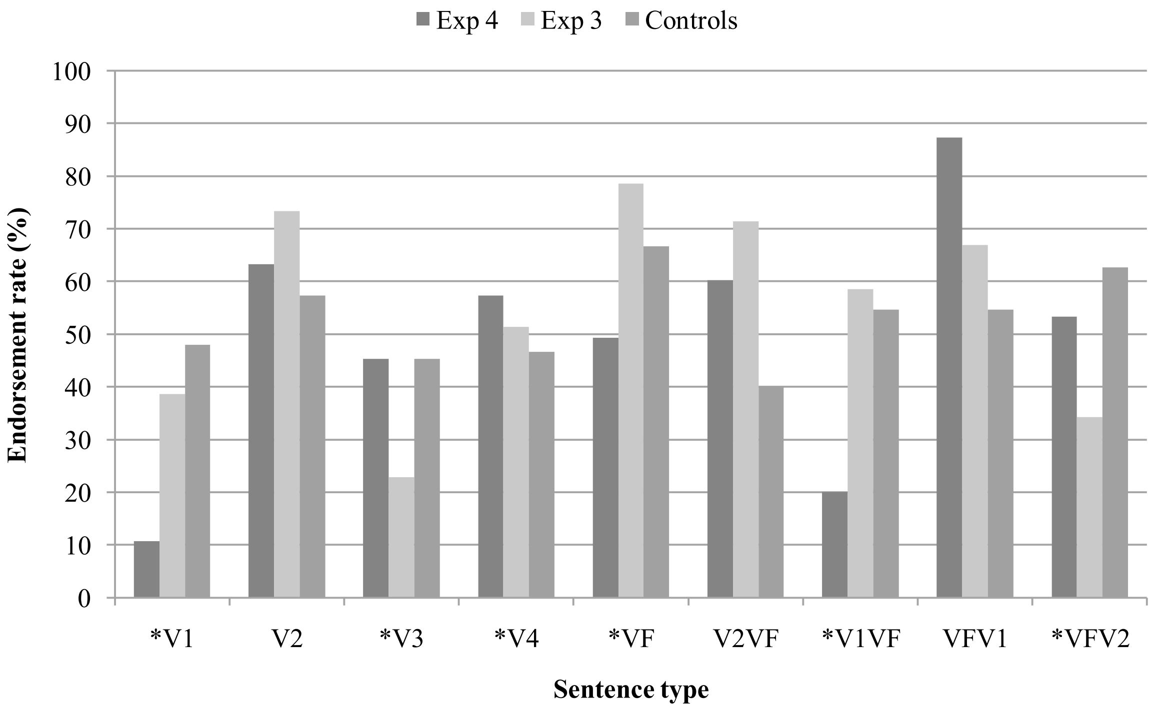

Figure 27. Classification performance of experimental group 3, experimental group 4 and control group across sentence types. Bars represent endorsement rates.

Table 25
Classification Performance (%) of Experimental and Control Subjects across Sentence Types

Group		Sentence type								
		*V1	V2	*V3	*V4	*VF	V2-VF	*V1-VF	VF-V1	*VF-V2
Exp 4	*M*	10.7**	63.3*	45.3	57.3	49.3	60.2^{+}	20**	87.3**	53.3
	SD	14.9	16.3	35	24.9	28.1	19.3	21.4	13.9	20.9
	SE	3.8	4.2	9	6.4	7.3	5	5.5	3.6	5.4
Exp 3	*M*	38.6	73.4*	22.9*	51.4	78.6**	71.4**	58.6	66.9*	34.3
	SD	36.3	27.6	35.8	33	22.8	18.8	35.5	29.2	30.8
	SE	9.7	7.4	9.6	8.8	6.1	5	9.5	7.8	8.2
Control	*M*	48	57.3	45.3	46.7	66.7*	40	54.7	54.7	62.7^{+}
	SE	31.9	23.1	22	24.7	28.9	21	26.7	18.8	23.7
	SD	8.2	6	5.7	6.4	7.5	5.4	6.9	4.9	6.1

Significance from chance: $^{+}p < .1$, * $p < .05$, ** $p < .001$.

5.3.1.2. Confidence ratings

Overall performance

The following analyses only report the results of the experimental group. The average confidence level in the experimental group 4 was 4.8 (*SD* = 0.8). In terms of proportion, experimental participants tended to select the option somewhat confident most frequently and the guess option least frequently. In terms of accuracy, the analysis indicated that experimental participants were most accurate when reporting to be very confident in their decisions and considerably less accurate when reporting to be somewhat confident. Accuracy was lowest for those grammaticality decisions in which subjects had no confidence whatsoever. Experimental participants scored significantly above chance when reporting to be somewhat confident and very confident. When participants reported to be guessing, performance was indistinguishable from chance. According to the guessing criterion, there was no evidence of unaware learning. Table 26 summarizes the findings.

Table 26
Accuracy and Proportions (%) across Confidence Ratings

	Guess	Somewhat confident	Very confident
Accuracy	50	65**	86**
Proportion	14	64	24

Significance from chance: * $p < .0005$.

In order to determine whether participants became more confident while performing on the testing phase, the confidence ratings were equally divided into 10 chronological stages, and the average confidence levels were calculated for each stage. A repeated measures ANOVA showed no significant differences in mean confidence across the stages, $F(9, 126) = 1.107$, $p > .05$. That is, experimental group 4 did not display significant changes in confidence levels while performing on the testing phase. In addition, the proportion of guess, somewhat confident and very confident responses were calculated for each of the 10 blocks. As figure 28 suggests, there were relatively few changes in the response proportions across the testing phase. The analysis confirms that participants did not become more confident while performing on the grammaticality judgment task.

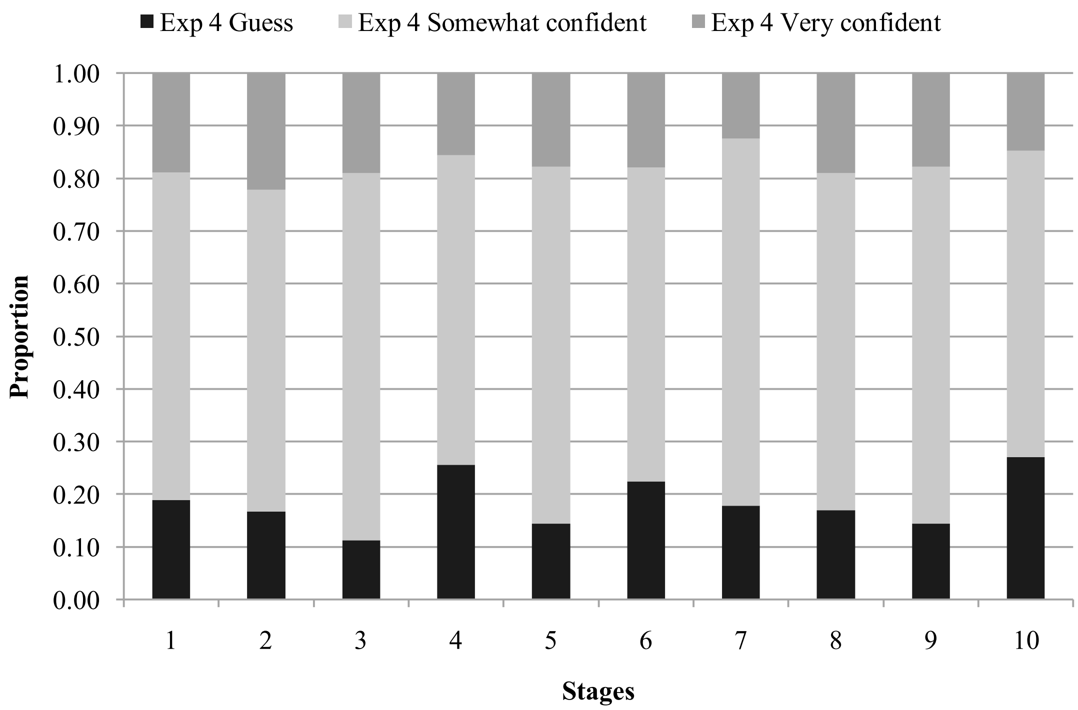

Figure 28. Proportions of guess, somewhat confident and very confident responses across 10 stages of the testing set.

Zero correlation criterion

The Chan difference score was computed in order to establish whether learning in the experimental group was implicit by the zero correlation criterion. The average confidence for correct grammaticality decisions was 5.2 ($SD = 0.9$) and the average confidence for incorrect decisions was 3.9 ($SD = 0.8$), i.e. experimental participants were more confident in correct decisions than in incorrect ones. The difference (1.3) was significant, $t(14) = 5{,}791$, $p < .0005$, which further confirms that participants were aware of acquiring knowledge. According to the zero correlation criterion, there was at least some conscious knowledge in the experimental group.

5.3.1.3. Source attributions

Overall performance

In terms of proportion, experimental participants most frequently believed their classification decisions to be based on intuition, followed by memory and guesses. The rule knowledge category was selected least frequently. In terms of accuracy, experimental participants scored highest when reporting to rely on memory for training sequences to guide their decisions, followed by the rule knowledge and intuition categories. Subjects were least accurate when basing decisions on guesses. Participants performed significantly above chance when basing their decisions on intuition, memory, or rule knowledge.

Table 27
Accuracy and Proportions (%) across Source Attributions

	Guess	Intuition	Memory	Rule
Accuracy	55	65**	74**	72**
Proportion	13	54	23	10

Significance from chance: * $p < .005$.

Relationship between confidence ratings and source attributions

Mean confidence ratings were calculated for each source attribution in order to assess the relationship between these types of judgment. Participants reported the highest confidence levels for decisions based on memory (M = 5.97), followed by decisions based on rule knowledge (M = 5.87), intuition (M = 4.76) and guesses (M = 1.93). Participants also gave significantly higher confidence levels for attributions associated with conscious structural knowledge (rules and memory, combined M = 5.92) rather than unconscious structural knowledge (guess and intuition, combined M = 3.35), $p < .05$.

Table 28 summarizes the mean accuracy score for pooled confidence ratings and source attributions. In terms of proportion, the analysis indicates that the majority of grammaticality judgments were either somewhat confident and based on intuition or very confident and based on rule knowledge. In terms of accuracy, participants performed significantly above chance when making somewhat confident decisions based on intuition, which further suggests that participants acquired unconscious structural knowledge. Participants also performed significantly above chance on somewhat confident decisions based on rule knowledge, and on very confident decisions based on rule knowledge. In the remaining cases, performance was not significantly different from chance.

Table 28
Accuracy and Proportions (%) for Combined Confidence Ratings and Source Attributions

		Source attributions			
Confidence ratings		Guess	Intuition	Memory	Rule
Guess	Accuracy	46.6	52.1	52.9	50
	Proportions	8.1	7.9	1.9	0.7
Somewhat confident	Accuracy	54.3	63.7**	68.4**	70.2*
	Proportions	5.1	39.6	13	6.3
Very confident	Accuracy		86.4**	87.3**	81.5*
	Proportions	0	6.6	7.9	3

Significance from chance: * $p < .01$. ** $p < .0005$.

5.3.2. Debriefing

5.3.2.1. Accuracy estimates

Participants' assessment of their performance on the grammaticality judgment task (*estimated accuracy*) was compared to their actual performance (*observed accuracy*). The comparison showed that experimental group 4 was very accurate in estimating their performance. The difference between estimated accuracy (M = 63.3%, SD = 13.6%) and observed accuracy (M = 65.5%, SD = 5.2%) was 2.2%. This difference is not significant, $t(14) = .651$, $p > .05$, which suggests that experimental subjects were aware of acquiring knowledge. The analysis thus supports the view that the experimental subjects were at least partially aware of having acquired the knowledge that guided their grammaticality decisions.

5.3.2.2. Verbal reports

The transcription of all verbal reports can be found in the appendix. The analysis showed that 12 out of 15 experimental subjects reported knowledge when prompted to describe any rules or regularities they might have noticed. Control subjects did not report noticing any rules. The verbalizations consisted of descriptions of bigrams and trigrams that were licensed by the grammar employed in this experiment. Subjects 10, 12 and 14 reported possible initial, middle and end sequences. Subject 7 only reported initial and end sequences and subject 5 only middle and end sequences. Interestingly, the reports of two subjects indicate that their knowledge was more abstract. Subject 11 noticed the frequent occurrence of "(f-) (c-) (s-) (p-) (t-)", and subject 7 reported that sequences often finished with a k-syllable and that the last three syllables featured "*p*, *t* and *k* sounds." Subject 10 reported the application of chunking as a strategy for training and testing tasks. She reported breaking training sequences down into two or three parts in order to facilitate recollection for the elicited imitation task; in the testing phase, she tried to recall the chunks and make classification decisions with basis on the presence or absence of recalled chunks. Interestingly, no subject reported noticing the fact that, in longer sequences, two k-syllables could follow each other, which supports the view that subjects are more likely to notice items in anchor positions.

5.4. Discussion

Experiment 4 investigated how linguistic knowledge affected the acquisition of the semi-artificial grammar by replacing English words with nonsense syllables. The results showed that this alteration did not result in a significant increase in overall performance (65.6%, compared to 61.6% in experiment 3). In contrast to the previous experiment, however, performance was not driven exclusively by the above-chance endorsement of grammatical structures. Experimental subjects in experiments 4 and 3 performed similarly on grammatical items (70.4% and 70.6%, respectively), but the former group was less likely to endorse ungrammatical stimuli (39.3% and 47.4%). While the difference between groups on ungrammatical items failed to reach significance, only experimental group 4 performed significantly below chance of ungrammatical items in terms of endorsement. Experiment 4 thus shows an improvement in rejection performance. Further analysis showed that subjects in experiments 3 and 4 performed surprisingly similarly across sentence patterns. Experimental group 4 was less likely than experimental group 3 to endorse *VF sequences, but performance was still at chance (49.3%), which indicates that subjects did not learn the VF rule. Both groups also performed similarly on complex sequences. Experimental group 4 was more likely to endorse VF-V1 sentences, but there were no differences on V2-VF, *V1-VF and *VF-V2.[26] Classification performance in experiment 4 thus suggests that pre-existing knowledge, in the form of metalinguistic categories, did not affect performance negatively in the previous experiments.

[26] Given that there were only two types of "k"-syllables, the improved performance on VF-V1 sentences is not particularly surprising, as subjects will have heard the possible repetitions (ki-ki, ki-ko, etc.) before.

In terms of awareness, experiment 4, like experiment 3, provided evidence for both conscious and unconscious knowledge. When subjects reported no confidence in their grammaticality judgments, they performed at chance, i.e. the guessing criterion for implicit learning was not met. Subjects also tended to be significantly more accurate in high confidence decisions, which suggests that they were partially aware of the knowledge they had acquired. The analysis of the accuracy estimates showed a good match between estimated and observed accuracy, which supports the view that subjects had developed conscious judgment knowledge. The source attributions, on the other hand, indicated that subjects were also partially unaware of the knowledge they had acquired. When attributing grammaticality judgments to intuition, subjects performed significantly above chance, i.e. they had developed unconscious structural knowledge. Experiment 4 found that the removal of semantic information from the semi-artificial language did not result in more implicit learning.

6. Experiment 5

6.1. Introduction

Experiment 5 investigates the acquisition of verb placement rules of a semi-artificial language under incidental learning conditions. The stimulus material and the testing procedure are the same as the ones used in experiment 3. In contrast to experiment 3, however, the present experiment utilizes a novel training task that was designed to bring the notions of clause type and clause sequence to subjects' attention. The aim was to determine whether the addition of this training task increased learning in general and the acquisition of the different verb placement rules in particular.

6.2. Method

6.2.1. Participants

Fifteen native speakers of English (8 men and 7 women, mean age = 22.4 years) were recruited to take part in this experiment. All participants were students at the University of Cambridge at the time of the experiment. Participants spoke, on average, 2.6 languages and none had a background in German or any other V2-language. Two participants were brought up bilingually. The data of the 15 untrained subjects from experiment 3 served as control data. Experimental and control groups did not differ significantly across age, $t(13.89) = 1.387$, $p > .05$, gender, $\chi^2(1, N = 30) = 3.589$, $p > .05$, occupation (student vs. nonstudent), $\chi^2(1, N = 30) = 2.143$, $p > .05$, and number of languages acquired, $\chi^2(3, N = 30) = 3.829$, $p > .05$. All subjects received £5 for participating in the experiment.

6.2.2. Stimulus material

The 180 stimulus sentences created for experiment 3 were used in this experiment. The linguistic focus was on the three verb placement rules described in table 14.

6.2.3. Procedure

The experiment consisted of (i) a training phase, (ii) a testing phase, and (iii) a debriefing session. Experimental participants were trained on the semi-artificial grammar by means of a sentence classification task and plausibility judgments. In the testing phase, participants were asked to classify novel instances of the semi-artificial grammar and to report confidence levels and the basis for each classification decision. The experiment concluded with a debriefing session during which subjects completed a questionnaire. All tasks were run on a Dell PC (Windows XP) and delivered via Cedrus SuperLab Pro (version 4.0). The instructions are reproduced in the appendix.

6.2.3.1. Training phase

At the beginning of the training phase, participants were told they were taking part in an experiment on sentence comprehension during which they would be exposed to 120 "scrambled" sentences. Subjects were then provided with a short briefing on sentence structure. Specifically, participants were told that sentences could consist of three types, namely (1) "main clause only", (2) "main clause, followed by a subordinate clause", and (3) "subordinate clause, followed by a main clause." The task instructions required participants to listen to the training sentences on an item-by-item basis, to classify each sentence according to its "type" (1-3), and to judge whether the statements were semantically plausible or not. The figure below illustrates the training procedure.

Figure 29. Training procedure in experiment 5.

The response keys for sentence classifications (*1* = main clause only, *2* = main > subordinate, *3* = subordinate > main) and plausibility judgments (*/* = plausible, *z* = implausible) were displayed on the screen below the relevant prompts ("Enter the sentence type." and "Do you think this is plausible?"). When participants classified the structure of a sentence incorrectly or misjudged its plausibility, a sound was played to indicate that a mistake had been made. The sentence was then repeated and participants were required to repeat both judgments. Participants only proceeded to the next sentence once they had correctly classified the sentence type and its plausibility. The presentation order of training items was determined randomly for each participant. The experiment began with a short practice session to familiarize participants with the training task. This consisted of four practice sentences which were not repeated in the actual training phase. The completion of the training phase took approximately 40 minutes.

6.2.3.2. Testing phase

The testing procedure was identical to the one employed in experiment 3.

6.2.3.3. Debriefing

At the end of the experiment, all participants were given the same debriefing questionnaire as participants in the previous experiments (see appendix).

6.3. Results

6.3.1. Training phase

Performance during the training phase was assessed by analyzing the sentence identifications and the plausibility judgments. In both cases, the proportion of correct judgments was calculated for each participant.

6.3.1.1. Sentence classifications

The overall accuracy for experimental group 5 was 93% (*SD* = 9.3%). Participants judged 94% (*SD* = 8.6%) of V2 sentences correctly, 92.4% (*SD* = 10.7%) of V2-VF sentences, and 94.1% (*SD* = 8.9%) of VF-V1 sentences. These differences were not significant, $F(2, 42) = .154$, $p > .05$, which suggests that participants did not find complex sentences more difficult to judge than simple sentences.

6.3.1.2. Plausibility judgments

In terms of frequency, participants committed, on average, 12.6 errors in their plausibility judgments. The analysis of the plausibility judgments showed that the experimental group judged 90.5% (*SD* = 5.3%) of sentences correctly. Participants were significantly more accurate when judging plausible sentences (*M* = 97%, *SD* = 2.8%) than implausible ones (*M* = 83.8%, *SD* = 9.4%), $t(12) = 5.308$, $p < 0.001$, which suggests that implausibility affected accuracy in the plausibility judgment task negatively.

6.3.2. Testing phase

6.3.2.1. Grammaticality judgments

Overall performance

The analysis of the grammaticality judgments showed that experimental group 5 classified 67.2% (*SD* = 6.4%) of the test items correctly. Overall performance was significantly above chance, $t(14) = 10.317$, $p < .001$. That is, the training phase produced a clear learning effect in the experimental group. A one-way ANOVA further showed that there were significant differences among experimental groups 5, 4, 3 and the control group, $F(3, 55) = 45.82$, $p < .0005$. The assumption of homogeneity of variance was violated. However, the Brown-Forsythe *F*-ratio was nonetheless significant, $F(3, 44.8) = 45.06$, $p < .0005$. Results of the Games-Howell post-hoc analysis indicate that experimental group 5 only significantly outperformed the control subjects, $p < .0005$, but not the other experimental groups in terms of overall accuracy.

In order to establish whether any improvements in accuracy occurred during the testing phase, the grammaticality judgment data was equally divided into 10 chronological stages. The mean accuracy for each block of six judgments was then computed. A repeated measures ANOVA showed no significant differences across the stages in the experimental group, $F(9, 126) = .586$, $p > .05$. That is, no online learning effect was observed during the testing phase. Figure 30 displays the mean accuracy across the 10 stages for experimental group 5.

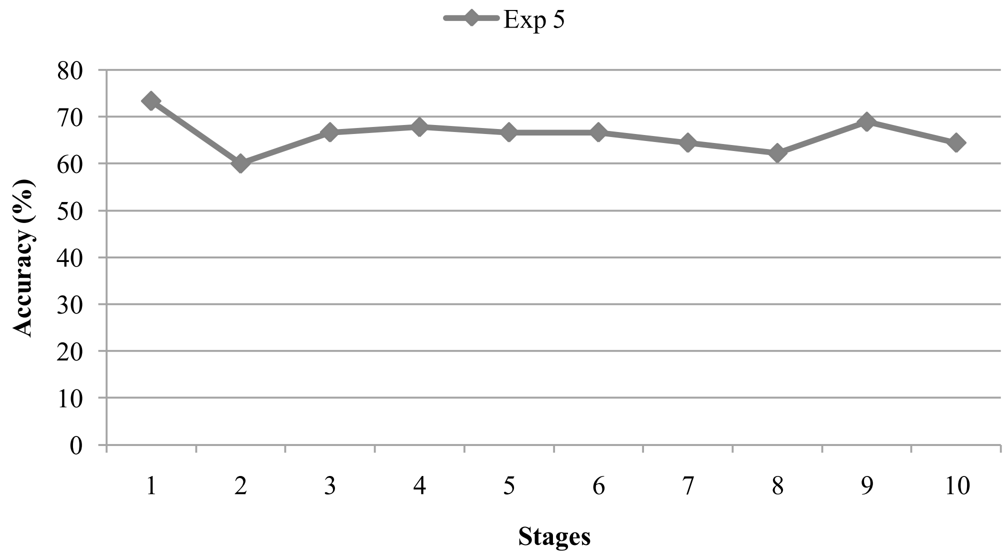

Figure 30. Classification accuracy of experimental group 5 across
10 chronological stages of the testing set.

Classification performance on grammatical and ungrammatical items

The following analyses report endorsement rates. The results showed that experimental group 5 endorsed 71% of all grammatical sentences and 37% of all ungrammatical ones, $t(14) = 10.260$, $p < .0005$. No subject endorsed more than 90% of grammatical sequences and less than 10% of ungrammatical ones, i.e. there was no evidence of categorical classification performance. Performance on grammatical items was significantly above chance, $t(14) = 5.585$, $p < .001$, while performance on ungrammatical items was significantly below chance, $t(14) = -3.931$, $p < .05$. A one-way ANOVA further showed that there were significant differences among experimental groups 5, 4, 3 and the control group on grammatical items, $F(3, 55) = 12.259$, $p < .0005$. The assumption of homogeneity of variance was violated. However, the Brown-Forsythe *F*-ratio was nonetheless significant, $F(3, 29.42) = 12.399$, $p < .0005$. Results of the Games-Howell post-hoc analysis indicate that experimental group 5 was significantly more likely to endorse grammatical items than the control group, $p < .05$, but not less likely than the other experimental groups. With regard to performance on ungrammatical items, a one-way ANOVA showed that there were no significant differences among the groups, $F(3, 55) = 1.782$, $p > .05$. The scores are summarized in table 29.

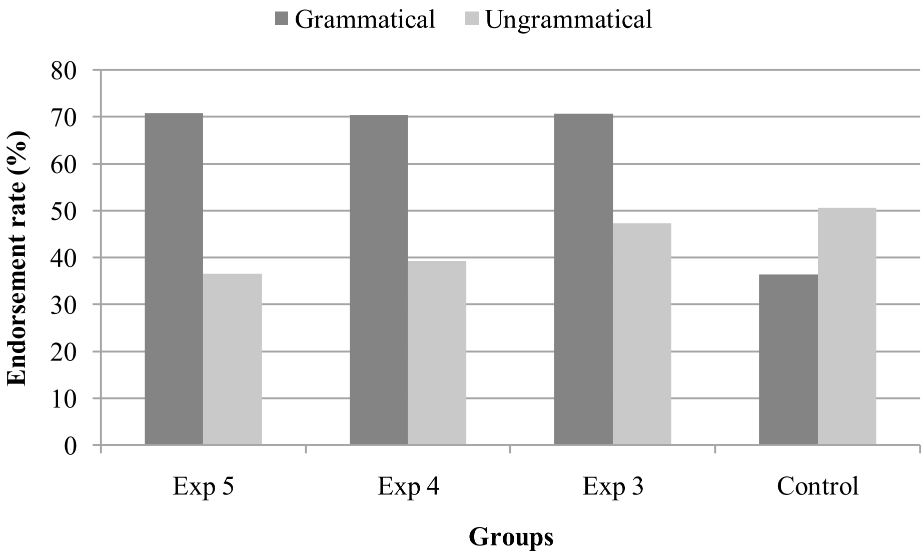

Figure 31. Classification performance of experimental and control subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

Table 29
Classification Performance (%) of Experimental and Control Subjects on Grammatical and Ungrammatical Stimuli

Group		Grammatical	Ungrammatical
Exp 5	*M*	70.8**	36.6**
	SD	14.4	13.2
	SE	3.7	3.4
Exp 4	*M*	70.4**	39.3**
	SD	9.4	10.5
	SE	2.4	2.7
Exp 3	*M*	70.6**	47.4
	SD	13.5	19.7
	SE	3.5	5.1
Control	*M*	36.4	50.6
	SD	30.3	28.4
	SE	7.8	7.3

Significance from chance: * $p < .05$. ** $p < .001$.

Classification performance across sentence types

Figure 32 illustrates participants' endorsement rates across the nine syntactic patterns employed in this experiment. The scores are summarized in table 30. The analysis indicated that experimental group 5 performed significantly above chance on all grammatical pattern, i.e. V2 sentences, $t(14) = 2.485$, $p < .05$, V2-VF sentences, $t(14) = 3.106$, $p < .05$, and VF-V1, $t(14) = 6.943$, $p < .001$. Experimental group 5 scored significantly below chance on *V3 sentences, $t(14) = 4.603$, $p < .001$, and *VF-V2 sentences, $t(14) = 4.171$, $p < .01$. In comparison to the other experimental groups, a one-way ANOVA showed significant effects between groups for the three grammatical sentence types – V2 items, $F(3, 55) = 3.347$, $p < .05$, V2-VF items, $F(3, 55) = 7.661$, $p < .0005$, VF-V1 items, $F(3, 55) = 14.474$, $p < .005$ – and for two ungrammatical sentence types, namely *V3 items, $F(3, 55) = 24.121$, $p < .0005$, and *VF-V2 items, $F(3, 55) = 3.348$, $p < .05$. The effects on *V1 sentences, $F(3, 55) = 2.631$, $p = .059$, and *VF sentences, $F(3, 55) = 2.39$, $p = .079$, approach significance. The assumption of homogeneity of variance was violated for V2, *V3, V2-VF, *VF-V2 and VF-V1 sentences. However, the Brown-Forsythe F-ratios were nonetheless significant on all constructions: V2, $F(3, 45.47) = 3.354$, $p < .05$, *V3, $F(3, 40.74) = 23.84$, $p < .0005$, V2-VF, $F(4, 47.69) = 7.711$, $p < .005$, VF-V1, $F(3, 37.17) = 14.348$, $p < .0005$, and *VF-V2, $F(3, 45.173) = 3.340$, $p < .05$. Results of the Games-Howell post-hoc analysis showed that experimental group 5 was significantly more likely than the controls to endorse V2 sentences, $p < .05$, V2-VF sentences, $p < .05$, and VF-V1 sentences, $p < .005$, and significantly less likely to endorse *V3 sentences, $p < .005$. Experimental group 5 was also significantly less likely than experimental group 4 to endorse *VF-V2 sentences, $p < .005$.

The performance on one-clause constructions indicated that experimental group 5 preferred V2 sentences over *V1 sentences, $t(14) = -2.942$, $p < .05$, *V3 sentences, $t(14) = 4.975$, $p < .0005$, and *V4 sentences, $t(14) = 3.686$, $p < .05$. However, experimental group 5 was just as likely to endorse V2 as *VF sentences, $t(14) = .496$, $p > .05$, which suggests that subjects did not know that verb-final is limited to subordinate clauses. The performance on complex constructions showed that experimental group 5 preferred V2-VF sentences over *V1-VF sentences, $t(14) = 2.849$, $p < .05$, and VF-V1 sentences over *VF-V2 sentences, $t(14) = 10.101$, $p < .0005$. That is, subjects had a clear preference for grammatical (i.e., trained) structures over ungrammatical ones.

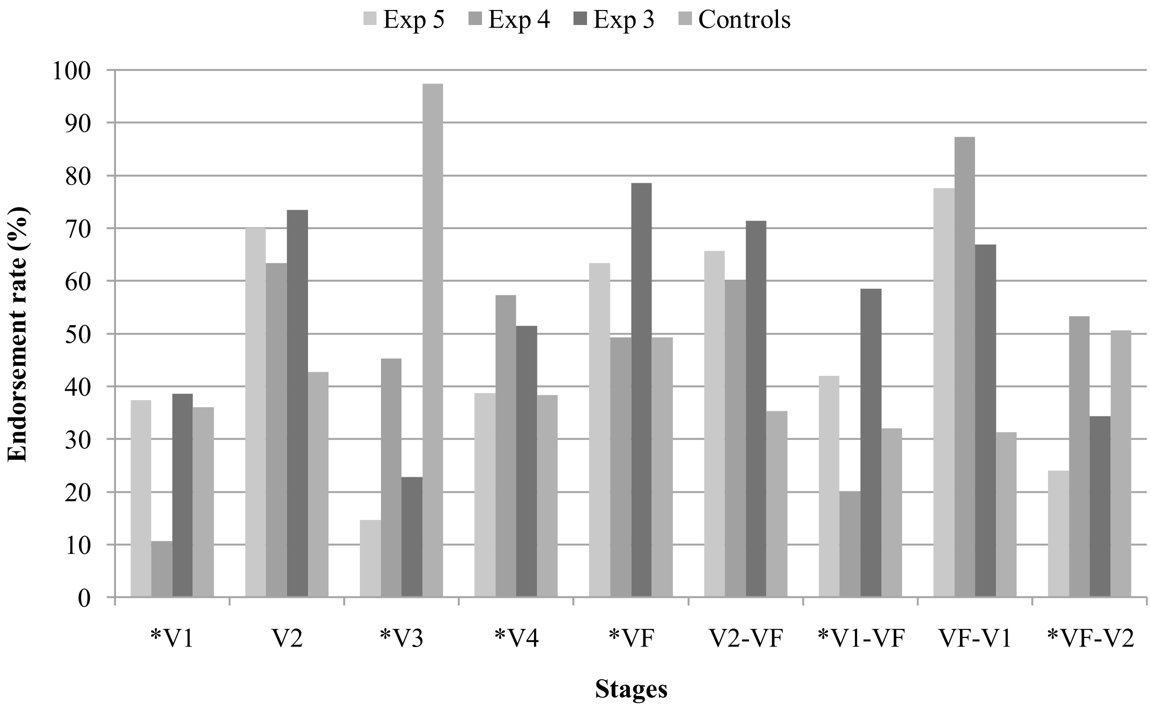

Figure 32. Classification performance of experimental and control subjects across sentence types. Bars represent endorsement rates.

Table 30
Classification Performance (%) of Experimental and Control Subjects across Sentence Types

		Sentence type								
Group		*V1	V2	*V3	*V4	*VF	V2-VF	*V1-VF	VF-V1	*VF-V2
Exp 5	*M*	37.3	70*	14.7**	38.7	63.3	65.7*	42	77.6**	24*
	SD	36.9	31.2	29.7	27.7	38.9	19.6	28.8	15.4	24.1
	SE	9.5	8	7.7	7.2	10	5.1	7.4	4	6.2
Exp 4	*M*	10.7**	63.3*	45.3	57.3	49.3	60.2^{+}	20**	87.3**	53.3
	SD	14.9	16.3	35	24.9	28.1	19.3	21.4	13.9	20.9
	SE	3.8	4.2	9	6.4	7.3	5	5.5	3.6	5.4
Exp 3	*M*	38.6	73.4*	22.9*	51.4	78.6**	71.4**	58.6	66.9*	34.3
	SD	36.3	27.6	35.8	33	22.8	18.8	35.5	29.2	30.8
	SE	9.4	7.1	9.3	8.5	5.9	4.8	9.2	7.5	8
Controls	*M*	36	42.7	97.3**	38.3	49.3	35.3	32	31.3+	50.7
	SD	34.0	36.7	7.0	37.0	42.7	28.8	35.3	34.8	38.4
	SE	8.8	9.5	1.8	9.6	11	7.4	9.1	9	9.9

Significance from chance: $^{+}p < .1$ * $p < .05$. ** $p < .001$.

6.3.2.2. Confidence ratings

Overall performance

The average confidence level in the experimental group was 5.2 (*SD* = 1.3). In terms of proportion, experimental participants tended to select the somewhat confident option most frequently and the guess option least frequently. In terms of accuracy, the analysis indicated that experimental participants were most accurate when reporting to be very confident in their decisions. Accuracy was lower for those grammaticality decisions in which subjects were only somewhat confident or not confident at all. Experimental participants scored significantly above chance on all three confidence categories, including guesses. According to the guessing criterion there was unconscious judgment knowledge in experimental group 5. Table 31 summarizes the scores.

Table 31

Accuracy and Proportions (%) across Confidence Ratings

	Guess	Somewhat confident	Very confident
Accuracy	63*	65**	77**
Proportion	14	64	24

Significance from chance: * $p < .05$. ** $p < .001$.

In order to determine whether participants became more confident while performing on the testing phase, the confidence ratings were equally divided into 10 chronological stages, and the average confidence levels were calculated for each stage. A repeated measures ANOVA showed no significant differences in mean confidence across the stages, $F(9, 126) = .678$, $p > .05$. That is, experimental group 5 did not display significant changes in confidence levels while performing on the testing phase. In addition, the proportion of guess, somewhat confident and very confident responses were calculated for each of the 10 stages. As figure 33 suggests, there were relatively few changes in the response proportions across the testing phase. The analysis confirms that participants did not become more confident while performing on the grammaticality judgment task.

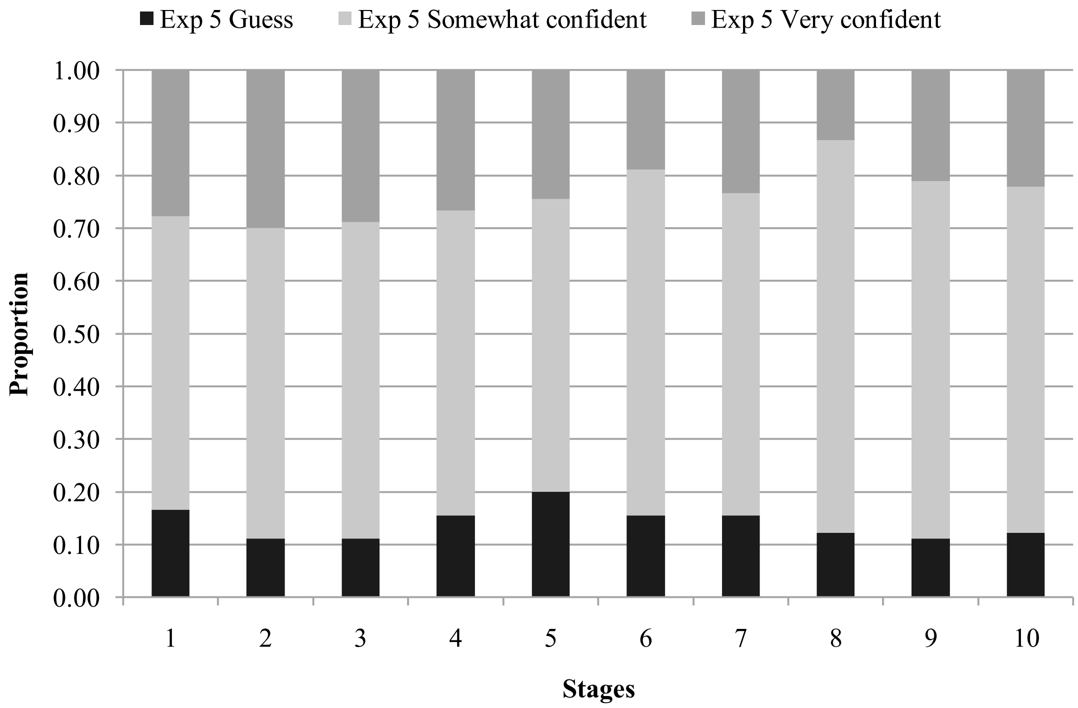

Figure 33. Proportions of guess, somewhat confident and very confident responses across 10 stages of the testing set.

Zero correlation criterion

The Chan difference score was computed in order to establish whether learning in the experimental group was implicit by the zero correlation criterion. The average confidence for correct grammaticality decisions was 5.3 (SD = 1.4) and the average confidence for incorrect decisions was 4.9 (SD = 1.2), i.e. experimental participants were more confident in correct decisions than in incorrect ones. The difference (0.4) was significant, $t(14) = 2.299$, $p < .05$, which indicates that participants were partially aware of the knowledge they acquired. According to the zero correlation criterion, there was at least some conscious judgment knowledge in the experimental group.

6.3.2.3. Source attributions

Overall performance

In terms of proportion, experimental participants most frequently believed their classification decisions to be based on intuition, followed by rule knowledge and memory. The guess category was selected least frequently. In terms of accuracy, experimental participants scored highest when reporting to use rule knowledge to guide their decisions, followed by the memory and intuition categories. Subjects were least accurate when basing decisions on guesses. Participants only performed significantly above chance when basing their decisions on intuition, memory or on rule knowledge. They did not perform significantly above chance on decisions based on guesses.

Table 32
Accuracy and Proportions (%) across Source Attributions

	Guess	Intuition	Memory	Rule
Accuracy	55.3	62.9*	69.5*	73.6*
Proportion	4.2	45.6	18.4	31.8

Significance from chance: * $p < .001$.

Relationship between confidence ratings and source attributions

Participants reported the highest confidence levels for decisions based on rule knowledge ($M = 6.31$), followed by decisions based on memory ($M = 6.05$), intuition ($M = 4.59$) and guesses ($M = 0.66$). Participants also gave significantly higher confidence levels for attributions associated with conscious knowledge (rules and memory, combined $M = 6.18$) rather than unconscious knowledge (guess and intuition, combined $M = 2.63$), $p < .05$.

Table 33 summarizes the mean accuracy score for pooled confidence ratings and source attributions. In terms of proportion, the analysis indicates that the majority of grammaticality judgments were either somewhat confident and based on intuition or rule knowledge or very confident and based on rule knowledge. In terms of accuracy, participants performed significantly above chance on classification decisions based on intuition, which suggests the presence of conscious judgment knowledge and unconscious structural knowledge. Participants also performed significantly above chance on somewhat confident and very confident grammaticality decisions based on intuition, memory or rule knowledge.

Table 33
Accuracy and Proportions (%) for Combined Confidence Ratings and Source Attributions

		Source attributions			
Confidence ratings		Guess	Intuition	Memory	Rule
Guess	Accuracy	60.6	63.1*	75	62.5
	Proportion	3.7	7.2	0.6	2.7
Somewhat confident	Accuracy	20	60.9**	60.9*	77.5**
	Proportion	0.6	34	12.3	15.3
Very confident	Accuracy		78.9**	88**	71.2**
	Proportion	0	4.3	5.6	13.8

Significance from chance: * $p < .05$. ** $p < .001$.

6.3.3. Debriefing

6.3.3.1. Accuracy estimates

Participants' assessment of their performance on the grammaticality judgment task was compared to their actual performance. The comparison showed that experimental group 5 was not very accurate in estimating its performance. The difference between estimated accuracy (M = 56.9%, SD = 21.1%) and observed accuracy (M = 67.2%, SD = 6.4%) was 10.3. This difference was significant, $t(14) = 2.170$, $p < .05$, i.e. participants in this experiment significantly underestimated their performance. The analysis of the accuracy estimates thus supports the view that subjects were partially unaware of the knowledge that guided their grammaticality decisions.

6.3.3.2. Verbal reports

The transcription of all verbal reports can be found in the appendix. The analysis showed that all subjects reported knowledge when prompted, at the end of the experiment, to describe any rules or regularities they might have noticed. Most participants did not rely on phrasal categories in their reports but employed functional categories and semantic categories instead. In comparison to experiment 3, subjects in this experiment were more likely to mention clause types in their reports, which suggests that the sentence identification task drew their attention to this linguistic element. Importantly, the analysis also showed that several subjects were able to report rules accurately or to provide sentences that were licensed by the semi-artificial grammar. Still, no subject was able to verbalize all rules that determined the placement of VP.

As in previous experiments, subjects frequently reported a “reversed” order of subject and verb and that adverbs (e.g. expressions of time) came at the beginning of sentences (see e.g. subjects 3, 9, 16, 18 and 19). Subjects 8 and 17 provided correct examples of simple clauses (= V2 sentences). Other subjects reported that a verb could be placed at the end but did not mention that this could only occur in subordinate clauses (5, 6, 7, 11 and 13).[27] Subject 7 reported the correct phrasal sequence for subordinate clauses but the incorrect one for main clauses. Only one subject (6) reported that only subordinate clauses are verb-final. Interestingly, this subject nonetheless endorsed 50% of *VF sentences as grammatical, which could suggest that the rule only occurred to her after the end of the testing phase. Two subjects (3, 18) reported the correct word order for VF-V1. Finally, the analysis also showed that several subjects also reported irrelevant rules. Subject 16 reported, for example, that type 2 sentences featured the subordinate “that”, while type 3 sentences featured the subordinators “since”, “when” and “after”. Subject 15, on the other hand, reported (erroneously) that “proper nouns” were used more frequently in the testing set.

6.4. Discussion

Experiment 5 introduced a new training task to the design of experiment 3. The stimulus material and the testing phase were the same in both experiments, but subjects in experiment 5 were required to identify the sentence structure of each stimulus before judging its semantic plausibility. As mentioned above, the aim was to establish whether the addition of a training task that required subjects to process clause type and clause sequence more directly had a positive effect on learning. The results showed that the introduction of the sentence identification task produced a greater learning effect in experimental group 5 (67.2%, compared to 61.6% in experiment 3). The analysis also indicated that the difference between experimental group 5 and experimental group 3 was due to better rejection of ungrammatical stimuli in the former group. Performance on grammatical items was practically identical (~70% in each group), but experimental group 5 only endorsed 36.6% of ungrammatical stimuli, whereas experimental group 3 endorsed 47.4%.[28] Further analysis showed that experimental group 5, like experimental group 3, tended to overendorse stand-alone *VF

[27] Interestingly, three of these subjects (5, 11 and 13) erroneously endorsed all *VF sentences in the testing phase.

[28] On grammatical stimuli, both groups performed similarly: experimental group 5 endorsed 70.8% of all grammatical stimuli and experimental group 3 70.6%.

constructions, i.e. subjects did not associate verb-final with subordinate clauses. The performance on complex sentences indicated that subjects in experiment 5, like their counterparts in experiment 3, had a clear preference for VF-V1 sentences over *VF-V2 sentences. In contrast to experimental group 3, however, experimental group 5 was significantly more likely to endorse V2-VF sentences over *V1-VF sentences. The results thus suggest that experimental group 5 knew that verbs could only occur in first position when preceded by a subordinate construction and that verbs had to be placed in second position of main clauses that were not preceded by a subordinate clause.

Did subjects learn the V1 and the V2 rules described in table 14? Their performance suggests that this is highly unlikely. As mentioned in chapter 2, the application of rule knowledge is characterized by categorical classification behaviour, e.g. by an endorsement rate of more than 90% for grammatical sequences and an endorsement rate of less than 10% for ungrammatical sequences. The analysis clearly showed that classification performance in experiment 5 was not categorical. Subjects might have acquired micro-rules or fragment knowledge that allowed them to endorse grammatical sentences and reject ungrammatical ones. In the case of the V1 rule, for example, subjects only needed to know that, in the middle of a long sentence, a verb could follow a verb. There was no need to learn that verbs had to be placed in first position in main clauses that were preceded by a subordinate clause. In the case of the V2 rule, it was sufficient to know that a verb could occur in the first part of the sentence and remember two restrictions: (i) a sentence could not begin with a verb (hence, reject *V1), and (ii) scrambled sentences do not sound like English (reject *V3). Again, there was no need to learn that verbs had to be placed in second position of a main clause. Considering the limited amount of exposure that subjects received in experiment 5, it is perhaps not surprising that there was no evidence of rule learning. The training phase in experiment 5 consisted of only 120 sentences and lasted, on average, 40 minutes. During training, subjects were exposed to 40 sentences in which a verb occurred in first position (VF-V1 sentences), to 60 sentences in which a verb occurred in second position (40 V2 sentences and 20 V2-VF sentences), and to 80 sentences in which the verb occurred in clause-final position (40 V2-VF sentences and 40 VF-V1 sentences). It is possible that a significant increase of exposure to the semi-artificial system would have resulted in more categorical classification performance.

Why did subjects fail to learn the VF rule? After all, the verbal reports showed that subjects were very likely to notice the occurrence of verbs at the end of sentences, and the rule seems surprisingly simple: In subordinate clauses, place the verb at the end. One explanation is that the sentence identification task directed participants' attention to grammatical aspects that were deemed relevant from the experimenter's perspective but were largely irrelevant from a learner's perspective. As Reber (1993) pointed out, the rules used by the experimenter to design the stimulus domain are unlikely to correspond to the rules employed by subjects to distinguish grammatical and ungrammatical sequences. Perhaps the sentence identification task simply directed attention to inappropriate cues in order for subjects to learn the VF rule.

An alternative explanation is that the German verb-final rule might simply be too difficult to acquire by native speakers of English under incidental learning conditions. The frequency and resilience of verb placement mistakes in subordinate clauses produced by English learners of German suggests that this is not an unreasonable assumption (Jansen, 2008). The syntactic rules of English could interfere negatively in the acquisition process and make it difficult to acquire the VF rule on positive evidence alone. Experiment 6 investigates whether subjects are able to acquire the rules of the semi-artificial grammar under intentional learning conditions. Subjects are asked to consciously search for word order rules in order to determine whether (i) the verb placement rules are learnable under intentional conditions and (ii) whether this results in more categorical, i.e. rule-like, classification performance.

In terms of awareness, experiment 5 provided evidence for both conscious and unconscious knowledge. In contrast to experiment 3, the confidence ratings showed that subjects were partially unaware of the knowledge they employed in making the grammaticality judgments. When subjects had no confidence in their grammaticality judgments, they actually performed significantly above chance. That is, the guessing criterion of implicit learning was met. The analysis of the accuracy estimates showed that subjects significantly underestimated their performance, which supports the view that subjects had developed unconscious judgment knowledge. The source attributions indicated that subjects were also partially unaware of the knowledge they had acquired. When attributing grammaticality judgments to intuition, subjects performed significantly above chance, i.e. they had developed unconscious structural knowledge. As in experiment 3, the analysis of confidence ratings and source attributions showed that subjects were significantly more accurate when reporting higher levels of confidence and when basing their decisions on explicit categories. Conscious knowledge was thus again associated with improved performance in the grammaticality judgment task. In terms of methodology, experiment 5 highlighted another problem in using verbal reports for assessing awareness. The analysis showed that sometimes subjects did not follow the rules they reported after the experiment, i.e. there was a discrepancy between what subjects reported and what they actually did. This further supports the view that post-experimental verbal reports are insufficient to assess whether learning occurred with or without awareness.

7. Experiment 6

7.1. Introduction

Experiment 6 investigates the acquisition of verb placement rules of a semi-artificial language under intentional learning conditions. The stimulus material and the testing procedure are the same as the ones used in experiments 3 and 5. In contrast to these experiments, however, subjects in experiment 6 were instructed to find the rules that governed the word order of the semi-artificial language. The aim was to investigate whether the verb placement rules are learnable under intentional conditions and to determine how explicit learning affects classification performance.

7.2. Method

7.2.1. Participants

Fifteen native speakers of English (11 women and 4 men, mean age = 23.1 years) were recruited to take part in this experiment. All participants were students at the University of Cambridge at the time of the experiment. Participants spoke, on average, 2.5 foreign languages and none had a background in German or any other V2-language. The data of the 15 untrained subjects from experiment 3 served as control data. Experimental and control groups did not differ significantly across age, $t(27) = 1.109 > .05$, gender, $\chi^2(1, N = 30) = 0.186, p > .05$, occupation (student vs. nonstudent), $\chi^2(1, N = 30) = 2.143, p > .05$, and number of languages acquired, $\chi^2(4, N = 30) = 3.565, p > .05$. All subjects received £5 for participating in the experiment.

7.2.2. Stimulus material

The 180 stimulus sentences created for experiment 3 were used in this experiment. The linguistic focus was on the three verb placement rules described in table 14.

7.2.3. Procedure

The experiment consisted of (i) a training phase, (ii) a testing phase, and (iii) a debriefing session. Participants were initially exposed to the semi-artificial grammar under intentional learning conditions by means of a rule-search task. In the testing phase, participants were asked to classify novel instances of the semi-artificial grammar and to report confidence levels and the basis for each classification decision. The experiment concluded with a debriefing session. All tasks were run on a Dell PC (Windows XP) and delivered via Cedrus SuperLab Pro (version 4.0). The instructions are reproduced in the appendix.

7.2.3.1. Training phase

Experimental participants were trained on the semi-artificial grammar by means of an auditory rule-search task. At the beginning, participants were told they were going to listen to 120 scrambled sentences and instructed to discover the rules that governed the placement of words. Participants were also told that they would subsequently be tested on a grammaticality judgment task. The experiment began with a short practice session to familiarize subjects with the training task. This consisted of four practice trials which were not repeated in the actual training set. Participants were then exposed to the stimulus sentences on an item-by-item basis and instructed to find the word-order rules. The screen was blank while the stimuli were being played. A new stimulus sentence was played each time participants pressed the space bar. Plausibility judgments were not used in this experiment. Participants were not allowed to take notes. The completion of the training phase took approximately 40 minutes.

7.2.3.2. Testing phase

The testing procedure was identical to the one used in experiments 3 and 5, with the exception that participants in experiment 6 knew in advance they were going to be tested.

7.2.3.3. Debriefing

At the end of the experiment, all participants were given the same debriefing questionnaire as participants in the previous experiments (see appendix).

7.3. Results

7.3.1. Testing phase

7.3.1.1. Grammaticality judgments

Overall performance

The analysis of the grammaticality judgments showed that the experimental group classified 76.6% (*SD* = 13.7%) of the test items correctly. Overall performance was significantly above chance, $t(14) = 7.515$, $p < .0005$. That is, the training phase produced a clear learning effect in experimental group 6. A one-way ANOVA further showed that there were significant differences among experimental groups 6, 5, 4, 3 and the control group, $F(4, 69) = 32.52$, $p < .0005$. The assumption of homogeneity of variance was violated, but the Brown-Forsythe *F*-ratio was nonetheless significant, $F(4, 40) = 32.52$, $p < .0005$. Results of the Games-Howell post-hoc analysis indicate that experimental group 6 only significantly outperformed the control subjects, $p < .0005$, but not the other experimental groups in terms of overall accuracy.

Table 34
Classification Performance (%) of Experimental and Control Subjects

Group		Overall accuracy
Exp 6	*M*	76.6*
	SD	13.7
	SE	3.5
Exp 5	*M*	67.2*
	SD	6.4
	SE	1.7
Exp 4	*M*	65.5*
	SD	5.2
	SE	1.4
Exp 3	*M*	61.6*
	SD	8.3
	SE	2.2
Control	*M*	42.9*
	SD	5.2
	SE	1.3

Significance from chance: * $p < .0005$.

In order to establish whether any improvements in accuracy occurred during the testing phase, the grammaticality judgment data was equally divided into 10 chronological stages. The mean accuracy for each block of six judgments was then computed. A repeated measures ANOVA showed no significant differences across the stages in the experimental group, $F(9, 126) = 1.084$, $p > .05$. That is, no online learning effect was observed during the testing phase. Figure 34 displays the mean accuracy across the 10 stages.

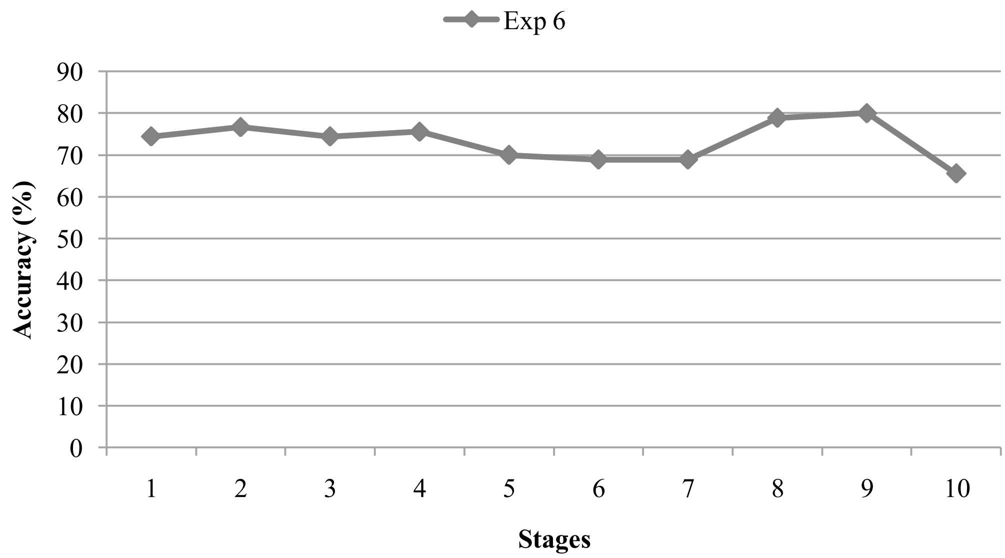

Figure 34. Classification accuracy of experimental group 6 across 10 chronological stages of the testing set.

Classification performance on grammatical and ungrammatical items

The following analyses report endorsement rates. The results showed that experimental group 6 endorsed 83.3% of all grammatical sentences and 30% of all ungrammatical ones, $t(14) = 7.531$, $p < .0005$. Performance on grammatical items was significantly above chance, $t(14) = 8.646$, $p < .0005$, and performance on ungrammatical items was significantly below chance, $t(14) = -3.967$, $p < .05$. A one-way ANOVA further showed that there were significant differences among experimental groups 6, 5, 4, 3 and the controls on grammatical items, $F(4, 69) = 14.128$, $p < .0005$. The assumption of homogeneity of variance was violated. However, the Brown-Forsythe F-ratio was nonetheless significant, $F(4, 37.46) = 14.238$, $p < .0005$. Results of the Games-Howell post-hoc analysis indicate that experimental group 6 was significantly more likely to endorse grammatical items than the control group, $p < .0005$, but not more likely than the other experimental groups. With regard to performance on ungrammatical items, a one-way ANOVA confirmed that there were significant differences among the experimental and the control groups, $F(4, 69) = 2.771$, $p < .05$. The assumption of homogeneity of variance was violated, but the Brown-Forsythe F-ratio was again significant, $F(4, 48.63) = 2.769$, $p < .05$. Results of the Games-Howell post-hoc analysis suggest that experimental group 6 was significantly less likely to endorse ungrammatical items than experimental group 3, $p < .05$, and the control group, $p < .05$, but not less likely than the other experimental groups. Figure 35 compares the performance of experimental and control subjects across grammaticality. The scores are summarized in table 35.

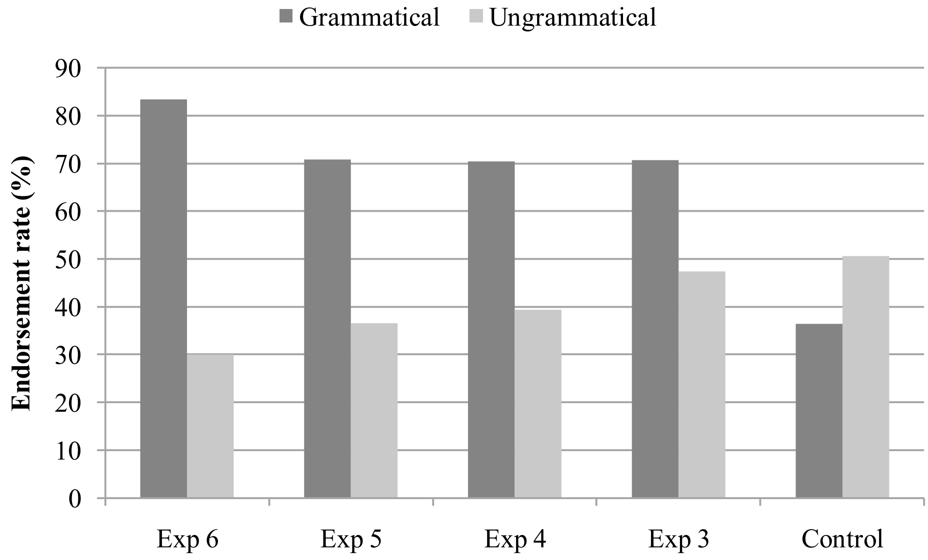

Figure 35. Classification performance of experimental and control subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

Table 35
Classification Performance (%) of Experimental and Control Subjects on Grammatical and Ungrammatical Stimuli

Group		Grammatical	Ungrammatical
Exp 6	*M*	83.3**	30**
	SD	14.8	19.4
	SE	3.8	5
Exp 5	*M*	70.8**	36.6**
	SD	14.4	13.2
	SE	3.7	3.4
Exp 4	*M*	70.4**	39.3**
	SD	9.4	10.5
	SE	2.4	2.7
Exp 3	*M*	70.6**	47.4
	SD	13.5	19.7
	SE	3.5	5.1
Control	*M*	36.4	50.6
	SD	30.3	28.4
	SE	7.8	7.3

Significance from chance: * $p < .05$. ** $p < .001$.

Interestingly, four subjects of experimental group 6 displayed categorical classification performance, endorsing 96.7% of grammatical sentences and only 5.8% of ungrammatical ones. The remaining 11 subjects endorsed 78.5% of grammatical items, which was lower than the endorsement rate of categorical subjects and 38.8% of ungrammatical items, which was higher. The performance of categorical and noncategorical subjects in experiment 6 will be compared in section 7.4. below.

Classification performance across sentence types

Figure 36 illustrates participants' endorsement rates across the nine syntactic patterns employed in this experiment. The scores are summarized in table 36. The analysis indicated that experimental group 6 performed significantly above chance on all grammatical patterns, i.e. V2 sentences, $t(14) = 15.029$, $p < .0005$, V2-VF sentences, $t(14) = 4.154$, $p < .001$, and VF-V1 sentences, $t(14) = 3.963$, $p < .001$. Experimental group 6 performed significantly below chance on *V3 sentences, $t(14) = -1.781$, $p < .0005$, and *VF-V2 sentences, $t(14) = -4.108$, $p < .001$. Performance on *V1 sentences approached significance from chance, $t(14) = -1.781$, $p = .097$. In comparison to the other groups, a one-way ANOVA showed significant effects between groups for the three grammatical sentence types – V2 items, $F(4, 69) = 6.884$, $p <. 0005$, V2-VF items, $F(4, 69) = 7.546$, $p < .0005$, VF-V1 items, $F(4, 69) = 10.971$, $p < .0005$ – and for several ungrammatical sentence types, namely *V3 items, $F(4, 69) = 25.912$, $p < .0005$, *VF items, $F(4, 69) = 2.851$, $p < .05$, *VF-V2 items, $F(4, 69) = 4.054$, $p < .01$, VF-V1 items, $F(4, 69) = 10.971$, $p < .0005$, and *V1-VF items, $F(4, 69) = 2.810$, $p < .05$. There were no effects on *V1 and V4 sentences. The assumption of homogeneity of variance was violated for V2, *V3, *VF, *VF-V2 and VF-V1 sentences. However, the Brown-Forsythe *F*-ratios were nonetheless significant on all constructions: V2, $F(4, 48.48) = 6.872$, $p < .0005$, *V3, $F(4, 64.16) = 1.06$, $p < .0005$, *VF, $F(4, 60.43) = 2.88$, $p < .05$, *VF, $F(4, 60.43) = 2.88$, $p < .05$, *VF-V2, $F(4, 58.39) = 4.046$, $p < .01$, and VF-V1, $F(4, 51.07) = 10.918$, $p < .0005$.

Results of the post-hoc analyses, using Tukey's HSD test for V2-VF and *V1-VF sentences and the Games-Howell procedure for the remaining items, showed that experimental group 6 was significantly more likely than the controls to endorse V2 sentences, $p < .005$, V2-VF sentences, $p < .0005$, VF-V1 sentences, $p < .0005$, and significantly less likely to endorse *V3 sentences, $p < .0005$. There were no significant differences between experimental groups 6 and 5 across patterns, but experimental group 6 was significantly more likely than experimental group 4 to endorse V2 sentences, $p < .0005$, and significantly less likely to endorse *V3 sentences, $p < .05$, and *VF-V2 sentences, $p < .01$. Finally, experimental group 6 was significantly less likely than experimental group 3 to endorse *VF sentences, $p < .05$.

The performance on one-clause constructions indicated that experimental group 6 preferred V2 sentences over *V1 sentences, $t(14) = -6.632$, $p < .0005$, *V3 sentences, $t(14) = 11.805$, $p < .0005$, *V4 sentences, $t(14) = 5.201$, $p < .0005$, and *VF sentences, $t(14) = 5.081$, $p < .0005$.[29] That is, when presented with a simple sentence, experimental group 6 endorsed the only verb placement option licensed by the grammar. The performance on complex sentences showed that experimental group 6 preferred V2-VF sentences over *V1-VF sentences, $t(14) = 3.212$, $p < .05$, and VF-V1 sentences over *VF-V2 sentences, $t(14) = 7.333$, $p < .0005$. Subjects had a clear preference for grammatical (i.e., trained) structures over ungrammatical ones.[30]

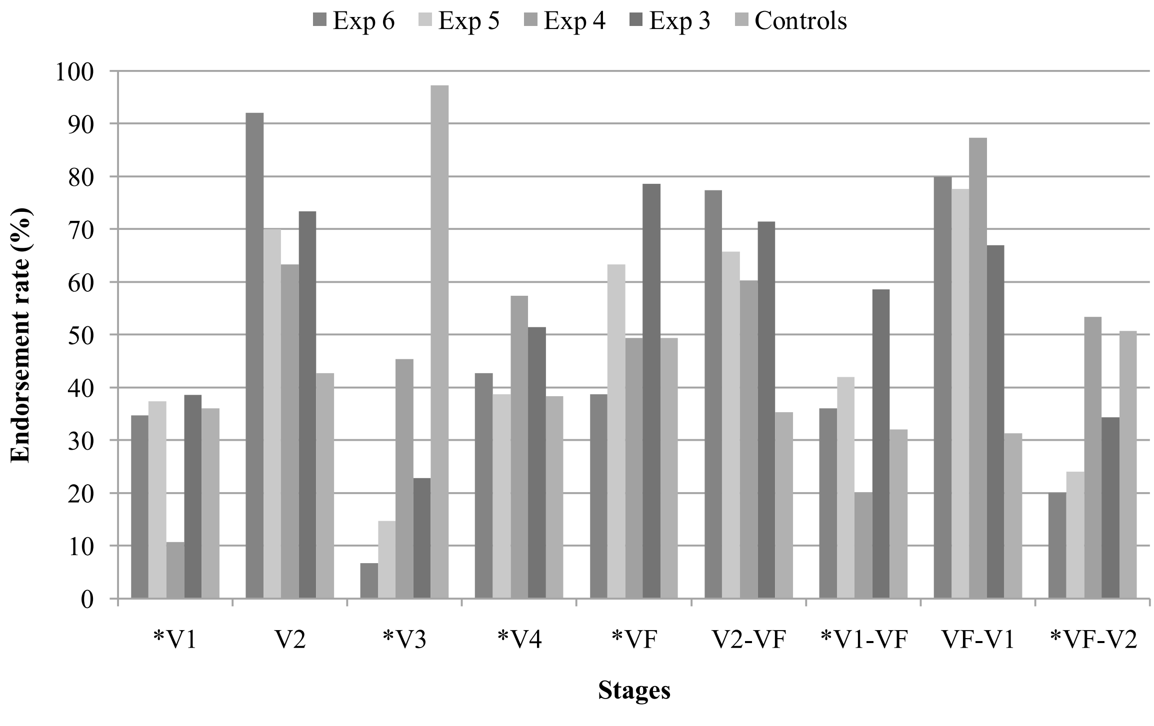

Figure 36. Classification performance of experimental and control subjects across sentence types. Bars represent endorsement rates.

[29] In the control group, there were no differences between performance on V2 sentences and *V1 sentences, $t(14) = -1.299$, $p > .05$, V2 and *V4 sentences, $t(14) = .540$, $p > .05$, and V2 and *VF sentences, $t(14) = -.856$, $p < .05$, but controls were more likely to endorse *V3 over V2 sentences, $t(14) = -5.942$, $p < .0005$.

[30] In the control group, there was no difference between V2-VF sentences and *V1-VF sentences, $t(14) = .617$, $p > .05$. Controls preferred the ungrammatical *VF-V2 sentences over VF-V1 sentences, $t(14) = -2.327$, $p < .05$.

Table 36
Classification Performance (%) of Experimental and Control Subjects across Sentence Types

Group		Sentence type								
		*V1	V2	*V3	*V4	*VF	V2-VF	*V1-VF	VF-V1	*VF-V2
Exp 6	M	34.7	92**	6.7**	42.7	38.7	77.3**	36	79.9*	20*
	SD	33.4	10.8	20.9	32.8	36.6	25.5	37.2	29.3	28.3
	SE	8.6	2.8	5.4	8.5	9.5	6.6	9.6	7.6	7.3
Exp 5	M	37.3	70*	14.7**	38.7	63.3	65.7*	42	77.6**	24*
	SD	36.9	31.2	29.7	27.7	38.9	19.6	28.8	15.4	24.1
	SE	9.5	8	7.7	7.2	10	5.1	7.4	4	6.2
Exp 4	M	10.7**	63.3*	45.3	57.3	49.3	60.2^{+}	20**	87.3**	53.3
	SD	14.9	16.3	35	24.9	28.1	19.3	21.4	13.9	20.9
	SE	3.8	4.2	9	6.4	7.3	5	5.5	3.6	5.4
Exp 3	M	38.6	73.4*	22.9*	51.4	78.6**	71.4**	58.6	66.9*	34.3
	SD	36.3	27.6	35.8	33	22.8	18.8	35.5	29.2	30.8
	SE	9.4	7.1	9.3	8.5	5.9	4.8	9.2	7.5	8
Controls	M	36	42.7	97.3**	38.3	49.3	35.3	32	31.3^{+}	50.7
	SD	34	36.7	7	37.0	42.7	28.8	35.3	34.8	38.4
	SE	8.8	9.5	1.8	9.6	11	7.4	9.1	9	9.9

Significance from chance: * $p < .05$. ** $p < .001$.

7.3.1.2. Confidence ratings

Overall performance

The average confidence level in the experimental group was 6.2 (*SD* = 1.1). In terms of proportion, experimental participants tended to select the option somewhat confident most frequently and the guess option least frequently. In terms of accuracy, the analysis indicated that experimental participants were most accurate when reporting to be very confident in their decisions. Accuracy was lower for those grammaticality decisions in which subjects were guessing and lowest for somewhat confident decisions. Experimental participants scored significantly above chance on all three confidence categories, including guesses. According to the guessing criterion there was unconscious judgment knowledge in experimental group 6. Table 37 summarizes the scores.

Table 37
Accuracy and Proportions (%) across Confidence Ratings

	Guess	Somewhat confident	Very confident
Accuracy	80*	74*	81*
Proportion	6	56	38

Significance from chance: * $p < .01$.

In order to determine whether participants became more confident while performing on the testing phase, the confidence ratings were divided into 10 chronological stages and the average confidence levels calculated for each stage. A repeated measures ANOVA showed no significant differences in mean confidence across the stages, $F(9, 126) = .755$, $p > .05$. That is, the experimental group did not display significant changes in confidence levels while performing on the testing phase. In addition, the proportion of guess, somewhat confident and very confident responses were calculated for each of the 10 blocks. As figure 37 suggests, there were relatively few changes in the response proportions across the testing phase. The analysis confirms that participants did not become more confident while performing on the grammaticality judgment task.

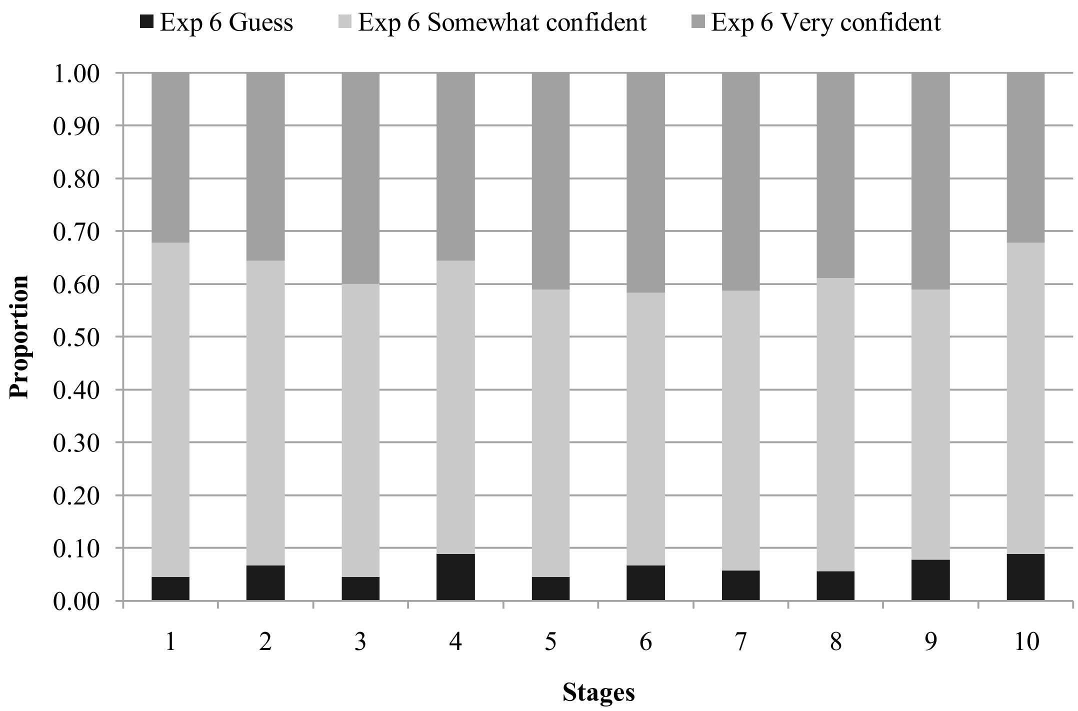

Figure 37. Proportions of guess, somewhat confident and very confident responses across 10 stages of the testing set.

Zero correlation criterion

The Chan difference score was computed in order to establish whether learning in the experimental group was implicit by the zero correlation criterion. The average confidence for correct grammaticality decisions was 6.3 (*SD* = 1.2) and the average confidence for incorrect decisions was 5.8 (*SD* = 0.8), i.e. experimental participants were more confident in correct decisions than in incorrect ones. The difference (0.4) approached significance, $t(14) = 1.855$, $p = .085$.

7.3.1.3. Source attributions

Overall performance

In terms of proportion, experimental participants most frequently believed their classification decisions to be based on rule knowledge, followed by intuition and memory. The guess category was selected least frequently. In terms of accuracy, experimental participants scored highest when reporting to use rule knowledge to guide their decisions, followed by the intuition and guess categories. Subjects were least accurate when basing decisions on memory. Participants performed significantly above chance when basing their decisions on intuition, memory or on rule knowledge. Performance on decisions based on guesses approaches significance, $p = .06$.

Table 38
Accuracy and Proportions (%) across Source Attributions

	Guess	Intuition	Memory	Rule
Accuracy	65.9^{+}	68.9**	61.9*	83.6**
Proportion	4.6	21.8	13.1	60.6

Significance from chance: $^{+}$ $p = .06$. * $p < .05$. ** $p < .001$.

Relationship between confidence ratings and source attributions

Participants tended to report significantly higher confidence levels for attributions associated with conscious knowledge (rules and memory, combined $M = 6.56$) rather than unconscious knowledge (guess and intuition, combined $M = 3.38$), $p < .05$.

Table 39 summarizes the mean accuracy score for pooled confidence ratings and source attributions. In terms of proportion, the analysis indicates that the majority of grammaticality judgments were either somewhat confident and based on intuition or rule knowledge, or very confident and based on rule knowledge. In terms of accuracy, participants performed significantly above chance on classification decisions made without confidence and based on guesses or intuition. Participants also performed significantly above chance on somewhat confident grammaticality decisions based on intuition or rule knowledge and on very confident grammaticality decisions based on memory or rule knowledge.

Table 39
Accuracy and Proportions (%) for Combined Confidence Ratings and Source Attributions

		Source attributions			
Confidence ratings		Guess	Intuition	Memory	Rule
Guess	Accuracy	73.1*	90.9**		75
	Proportion	2.9	2.4	0	0.9
Somewhat confident	Accuracy	50	66.2**	55.8	85.5**
	Proportion	1.6	16.8	9.6	27.7
Very confident	Accuracy	100	65.2	78.1*	82.2**
	Proportion	0.1	2.6	3.6	31.9

Significance from chance: * $p < .05$. ** $p < .001$.

7.3.2. Debriefing

7.3.2.1. Accuracy estimates

Participants' assessment of their performance on the grammaticality judgment task was compared to their actual performance. The comparison showed that the experimental group was not very accurate in estimating its performance. The difference between estimated accuracy (M = 64.3%, SD = 12.4%) and observed accuracy (M = 76.6%, SD = 13.7%) was 12.4. This difference was significant, $t(14) = 2.605$, $p < .05$, i.e. participants in this experiment significantly underestimated their performance. The analysis of the accuracy estimates thus supports the view that subjects were partially unaware of the knowledge that guided their grammaticality decisions.

7.3.2.2. Verbal reports

A transcription of the verbal reports can be found in the appendix. The analysis showed that all subjects reported knowledge when prompted, at the end of the experiment, to describe any rules or regularities they might have noticed. Since subjects in experiment 6 were explicitly instructed to discover rules, this result is not particularly surprising. As in the previous experiment, closer analysis of the reports showed that subjects did not employ phrasal categories such as NP, PP or VP in order to describe what they notice but relied on functional categories and semantic categories instead. Again, the most frequently remarked regularity concerned the order of NP_{subj} and VP, with subjects commenting on the "reversed" order of subject and verb (see experimental subjects 3, 6, 8, 9 and 10). Several participants also commented on the fact the verb was often placed at the end of a sentence (e.g., subjects 2, 3, 9 and 10). The descriptions of subjects 2, 4 and 10 suggested that they were aware of V1 rule. Still, despite listening to 120 sentences with the instruction to discover the rules that govern the placement of words, only two subjects (4 and 10) were able to do so for all three rules.[31]

7.4. Comparison of categorical and noncategorical subjects

As mentioned above, the analysis showed that four subjects performed categorically in their grammaticality decisions, i.e. they endorsed more than 90% of grammatical sentences and less than 10% of ungrammatical sentences. The following subsection will compare the performance of these subgroups on grammaticality judgments, confidence ratings and source attributions.

7.4.1. Grammaticality judgments

Overall performance

The reanalysis of the grammaticality judgments showed that the categorical subjects classified 95.4% (*SD* = 2.5%) of all sentences correctly and the noncategorical subjects 69.8% (*SD* = 8.3%). The performance of categorical subjects was higher than that of experimental groups 3, 4 5. Noncategorical subjects outperformed experimental group 3 but not experimental groups 4 and 5.

Classification performance

The performance across grammaticality showed that categorical subjects endorsed 96.6% (*SD* = 2.8%) of grammatical sentences and 5.8% (*SD* = 5%) of ungrammatical ones, whereas noncategorical subjects endorsed 78.5% (*SD* = 14.4%) of grammatical sentences and 38.8% (*SD* = 14.1%) of ungrammatical ones. The analysis also showed that categorical subjects were more likely to endorse grammatical sentences and less likely to endorse ungrammatical sentences than subjects in experimental group 3, 4 and 5. Noncategorical subjects were more

[31] Neither subject had a background in linguistics. Both belong to the "categorical" subgroup.

likely to endorse grammatical items than subjects in experimental groups 3, 4 5, but there was little difference on ungrammatical items.

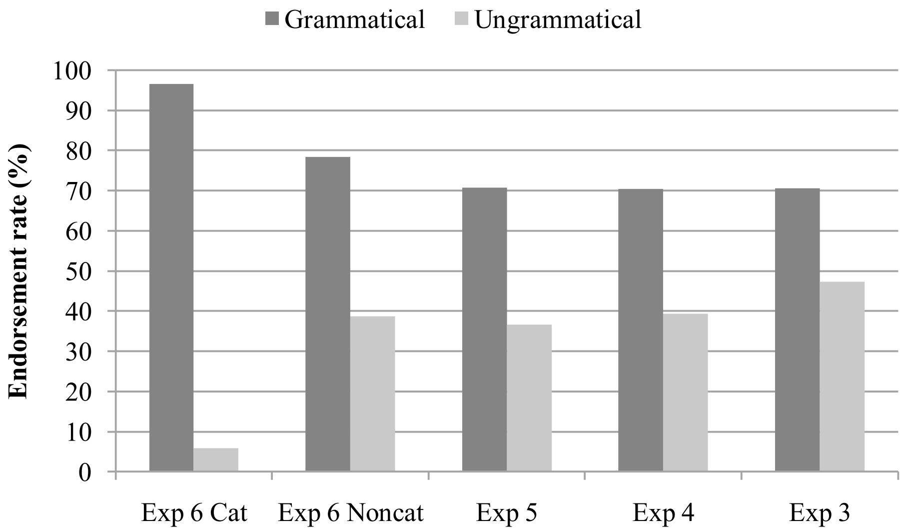

Figure 38. Classification performance of experimental subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

Classification performance across sentence types

Figure 39 compares the endorsement rates of categorical and noncategorical subjects across the syntactic patterns employed. Table 40 summarizes the scores. The analysis indicated that categorical subjects were more likely than noncategorical subjects to endorse V2-VF and VF-V1 sentences and less likely to endorse *V1, *V4, *VF, *V1-VF and *VF-V2 sentences.

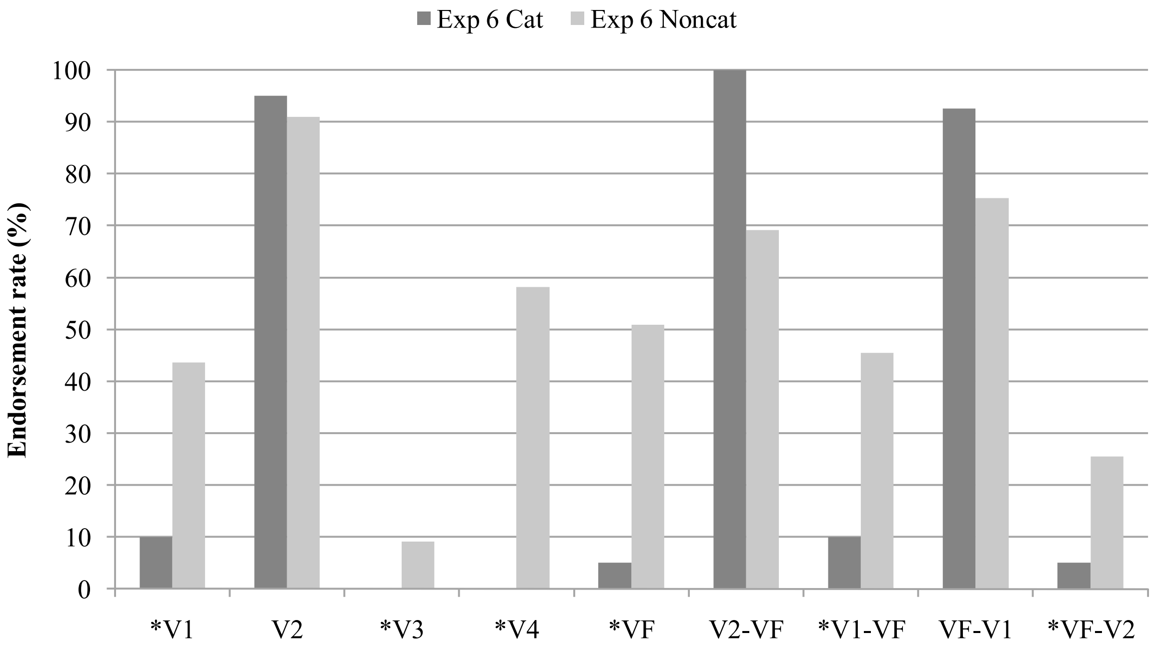

Figure 39. Classification performance of experimental subjects across sentence types.
Bars represent endorsement rates.

In comparison to the other experimental groups, the analysis showed that categorical subjects were more likely than experimental group 3 to endorse V2-VF sentences and less likely to endorse *V4, *VF, *V1-VF and *VF-V2 sentences. Categorical subjects were more likely than experimental group 4 to endorse V2 and V2-VF sentences and less likely to endorse *V3, *V4, *VF and *VF-V2 sentences. Categorical subjects were more likely than experimental group 5 to endorse V2-VF and VF-V1 sentences and less likely to endorse *V4, *VF and *V1-VF sentences.

The noncategorical subjects were more likely than experimental group 3 to endorse V2 sentences and less likely to endorse *VF sentences. Noncategorical subjects were also more likely than experimental group 4 to endorse V2 sentences and less likely to endorse *V3 and *VF-V2 sentences. They were more likely to erroneously endorse *V1 sentences and *V1-VF sentences. Finally, noncategorical subjects were more likely than experimental group 5 to endorse V2 sentences but also *V4 sentences.

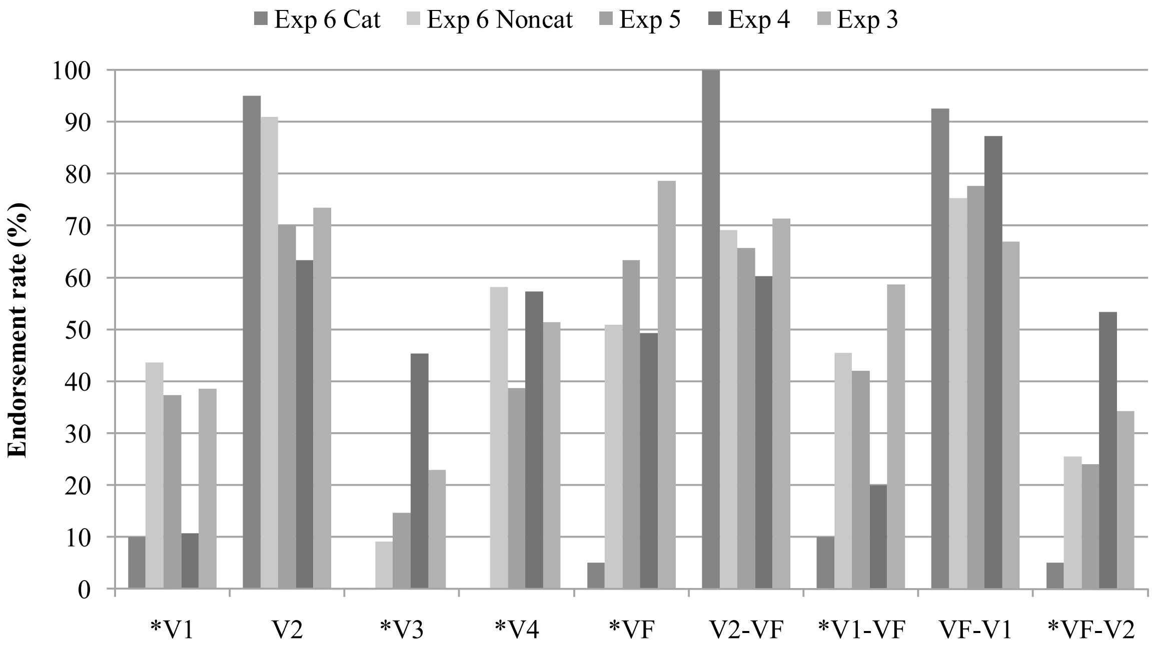

Figure 40. Classification performance of experimental subjects across sentence types. Bars represent endorsement rates.

Table 40
Classification Performance (%) of Experimental Subjects across Sentence Types

Group		Sentence type								
		*V1	V2	*V3	*V4	*VF	V2-VF	*V1-VF	VF-V1	*VF-V2
Exp 6 Cat	*M*	10*	95*	0	0	5*	100	10*	92.5*	5*
	SD	11.5	10	0	0	10	0	11.5	9.6	10
	SE	5.8	5	0	0	5	0	5.8	4.8	5
Exp 6 Noncat	*M*	43.6	90.9**	9.1**	58.2	50.9	69.1*	45.5	75.4*	25.5*
	SD	34.4	11.4	24.3	22.7	35.1	25.1	39.1	32.9	31.1
	SE	10.4	3.4	7.3	6.9	10.6	7.6	11.8	9.9	9.4
Exp 5	*M*	37.3	70*	14.7**	38.7	63.3	65.7*	42	77.6**	24*
	SD	36.9	31.2	29.7	27.7	38.9	19.6	28.8	15.4	24.1
	SE	9.5	8	7.7	7.2	10	5.1	7.4	4	6.2
Exp 4	*M*	10.7**	63.3*	45.3	57.3	49.3	60.2^{+}	20**	87.3**	53.3
	SD	14.9	16.3	35.0	24.9	28.1	19.3	21.4	13.9	20.9
	SE	3.8	4.2	9	6.4	7.3	5	5.5	3.6	5.4
Exp 3	*M*	38.6	73.4*	22.9*	51.4	78.6**	71.4**	58.6	66.9*	34.3
	SD	36.3	27.6	35.8	33.0	22.8	18.8	35.5	29.2	30.8
	SE	9.4	7.1	9.3	8.5	5.9	4.8	9.2	7.5	8

Significance from chance: * $p < .05$. ** $p < .001$.

7.4.2. Confidence ratings

In terms of proportion, both groups selected the „somewhat confident' option most frequently, followed by the options „very confident' and „guess.' In terms of accuracy, the analysis indicated that categorical subjects were most accurate when reporting to be very confident in their decisions, followed by somewhat confident decisions. Accuracy was lowest for guess decisions. In all three cases, subjects performed above 90% and significantly above chance. Noncategorical subjects were most accurate when reporting to be guessing, followed by very confident decisions. Accuracy was lowest for somewhat confident decisions. Performance was above chance across the three response options. Table 41 summarizes the scores.

Table 41
Accuracy and Proportions (%) across Confidence Ratings

Group		Guess	Somewhat confident	Very confident
Categorical	Accuracy	92.9*	95.2*	96.3*
	Proportion	5.8	60.8	33.3
Noncategorical	Accuracy	76.2*	64.6*	76*
	Proportion	6.4	53.7	39.9

Significance from chance: * $p < .01$.

Zero correlation criterion

Both groups were more confident in correct decisions than in incorrect ones. In the categorical subjects, the average confidence for correct grammaticality decisions was 6.1 (*SD* = 1.2) and the average confidence for incorrect decisions was 5.4 (*SD* = 0.5). In the noncategorical subjects, the average confidence for correct grammaticality decisions was 6.4 (*SD* = 1.3) and the average confidence for incorrect decisions was 6 (*SD* = 0.9). The difference was not significant for categorical, $t(3) = -1.197$, $p > .05$, or noncategorical subjects, $t(10) = -1.369$, $p > .05$.

7.4.3. Source attributions

In terms of proportion, both groups attributed their decisions most frequently to rule knowledge, followed by memory and intuition. The guess category was selected least frequently. In terms of accuracy, categorical subjects scored highest when reporting to base decisions on guessing, followed by the rule knowledge and memory. Categorical subjects were least accurate when basing decisions on intuition. Performance was above chance on grammaticality judgments based on intuition and on rule knowledge. Noncategorical subjects scored highest when reporting to base decisions on rule knowledge, followed by memory and guesses. Noncategorical subjects were least accurate when basing decisions on intuition. Performance was significantly above chance on grammaticality judgments based on intuition, memory and rule knowledge.

Table 42
Accuracy and Proportions (%) across Source Attributions

Group		Guess	Intuition	Memory	Rule
Categorical	Accuracy	100	80**	88.9	96.6**
	Proportion	0.8	2.1	11.3	85.8
Noncategorical	Accuracy	64.1	61.1**	65.7*	75.7**
	Proportion	5.9	17.1	25.6	51.4

Significance from chance: * < p .05. ** p < .01.

7.4.4. Accuracy estimates

The comparison between estimated and observed accuracy showed that categorical and noncategorical subjects underestimated their performance. However, this difference was only significant in the case of categorical subjects. Categorical subjects estimated to be 66.3% (SD = 13.8%) accurate but actually performed at 95.4% (SD = 2.5%) accuracy, $t(3) = -3.672$, $p < .05$). Noncategorical subjects believed to be 63.5% accurate (SD = 12.5%) and were actually 69.7% (SD = 8.1%) accurate, $t(10) = -1.355$, $p > .05$. The reanalysis of the accuracy suggests that the categorical subjects were partially unaware of the knowledge that guided their grammaticality decisions.

7.5. Discussion

Experiment 6 investigated whether the semi-artificial grammar could be learned by means of a rule-search task. The results showed that asking subjects to consciously search for word order rules produced a greater learning effect in experimental group 6 (76.6%, compared to 67.2% in experiment 5 and 61.6% in experiment 3). The analysis also indicated that experimental group 6 was more likely to endorse grammatical patterns than the previous experimental groups and less likely to endorse ungrammatical patterns. Further analysis showed that experimental group 6, in contrast to experimental groups 3 and 5, did not overendorse stand-alone *VF constructions. When presented with simple sentences, subjects in experiment 6 tended to accept V2 sentences and reject *V1, *V3, *V4 and *VF sentences. The performance on complex sentences indicated that experimental group 6 had a clear preference for VF-V1 over *VF-V2 sentences and for V2-VF over *V1-VF sentences. The results thus showed that experimental group 6 was more likely to endorse grammatical patterns and reject ungrammatical ones than other groups.

In terms of awareness, experiment 6 provided evidence for both conscious and unconscious knowledge. The confidence ratings provided no evidence for unconscious judgment knowledge since the guessing criterion for implicit learning was not met. When subjects reported no confidence in their grammaticality judgments, they performed at chance level. Interestingly, however, the analysis of the accuracy estimates showed that subjects significantly underestimated their performance, which supports the view that subjects had developed at least some unconscious judgment knowledge. The source attributions indicated that subjects were also partially unaware of the knowledge they had acquired. When attributing grammaticality judgments to intuition, subjects performed significantly above chance, i.e. they had developed unconscious structural knowledge. The analysis of confidence ratings showed no significant correlation between confidence and accuracy. Subjects were similarly accurate when making guess decisions and very confident decisions (80% and 81% accuracy, respectively). The fact that experimental group 6 significantly underestimated their performance in terms of accuracy confirms that learning in this experiment proceeded in part without awareness.

Closer analysis of the grammaticality judgments indicated that experimental group 6 could be dissociated on the basis of their classification performance into categorical and noncategorical subjects. The reanalysis of the judgment data showed that only categorical subjects were performing significantly different from the remaining experimental groups. Noncategorical subjects outperformed experimental group 3 in terms of overall accuracy, but there were no differences between noncategorical subjects and experimental groups 3, 4 and 5 across grammaticality. Noncategorical subjects performed similarly to experimental groups 3 and 5 across syntactic patterns, including *VF sentences, which they tended to overendorse. Only categorical subjects preferred all three grammatical patterns (V2, V2-VF, VF-V1) over ungrammatical ones (*V1, *V3, *V4, *VF, *V1-VF, *VF-V2). The performance of the categorical subjects suggests that the semi-artificial grammar was learnable under intentional learning conditions, without the benefit of feedback and after a brief exposure period. In noncategorical subjects, learning appears to have been limited. However, the reanalysis of the confidence ratings and source attributions showed that the division of experimental group 6 into categorical and noncategorical did not affect the conclusion that learning in this experiment was at least partially unconscious. Both groups performed significantly above chance when guessing and when basing decisions on intuition or on rule knowledge. The reanalysis of the accuracy estimates indicated that only categorical subjects significantly underestimated their performance, namely by 29.2%. The reanalysis of the verbal reports showed that the two subjects that were able to provide examples of all three rules belonged to the categorical group.

Why did only four subjects manage to acquire the verb placement rules of the semi-artificial grammar under intentional learning conditions? After all, categorical and noncategorical subjects were given the same task instructions and received the same amounts of exposure. One explanation is the wording of the task instructions. As mentioned above, subjects were told to pay attention to the sentences and to find the rules that determined word order. It is likely that more subjects would have succeeded if the instructions had prompted them to search specifically for verb placement rules, rather than word order rules in general. A different possibility is that categorical and noncategorical subjects differed across variables that have been shown to affect explicit learning processes. These variables include, for example, age (Reber, 1993), working memory capacity (Frackowiak et al., 2004) and intelligence quotient (Reber, Walkenfeld, & Hernstadt, 1991). The analysis showed that categorical and noncategorical subjects did not differ across the variables age, gender, occupation or number of languages acquired.[32] Subjects' working memory capacity and IQ were not assessed in this experiment, which makes it impossible to determine whether categorical subjects outperformed noncategorical subjects on these factors. A future experiment would benefit from the addition of a standard IQ measure, e.g. the Wechsler Adult Intelligence Scale-Revised (WAIS-R), and a working memory measure, e.g. Majerus et al.'s (2006) serial order reconstruction task, to the experimental procedure.

[32] The small sample size of categorical subjects constitutes a problem here.

8. Summary and Conclusion

The present dissertation reported the results of six experiments on the implicit learning of natural language syntax. The experiments followed the conceptual design of AGL research but employed stimuli that resemble natural languages more closely. The stimulus sentences featured word order patterns of a natural language and contained semantic information in most experiments. In contrast to studies conducted in the statistical learning tradition, these experiments added measures of awareness to the experimental design in order to determine whether subjects became conscious of the knowledge they had acquired. Unlike SLA studies on implicit and explicit learning, the experiments did not simply rely on verbal reports to measure awareness but on a variety of tests, including confidence ratings and source attributions. The aims were (i) to investigate how adult learners acquire syntactic knowledge of a new linguistic system without intending to, (ii) to determine whether adult learners can acquire non-native syntax without becoming of the aware of the knowledge they have acquired, and (iii) to establish what role awareness plays in the acquisition process.

The results of the experiments indicate that adult learners are able to acquire syntactic structures of a novel language under both incidental and intentional learning conditions, while processing sentences for meaning, without the benefit of corrective feedback and after short exposure periods. That is, the findings demonstrate that the implicit learning of natural language is not restricted to infants and child learners. In addition, the experiments also show that subjects are able to transfer their knowledge to stimuli with the same underlying structure but new surface features. The measures of awareness further suggest that, in experiments 3 to 6 at least, learning resulted in both conscious and unconscious knowledge. While subjects did not become aware of all the information they have acquired, it was clear that higher leves of awareness were associated with improved performance.

In experiment 1, subjects were initially exposed to a semi-artificial grammar under incidental learning conditions by means of a plausibility judgment task. The training task required subjects to listen to each stimulus sentence and to indicate, on an item-by-item basis, whether the statements made were semantically plausible. In order to complete the task, subjects had to focus primarily on the linguistic meaning of the stimuli. After the exposure phase, learning was assessed by means of grammaticality judgments and awareness by means of binary confidence ratings. In the debriefing session, subjects also judged their overall accuracy and completed a verbal report questionnaire. The results showed that learning was restricted to those subjects that were aware of the knowledge that guided their performance on the classification task. The verbal reports indicated that subjects were not able to explicitly state the rules that governed verb placement. However, subjects' knowledge was conscious according to the zero correlation criterion, as assessed by Kunimoto et al.'s (2001) binary confidence technique. In addition, the results suggested that classification performance was largely driven by the above chance endorsement of grammatical sentences. Experimental subjects were significantly more likely to endorse previously encountered patterns than untrained controls. However, when presented with ungrammatical items, subjects were reduced to guessing.

Experiment 2 investigated whether increasing the variety of training patterns would produce a greater learning effect. The training and testing procedures in experiment 2 were the same as those in experiment 1. The training set in experiment 2 contained the same sentence types (V2, V2VF, V2-VF and VF-V1), but there were more patterns associated with each sentence type. The same testing set as in experiment 1 was used. The results showed that presenting subjects with more varied sentences did not increase learning. In addition, the results showed that learning was again restricted to those subjects that were aware of the knowledge they had acquired. Performance was again driven by the above chance endorsement of grammatical items. Subjects' performance on ungrammatical stimuli was not significantly different from chance. The analysis of the verbal reports indicated that subjects were unable to articulate the rules that governed verb placement in the semi-artificial language.

Experiment 3 introduced several methodological changes. Three, rather than four, rules determined the placement of VPs in the semi-artificial grammar. An elicited imitation task was added to the training phase, i.e. subjects had to first repeat each sentence and then judge whether it was semantically plausible. Learning was still assessed by means of grammaticality judgments, but the measures of awareness were modified by adding a guess category to the confidence ratings and by adding source attributions. The debriefing session still required subjects to judge their overall accuracy and to report any word order rules they might have noticed. The results showed that the modifications resulted in an increased learning effect, but closer analysis suggested that, again, performance was driven by the endorsement of grammatical items. The confidence ratings and the source attributions showed that subjects developed both conscious and unconscious knowledge, even though their verbal reports show no knowledge of the verb placement rules.

Experiment 4 investigated to what extent the limited learning was a result of pre-existing linguistic knowledge. The semantic information was removed from the semi-artificial grammar by replacing English phrases with nonsense syllables, i.e. the training and the testing sets were different from those employed in experiment 3. Subjects were trained under incidental learning conditions by means of elicited imitations. Learning and awareness were measured by the same tasks as in experiment 3. The results showed that removing the semantic layer from the artificial language did not result in a substantial increase in learning. Classification performance on grammatical items was similar to the one observed in experiment 3, but performance on ungrammatical items improved. In terms of awareness, experiment 4, like experiment 3, provided evidence for both conscious and unconscious knowledge.

Experiment 5 introduced a new training task. Before judging semantic plausibility, subjects were asked to identify the syntactic structure of the stimulus sequence. It was assumed that directing subjects' attention to sentence structure would increase learning because clause type was an important cue to verb placement. The training and the testing items as well as the tasks of the testing phase were the same as those employed in experiment 3. The results indicated that the introduction of the sentence identification task produced a greater learning effect compared to the previous experiments. Further analysis showed that this was due to an improvement in rejection performance. At the same time, learning was still somewhat limited. For example, subjects did not associate the verb final position with subordinate clauses, despite the fact that the training task involved processing sentence structure more directly. No subject was able to report the verb placement rules in the debriefing session, but the analysis of confidence ratings and source attributions suggested that subjects acquired both conscious and unconscious knowledge.

Experiment 6 differed from the previous five experiments in that subjects were exposed to the semi-artificial grammar under intentional learning conditions. The training and testing sets as well as the test tasks were the same as those employed in experiment 3. In contrast, however, subjects in experiment 6 were asked to discover word order rules. Subjects could focus on this task exclusively and were not required to enter plausibility judgments. The results showed that the intentional learners outperformed all other groups. Interestingly, only four subjects seemed to have acquired the three rules that determined verb placement. Their classification performance was categorical, i.e. they endorsed over 90% of all grammatical sentences and less than 10% of all ungrammatical ones. Two of these categorical subjects were also able to report all rules in the debriefing session. The performance of the noncategorical subjects was surprisingly similar to the performance of the subjects of previous experiments. The measures of awareness indicated that both categorical and noncategorical subjects acquired conscious and unconscious knowledge.

The findings reported in this dissertation have several implications for our understanding of language acquisition and for future research. Firstly, while the precise form of the knowledge acquired in these experiments is unclear, the findings provided no evidence for rule learning in the vast majority of subjects. It suggests that subjects in these types of experiments (and perhaps in natural language acquisition) do not acquire linguistic rules. The results support Shanks (1995; Johnstone & Shanks, 2001), who argues against the possibility of implicit rule learning. Rule learning was only observed in the categorical subjects of experiment 6 and clearly associated with explicit learning. Additional research is necessary to characterize the nature of what was learned incidentally and to determine more precisely what conditions lead to successful rule acquisition. For example, it would be important to establish whether increased exposure would lead to the development of rule knowledge under both incidental and intentional learning conditions. Given that only six subjects appear to have acquired rules, it would also be of interest to investigate whether individual differences (e.g., working memory capacity) play a significant role in predicting success at rule acquisition.

Secondly, while adults can acquire knowledge implicitly, the work reported in this dissertation also demonstrates that adult syntactic learning results predominantly in a conscious (but largely unverbalizable) knowledge base. This might, in fact, explain why learning was very much constrained across all experiments. *VF sentences, for example, were generally accepted as grammatical, even though this option was restricted to subordinate clauses. It would be of interest to run the same, or a slightly adapted, version of experiments 3 and 6 with children in order to determine whether there are child-adult differences in syntactic learning. Young learners, especially pre-schoolers without extensive metalinguistic knowledge, might display more implicit learning than adults. It would be also interesting to establish whether this would lead to greater learning effects across patterns. If this were the case, then the fact that adults are potentially less likely to engage in implicit learning of a novel language might help explain why they frequently fail to achieve native-like levels of proficiency in a novel language, despite prolongued periods of exposure.

From a methodological perspective, the results of the experiments confirm that relying on verbal reports as a measure of awareness is not sufficient. The verbal reports collected at the end of the experiment were helpful in determining what aspects of the semi-artificial grammar subjects had consciously noticed. At the same time, verbal reports were clearly not sensitive enough to assess whether subjects were aware of the knowledge they had acquired. Confidence ratings and source attributions provided a very useful method for capturing low levels of awareness and to observe the conscious status of both structural and judgment knowledge. Future experiments in the statistical learning tradition and in the field of SLA would benefit from the introduction of this relatively simple, but effective way of assessing awareness. As far as the assessment of learning effects is concerned, it would be useful to explore whether the binary grammaticality judgments used in this dissertation are an adequate measure of learning. It could be, for example, that learning would not have appeared as constrained if more sensitive measures had been used. Scott & Dienes (2008) have shown that familiarity is the essential source of knowledge in AGL, which suggests that, in terms of offline measures, familiarity or preference judgments might be more suitable. The use of online measures in particular, e.g. tracking eye movements, recording response latency (e.g. in a RSVP task) or measuring event-related potentials (Tokowicz & MacWhinney, 2006), seems to be a especially promising way to assess the knowledge underlying native-speaker intuitions. Finally, it should also be noted that all experiments in this dissertation focused on comprehension tasks. Given that communication usually consists of both comprehension and production of language, it would be important to determine whether the knowledge acquired in artificial language experiments transfers successfully to production tasks. However, further research is necessary to address the shortcomings and the unanswered questions of the present dissertation.

Appendix

1. Stimulus material

Following convention in linguistics, the question marks (?) indicate implausible sentences and asterisks (*) ungrammatical sentences.

1.1. Experiment 1

1.1.1. Training set

V2 sentences:

1. Peter entertained today his colleagues with an interesting performance.
2. Brian reacted usually badly against unfounded accusations.
3. Jack operated today on his patient for many hours.
4. Mike saved usually a lot of money during the vacation period.
5. Joanna heaved today the boxes onto the table.
6. Sarah satisfied usually all the requirements promptly.
7. Rose gambled today with her savings at the casino.
8. Jessica played usually an important part in the school plays.
9. Usually forgot Peter all of his problems during the vacation.
10. Today justified Brian the investments during the meeting.
11. Usually drew Jack his clients in a realistic fashion.
12. Today executed Mike the plan with efficiency.
13. Usually sat Joanna in the seminar next to the door.
14. Today ranked Sarah her employees according to their performance.
15. Usually rebutted Rose her employee’s claim for compensation.
16. Today ranted Jessica about the government’s plans.
17. ? Peter whispered usually all evening to a potato.
18. ? Brian gossiped today in his office with a Martian.
19. ? Jack scrutinized usually the old flat on planet Mars.
20. ? Mike reported today to his friends the great haystack.
21. ? Joanna laminated usually the recipe cards with peaches.
22. ? Sarah lectured today for a long time to the doorknob.

23. ? Rose covered usually the fields with gold watches.
24. ? Jessica wrote today a book with her poodle.
25. ? Today abandoned Peter his cats on planet Venus.
26. ? Usually held Brian important discussions with a fork.
27. ? Today bit Jack the vicar in the foot.
28. ? Usually earned Mike a melon for his birthday.
29. ? Today met Joanna in Buckingham the President of England.
30. ? Usually preached Sarah with Elvis Presley in Memphis.
31. ? Today matriculated Rose at university on planet Saturn.
32. ? Usually repaired Jessica the guitar with oranges and lemons.

V2VF sentences:

1. Peter has today the college statutes in his speech challenged.
2. Brian has usually many shots during his matches defended.
3. Jack has today excellent dessert at a restaurant eaten.
4. Mike has usually with much zest in the varsity match competed.
5. Joanna has today medication at the chemist's nicked.
6. Sarah has usually surprise parties for her husband planned.
7. Rose has today her PhD in Linguistics finished.
8. Jessica has usually every night in a club danced.
9. Today were his new designs in a fashion show presented.
10. Usually were his parents to a free lunch treated.
11. Today were his children at the hospital vaccinated.
12. Usually were the ambassadors for important discussions summoned.
13. Today were her shares at the stock exchange sold.
14. Usually were her friends to South America flown.
15. Today were her parents at home telephoned.
16. Usually were her collections to a foundation donated.
17. ? Peter has today at midnight by himself erupted.
18. ? Brian has usually the course assignments with his fist pulverized.
19. ? Jack has today with an airplane to Norway sailed.
20. ? Mike has usually the window with a budgie shattered.
21. ? Joanna has today a tasty cake in the town centre killed.
22. ? Sarah has usually in the afternoon a soup e-mailed.

23. ? Rose has usually California for breakfast devoured.
24. ? Jessica has usually three apples with a knife assassinated.
25. ? Today were his parents on their wedding anniversary watered.
26. ? Usually were his parents by a table caressed.
27. ? Today were medieval languages in Essex detonated.
28. ? Today were many psychologists in her garden imbedded.
29. ? Usually were her friends in a swimming pool stamped.
30. ? Today were her children in the astronomy class abbreviated.
31. ? Usually were her friends to a salad married.
32. ? Today were his sculptures at Sotheby's hiked.

V2-VF sentences:

1. Peter repeated today that the movers his furniture scratched.
2. Brian reckoned usually that his students about their classes cared.
3. Jack asserted today that his father interesting appliances created.
4. Mike stressed usually that his children their teeth brushed.
5. Joanna contested today that her husband firearms possessed.
6. Sarah insisted usually that the judge innocent people protected.
7. Rose realized today that her neighbour with heavy drugs dealt.
8. Jessica insinuated usually that all jurors against her defendants judged.
9. Usually advised Peter that his students the vocabulary memorized.
10. Today conceded Brian that the evidence his brother implicated.
11. Usually swore Jack that this ointment all wounds healed.
12. Today reflected Mike that the fire department too few hours trained.
13. Usually alleged Joanna that her husband their bank account overdrew.
14. Today opined Sarah that the storm his parents' boat overturned.
15. Usually yelled Rose that the government drinking shamelessly encouraged.
16. Today figured Jessica that her husband an affair conducted.
17. ? Peter conjectured usually that his professor a cucumber massacred.
18. ? Brian presumed today that his supervisor the best university swallowed.
19. ? Jack hypothesized usually that his parents the banknotes burnt.
20. ? Mike inferred today that the new computer in his office fainted.
21. ? Joanna posited usually that the new DVD in her office jumped.
22. ? Sarah denied today that her mother the roast beef attacked.

23. ? Rose assumed usually that the new administrator many debates exploded.
24. ? Jessica confessed today that her horse the corridor murdered.
25. ? Today testified Peter that his horses the plough burned.
26. ? Usually guessed Brian that the UN a new country transpired.
27. ? Today beheld Jack that the earth around the tomato rotated.
28. ? Usually reasoned Mike that the company cash digested.
29. ? Today proposed Joanna that the moment for retirement crashed.
30. ? Usually accepted Sarah that her investment in sports facilities drowned.
31. ? Today verified Rose that her office desk English mumbled.
32. ? Usually admitted Jessica that the university in the summer melted.

VF-V1 sentences:

1. Since his teacher criticism voiced, put Peter more effort into his homework.
2. After the instructor a sword brandished, focused Brian more on his defensive stance.
3. Since a silver chandelier after dinner vanished, accused Jack his guests of theft.
4. After his father a private tutor appointed, concentrated Jim harder on his assignments.
5. Since her team against the weak opponents lost, fired Joanna the manager.
6. After her mother the house cleaned, admired Sarah the state of the living-room.
7. Since her car in the middle of the road stopped, began Rose to take the tube.
8. After the police her car apprehended, expected Jessica a heavy fine.
9. After the vending machine fifty pence withheld, complained Peter to the company.
10. Since the famous journalist his daughters interviewed, bragged Brian about his offspring.
11. After the university press his book published, sent Jack many copies to his friends.
12. Since the storm many trees uprooted, planted Jim a dozen saplings.
13. After the lottery the winning numbers announced, looked Joanna for her ticket.
14. Since the company their job offer withdrew, searched Sarah the employment listings.
15. After her friend four goals in a match scored, viewed Rose him with different eyes.
16. Since the factory the river polluted, avoided Jessica contact with tap water.
17. ? Since his boss many sessions scheduled, squandered Peter his morning with a pumpkin.
18. ? After his dog the cats chased, assembled Brian a new apricot.
19. ? Since his children the old car dirtied, invested Jack in a new appetizer.
20. ? After his wife a sandwich craved, rushed Mike to Japan.
21. ? Since her employer generous payments promised, took Joanna a holiday on the moon.
22. ? After her daughter to the museum walked, phoned Sarah their collie.

23. ? Since her mother sculptures inherited, showed Rose the collection to a bear.
24. ? After her family the large mansion inhabited, painted Jessica a lot of cucumbers.
25. After his children for the bus waited, boarded Peter the tube to New Zealand.
26. ? Since his family a boutique owned, cycled Brian on a luxurious camel.
27. ? After his wife a thief surprised, communicated Jack with the police banana.
28. ? Since his friend in Nigeria dwelled, received Mike postcards from Rome.
29. ? After her husband in many films participated, accommodated Joanna famous caterpillars.
30. ? Since her friend to Ely hitchhiked, spoke Sarah to a lot of sausages.
31. ? After her mother from the company resigned, erased Rose the nearby satellite.
32. ? Since her son a medical doctor became, celebrated Jessica a surprise doorbell.

1.1.2. Testing set

V2 sentences:

1. Yesterday scribbled David a long letter to his family.
2. Recently recognized John the stolen paintings in a museum.
3. Yesterday loaded Jim the wagon with hay.
4. Recently told Paul his parents the good news.
5. Yesterday decided Emma to eat in an Italian restaurant.
6. Recently drank Susan a glass of wine in the local bar.
7. Yesterday enjoyed Chloe the food in the dining hall.
8. Recently spent Jennifer her day at the library.

V2VF sentences:

1. David was yesterday by a new company hired.
2. John was recently by his employees imitated.
3. Emma was yesterday to a birthday party invited.
4. Susan was recently to a different office transferred.
5. Yesterday have his students with the American researcher chatted.
6. Recently have his parents an accountant consulted.
7. Yesterday have her parents the homework inspected.
8. Recently have her students the library books returned.

V2-VF sentences:

1. David speculated yesterday that the suspect from prison escaped.
2. John disclosed recently that the company sufficient funds generated.
3. Jim said yesterday that his parents the pub rented.
4. Paul argued recently that the chairman the wrong figures displayed.
5. Yesterday suspected Emma that the old professor the computer broke.
6. Recently lamented Susan that her son the baptism missed.
7. Yesterday learned Chloe that the university her application rejected.
8. Recently acknowledged Jennifer that her children to England moved.

VF-V1 sentences:

1. When his parents groceries needed, purchased David everything necessary.
2. Because his children fairy tales loved, invented John many stories.
3. When his wife the office building left, prepared Jim her favourite dinner.
4. Because his friend very intelligent appeared, asked Paul him for advice.
5. Because her daughter dogs adored, gave Emma her a puppy.
6. When her daughter in Manchester worked, visited Susan this city many times.
7. When her children from school arrived, interrupted Chloe her work.
8. Because her company capital lacked, organized Jennifer a fundraiser

*VF sentences:

1. Recently David to his son’s house in Wales drove.
2. Yesterday John a meeting with his tutor arranged.
3. Recently Jim the Boston Marathon in four hours ran.
4. Yesterday Paul with distinction from Cambridge University graduated.
5. Recently Emma a wonderful meal for her in-laws cooked.
6. Yesterday Susan an old car with her savings bought.
7. Recently Chloe to New York with her husband travelled.
8. Yesterday Jennifer the new CD with her friend discussed.

*V3VF sentences:

1. Yesterday the guitar was by David smashed.
2. Recently the bucket was by John filled.
3. Yesterday a ball was by Emma juggled.

4. Recently much furniture was by Susan imported.
5. Yesterday his parents have his decision undermined.
6. Recently his parents have the envelope sealed.
7. Yesterday her parents have the payments suspended.
8. Recently her parents have the television show recorded.

*V2-V1 sentences:

1. Recently maintained David that abstained his father from unhealthy food.
2. Yesterday claimed John that liked his mother Melbourne.
3. Recently thought Jim that slept the new CEO in his office.
4. Yesterday mentioned Paul that designed the architect a new building.
5. Emma emphasized recently that offered the university an interesting career.
6. Susan explained yesterday that remained the profit below the estimate.
7. Chloe remarked recently that resided her son in Portugal.
8. Jennifer suggested yesterday that declared the company bankruptcy.

*VF-VF sentences:

1. Because his son an instrument wanted, David with the music teacher talked.
2. When his parents to Paris retired, John a lot to France flew.
3. Because his children a calculator required, Jim the electronics store called.
4. When his director confidential information divulged, Paul the department quit.
5. Because her husband in Oxford taught, Emma the job transfer declined.
6. When her husband a lot of money made, Susan an extravagant necklace acquired.
7. When her daughter in Australia lived, Chloe Melbourne very well knew.
8. Because her husband a boring conference attended, Jennifer at the hotel stayed.

1.2. Experiment 2

1.2.1. Training set

V2 sentences:

1. A thousand pounds won Peter today in the lottery.
2. His lunch ate Brian usually at a restaurant.
3. Her art collection donated Joanna some time ago to a charitable foundation.
4. A surprise trip planned Sarah in the evening for her husband.

5. During the meeting justified today Jack the new investment plan.
6. In a realistic fashion sketched usually Mike his clients.
7. In the casino jeopardized some time ago Jessica her savings.
8. In the bingo team played in the evening Rose an important part.
9. Today executed with efficiency Peter the plan.
10. Usually placed on the desk Brian his writing material.
11. Some time ago forgot in a hotel room Joanna her bathing suit.
12. In the evening ranked according to their performance Sarah her employees.
13. Jack entertained his friends today with an interesting performance.
14. Mike satisfied all the requirements usually in a timely fashion.
15. Jessica telephoned her lawyer some time ago about the unfounded accusations.
16. Rose heaved the boxes in the evening onto the table.
17. ? The meadows covered Peter today with gold watches.
18. ? His guitar repaired Brian usually with oranges and lemons.
19. ? The document e-mailed Joanna some time ago to a tomato.
20. ? The recipe cards laminated in the evening with peaches.
21. ? To a meal treated today Jack his doorknob.
22. ? With a fork held usually Mike important meetings.
23. ? On the moon abandoned some time ago her cats.
24. ? At the restaurant devoured in the evening Rose a tasty computer.
25. ? Today bit in the foot Peter the vicar.
26. ? Usually earned for his birthday Brian a melon.
27. ? Some time ago wrote with her pen Joanna a poodle.
28. ? In the evening met in London Sarah the President of England.
29. ? Jack reported the great haystack today to his friends.
30. ? Mike whispered all evening usually to a potato.
31. ? Rose sailed her boat some time ago into the stratosphere.
32. ? Jessica scrutinized an old flat in the evening on planet Mars.

V2VF sentences:

1. With much zest has Peter today in the varsity match competed.
2. Many shots has Brian usually during his matches defended.
3. The college statutes has Joanna today in her speech challenged.
4. Excellent dessert has Sarah some time ago at a restaurant eaten.

5. At the chemist's has today Jack medication nicked.
6. At the club has usually Mike with complete strangers danced.
7. In the linguistics department has some time ago Jessica her presentation given.
8. For her husband has in the evening Rose a surprise party planned.
9. Today were in a fashion show his new designs presented.
10. Usually were to a free lunch his parents treated.
11. Some time ago were at the hospital her children vaccinated.
12. In the evening were for important discussions the ambassadors summoned.
13. His shares were at the stock exchange today sold.
14. His friends were to South America usually flown.
15. Her collections were to a foundation some time ago donated.
16. Her parents were at home in the evening telephoned.
17. ? A tasty cake has Peter today in a cafe killed.
18. ? The course assignments has Brian usually with his fist pulverized.
19. ? A soup has Joanna some time ago e-mailed.
20. ? The window has Sarah in the evening with a budgie shattered.
21. ? At midnight has today Jack by himself erupted.
22. ? For breakfast has usually Mike California devoured.
23. ? With an airplane has some time ago Jessica to Norway sailed.
24. ? With a knife has in the evening Rose three apples assassinated.
25. ? Today were on their wedding anniversary his parents watered.
26. ? Usually were by a table his parents caressed.
27. ? Some time ago were in Essex medieval languages detonated.
28. ? In the evening were in her garden many psychologists embedded.
29. ? His friends were in a swimming pool today stamped.
30. ? His children were in the astronomy class usually abbreviated.
31. ? Her sculptures were at Sotheby's some time ago hiked.
32. ? Her friends were to a salad in the evening married.

V2-VF sentences:

1. Peter repeated today that his furniture the movers scratched.
2. Brian insisted some time ago that the judge innocent people protected.
3. Joanna contested some time ago that her neighbour illegal firearms possessed.
4. Sarah asserted today that her father interesting appliances created.

5. Jack realized today that with heavy drugs his friend dealt.
6. Mike reckoned usually that about their classes his students cared.
7. Jessica figured some time ago that in nightclub her daughter danced.
8. Rose insinuated in the evening that against her defendant all jurors judged.
9. Today conceded Peter that the evidence his brother implicated.
10. Usually advised Brian that his students the vocabulary memorized.
11. Some time ago stressed Joanna that her children their teeth brushed.
12. In the evening opined Sarah that her husband their bank account overdrew.
13. Today reflected Jack that too few hours the fire department trained.
14. Usually swore Mike that all wounds this ointment healed.
15. Some time ago alleged Jessica that to California the convict escaped.
16. In the evening yelled Rose that an affair her husband conducted.
17. ? Peter hypothesized today that his parents the bank notes burned.
18. ? Peter presumed usually that his supervisor the best university swallowed.
19. ? Joanna denied some time ago that her horse the corridor murdered.
20. ? Sarah conjectured in the evening that her professor a cucumber massacred.
21. ? Jack posited today that in his office the new chair snored.
22. ? Mike assumed usually that many debates the new administrator exploded.
23. ? Jessica proposed some time ago that for retirement the moment crashed.
24. ? Rose inferred in the evening that in her bedroom the new computer fainted.
25. ? Today testified Peter that his horses the plough burned.
26. ? Usually guess Brian that the UN a new country transpired.
27. ? Some time ago confessed Joanna that the roast beef her mother attacked.
28. ? In the evening reasoned Sarah that the company cash digested.
29. ? Today beheld Jack that around the tomato the earth rotated.
30. ? Usually accepted Mike that in sports facilities his investment drowned.
31. ? Some time ago admitted Jessica that in the summer the university melted.
32. ? In the evening verified Rose that English her office desk mumbled.

VF-V1 sentences:

1. Since his father a private tutor appointed, concentrated Peter all day on his assignments.
2. After the police his car apprehended, expected a heavy fine Mike from the authorities.
3. Since the river the factory polluted, avoided Joanna contact with tap water.
4. After her mother the house cleaned, admired Sarah the living room with her husband.

5. Since criticism his teacher voiced, put more effort Jack into his homework.
6. After the university press his book published, sent many copies Mike to his friends.
7. Since the famous journalist her daughters interviewed, bragged about her offspring Jessica to many friends.
8. After the lottery the winning numbers announced, looked for her ticket Rose in the cupboard.
9. Since on the motorway his car stopped, hiked Peter to the next petrol station with his bags.
10. After a sword the instructor brandished, focused Brian on his defensive stance with intensity.
11. Since during dinner a silver chandelier vanished, accused Joanna her guests in the living room.
12. After against the weak opponents her team lost, fired Sarah the manager for the first time.
13. Since their job offer the company withdrew, searched the employment listings Jack for an alternative.
14. After four goals his friend scored, viewed him Mike with different eyes.
15. Since many trees the storm uprooted, planted a dozen saplings Jessica in the garden.
16. After the vending machine fifty pence withheld, complained about this Rose to the company.
17. ? Since his boss many sessions scheduled, squandered Peter his morning with a pumpkin.
18. ? After his dog the cats chased, assembled Brian a new apricot with a screwdriver.
19. ? Since her children the old car dirtied, invested Joanna with a business acquaintance in a new appetizer.
20. ? After her husband a sandwich craved, rushed Sarah to Japan by plane.
21. ? Since his employer generous payments promised, took a holiday Jack on the moon.
22. ? After his wife a thief surprised, contacted the police Mike with a banana.
23. ? Since her family the large mansion inhabited, painted the gardens Jessica with a cucumber.
24. ? After her mother sculptures inherited, showed the collection Rose to a bear.
25. ? Since a boutique his family owned, cycled Peter to London on a luxurious camel.
26. ? After for the bus his children waited, boarded Brian the tube to New Zealand.
27. ? Since in Nigeria her friend dwelled, received Joanna postcards from Rome.
28. ? After to the museum her daughter walked, phoned Sarah their collie with her mobile phone.
29. ? Since to Ely his son hitchhiked, spoke on the phone Jack to many sausages.

30. ? After from the family company his mother resigned, erased the nearby satellite Mike with a laugh.
31. ? Since a medical doctor her son became, celebrated a surprise party Jessica with her doorbell.
32. ? After in many films her husband participated, accommodated many celebrities Rose in her caterpillars.

1.2.2. Testing set

Same as experiment 1 (see 1.1.2).

1.3. Experiment 3, 5, and 6

1.3.1. Training set

V2 sentences:

1. Chris entertained today his colleagues with an interesting performance.
2. Brian played usually an important part in the school plays.
3. Jack reacted in the evening badly against unfounded accusations.
4. Mike operated last June on his patient for many hours.
5. George saved a few months ago a lot of money during the vacation period.
6. Joanna heaved today the boxes onto the table.
7. Sarah forgot usually all of her problems during the vacation.
8. Rose satisfied in the evening all the requirements promptly.
9. Jessica gambled last June with her savings at the casino.
10. Cate justified a few months ago the investments during the meeting.
11. Today executed George the plan with efficiency.
12. Usually drew Mike his clients in a realistic fashion.
13. In the evening ranked Jack his employees according to their performance.
14. Last June rebutted Brian his employee’s claim for compensation.
15. A few months ago ranted Chris about the government’s plans.
16. Today challenged Cate the college statutes in her speech.
17. Usually sat Jessica in the seminar next to the door.
18. In the evening ate Rose excellent dessert at a restaurant.
19. Last June defended Sarah many shots during her matches.

20. A few months ago competed Joanna with much zest in the varsity match.
21. ? Chris gossiped today in his office with a Martian.
22. ? Brian whispered usually all evening to a potato.
23. ? Jack scrutinized in the evening the old flat on planet Mars.
24. ? Mike reported last June to his friends the great haystack.
25. ? George laminated a few months ago the recipe cards with peaches.
26. ? Joanna lectured today for a long time to the doorknob.
27. ? Sarah covered usually the fields with gold watches.
28. ? Rose abandoned in the evening her cats on planet Venus.
29. ? Jessica wrote last June a book with her poodle.
30. ? Cate held a few months ago important discussions with a fork.
31. ? Today bit George the vicar in the foot.
32. ? Usually earned Mike a melon for his birthday.
33. ? In the evening met Jack in Buckingham the President of England.
34. ? Last June preached Brian with Elvis Presley in Memphis.
35. ? A few months ago matriculated Chris at university on planet Saturn.
36. ? Today erupted Cate at midnight by herself.
37. ? Usually repaired Jessica the guitar with oranges and lemons.
38. ? In the evening pulverized Rose the course assignments with her fist.
39. ? Last June sailed Sarah with an airplane to Norway.
40. ? A few months ago shattered Joanna the window with a budgie.

V2-VF sentences:

1. George repeated today that the movers his furniture scratched.
2. Mike reckoned usually that his students about their classes cared.
3. Jack asserted in the evening that his father interesting appliances created.
4. Brian contested last June that his friend firearms possessed.
5. Chris stressed a few months ago that his children their teeth brushed.
6. Cate realized today that her neighbour with heavy drugs dealt.
7. Jessica insisted usually that the judge innocent people protected.
8. Rose insinuated in the evening that all jurors against her defendants judged.
9. Sarah advised last June that her students the vocabulary memorized.
10. Joanna conceded a few months ago that the evidence his brother implicated.
11. Today reflected Chris that the fire department too few hours trained.

12. Usually swore Brian that this ointment all wounds healed.
13. In the evening alleged Jack that his wife their bank account overdrew.
14. Last June yelled Mike that the government drinking shamelessly encouraged.
15. A few months ago opined George that the storm his parents' boat overturned.
16. Today figured Joanna that her husband an affair conducted.
17. Usually emphasized Sarah that the university an interesting career offered.
18. In the evening explained Rose that the profit below the estimate remained.
19. Last June remarked Jessica that her son in Portugal resided.
20. A few months ago suggested Cate that the company bankruptcy declared.
21. ? George presumed today that his supervisor the best university swallowed.
22. ? Mike conjectured usually that his professor a cucumber massacred.
23. ? Jack hypothesized in the evening that his parents the banknotes burnt.
24. ? Brian denied last June that his mother the roast beef attacked.
25. ? Chris inferred a few months ago that the new computer in his office fainted.
26. ? Cate confessed today that her horse the corridor murdered.
27. ? Jessica posited usually that the new DVD in her office jumped.
28. ? Rose assumed in the evening that the new administrator many debates exploded.
29. ? Sarah testified last June that her horses the plough burned.
30. ? Joanna guessed a few months ago that the UN a new country transpired.
31. ? Today beheld Chris that the earth around the tomato rotated.
32. ? Usually reasoned Brian that the company cash digested.
33. ? In the evening proposed Jack that the moment for retirement crashed.
34. ? Last June accepted Mike that his investment in sports facilities drowned.
35. ? A few months ago verified George that his office desk English mumbled.
36. ? Today doubted Joanna that her parents the bookshelf watered.
37. ? Usually deliberated Sarah that the university in the summer melted.
38. ? In the evening questioned Rose that her parents a table with a beagle caressed.
39. ? Last June thought Jessica that the new CEO in his office slept.
40. ? A few months ago mentioned Cate that the architect a new building designed.

VF-V1 sentences:

1. Since his teacher criticism voiced, put Chris more effort into his homework.
2. After the instructor a sword brandished, focused Brian more on his defensive stance.
3. Since his team against the weak opponents lost, fired Jack the manager.

4. After his father a private tutor appointed, concentrated Mike harder on his assignments.
5. Since a silver chandelier after dinner vanished, accused George his guests of theft.
6. After her mother the house cleaned, admired Joanna the state of the living-room.
7. Since her car in the middle of the road stopped, began Sarah to take the tube.
8. After the police her car apprehended, expected Rose a heavy fine.
9. After the vending machine fifty pence withheld, complained Jessica to the company.
10. Since the famous journalist her daughters interviewed, bragged Cate about her offspring.
11. After the university press his book published, sent George many copies to his friends.
12. Since the storm many trees uprooted, planted Mike a dozen saplings.
13. After the lottery the winning numbers announced, looked Jack for his ticket.
14. Since the company their job offer withdrew, sought Brian the employment listings.
15. After his friend four goals in a match scored, viewed Chris him with different eyes.
16. Since the factory the river polluted, avoided Cate contact with tap water.
17. After the magazine her new designs presented, had Jessica many phone calls from interested buyers.
18. Since her parents the grandchildren to a free lunch treated, passed Rose her day in bed.
19. After the nurse her children vaccinated, texted Sarah her husband on his mobile phone.
20. Since many demonstrators the company offices occupied, summoned Joanna her management team for important discussions.
21. ? Since his boss many sessions scheduled, squandered Chris his morning with a pumpkin.
22. ? After his cats the rodents hunted, assembled Brian a new apricot.
23. ? Since his children the old car dirtied, invested Jack in a new appetizer.
24. ? After his wife a sandwich craved, rushed Mike to Japan.
25. ? Since his employer generous payments promised, enjoyed George a holiday on the moon.
26. ? After her daughter to the museum walked, phoned Joanna their collie.
27. ? Since her mother sculptures inherited, showed Sarah the collection to a bear.
28. ? After her family the large mansion inhabited, painted Rose a lot of cucumbers.
29. ? After her children for the bus waited, boarded Jessica the tube to New Zealand.
30. ? Since her family a boutique owned, cycled Cate on a luxurious camel.
31. ? After his wife a thief surprised, communicated George with the police banana.
32. ? Since many people carrots in their garden imbedded, issued Mike a vegetable warning.
33. ? After his wife in many films participated, accommodated Jack famous caterpillars.

34. ? Since his friend to Ely hitchhiked, spoke Brian to a lot of sausages.
35. ? After his mother from the company resigned, erased Chris the nearby satellite dish.
36. ? Since her son a medical doctor became, celebrated Cate a surprise doorbell.
37. ? After the bakery in her neighbourhood closed, killed Jessica a tasty cake in the town centre.
38. ? Since her friends often hunger feigned, e-mailed Rose in the afternoon a soup.
39. ? After her children the United States toured, devoured Sarah California for breakfast.
40. ? Since many apples to the ground fell, assassinated Joanna three of them with an expensive knife.

1.3.2. Testing set

V2 sentences:

1. Yesterday scribbled David a long letter to his family.
2. Recently recognized John the stolen paintings in a museum.
3. In the afternoon loaded Jim the wagon with hay.
4. After dinner told Paul his parents the good news.
5. Some time ago decided Peter during a meal on a business proposal.
6. Yesterday enjoyed Emma the food in the dining hall.
7. Recently spent Chloe her day at the library.
8. In the afternoon drank Jennifer a glass of wine in the local bar.
9. After dinner drove Susan to his son's house in Wales.
10. Some time ago arranged Janet a meeting with her tutor.

V2-VF sentences:

1. Paul speculated yesterday that the suspect from prison escaped.
2. Peter disclosed recently that the company sufficient funds generated.
3. Emma said in the afternoon that her parents the pub rented.
4. Chloe argued after dinner that the chairman the wrong figures displayed.
5. Jennifer suspected some time ago that the old professor the computer broke.
6. Yesterday lamented Susan that her son the baptism missed.
7. Recently learned Janet that the university her application rejected.
8. In the afternoon acknowledged David that her children to England moved.
9. After dinner maintained John that his father in French cuisine indulged.

10. Some time ago claimed Jim that his mother Melbourne liked.

VF-V1 sentences:

1. When his parents yesterday groceries needed, purchased David everything necessary.
2. Because his collie recently very annoyingly acted, asked John the vet for advice.
3. When his wife in the afternoon the office building left, prepared Jim dinner for the entire family.
4. Because his children after dinner fairy tales loved, invented Paul many stories.
5. When his daughter some time ago in Manchester worked, visited Peter this city many times.
6. When the children yesterday the new flowerbed destroyed, chased Emma the culprits around the garden.
7. Because her company recently capital lacked, organized Chloe a fundraiser.
8. When her children in the afternoon from school arrived, interrupted Jennifer her work.
9. Because her daughter after dinner sweets adored, brought Susan many desserts to the table.
10. When her friend some time ago in Italy dwelled, received Janet postcards from Rome.

*VF sentences:

1. Yesterday David with distinction from Cambridge University graduated.
2. Recently John the Boston Marathon in four hours ran.
3. In the afternoon Emma a wonderful meal for her in-laws cooked.
4. After dinner Chloe an old car with her savings bought.
5. Some time ago Jennifer to New York with her husband travelled.

*VF-V2 sentences:

1. Because his son yesterday an instrument wanted, Jim talked with the music teacher.
2. When his parents recently to Paris retired, Paul flew a lot to France.
3. Because his children in the afternoon a calculator required, Peter called the electronics store.
4. When her director after dinner confidential information divulged, Susan quit the department.
5. Because her husband some time ago in Oxford taught, Janet declined the job transfer.

*V1-VF sentences:

1. Acquired David yesterday an extravagant watch when his partner a lot of money made.
2. Went John recently to the cinema on his own because his favourite actress in a new film starred.
3. Discussed Jim after dinner the new CD with his friend after his wife the kids to bed took.
4. Stayed Emma in the afternoon at the hotel because her husband a boring conference attended.
5. Knew Chloe some time ago Melbourne very well when her daughter in Australia lived.

*V1 sentences:

1. Hired Paul yesterday two new chefs for his restaurant.
2. Imitated Peter recently his best employee during the Christmas dinner.
3. Transferred Jennifer in the afternoon three employees to a different office.
4. Invited Susan after dinner some colleagues to her birthday party.
5. Chatted Janet some time ago with her new students for a long time.

*V3 sentences:

1. Recently David consulted an accountant during a five-hour meeting.
2. Yesterday John inspected the homework with increased rigor.
3. In the afternoon Emma returned the library books to the stacks.
4. After dinner Chloe smashed the guitar without any warning.
5. Some time ago Jennifer filled the bucket with apples.

*V4 sentences:

1. Yesterday Jim the television show recorded with their new VCR.
2. Recently Paul much furniture imported for her new weekend retreat.
3. In the afternoon Peter his decision undermined with poignant arguments.
4. After dinner Susan the envelope sealed with wax.
5. Some time ago Janet the payments suspended for an indefinite period.

1.4. Experiment 4

1.4.1. Training set

V2 sequences:

1.	Se	Ko	Fi	Tu	Pu
2.	Se	Ko	Fa	Pa	Pa
3.	Se	Ki	Fi	Tu	Pu
4.	Se	Ki	Fa	Pa	Tu
5.	Sa	Ki	Fi	Pa	Tu
6.	Sa	Ko	Fi	Te	Pa
7.	Sa	Ki	Fa	Pu	Pa
8.	Sa	Ko	Fa	Pa	Tu
9.	Fa	Ki	Se	Pa	Pa
10.	Fa	Ko	Sa	Te	Tu
11.	Fa	Ki	Sa	Te	Te
12.	Fa	Ko	Se	Pu	Tu
13.	Fi	Ko	Se	Pa	Pu
14.	Fi	Ki	Se	Pa	Te
15.	Fi	Ko	Sa	Pa	Te

V2-VF sequences:

1.	Se	Ki	Fi	Ra	Se	Pu	Tu	Ki
2.	Se	Ko	Fi	Ra	Se	Pa	Tu	Ko
3.	Se	Ko	Fa	Ra	Sa	Pu	Te	Ki
4.	Se	Ki	Fi	Ra	Sa	Pa	Te	Ko
5.	Sa	Ki	Fi	Ra	Sa	Pu	Tu	Ko
6.	Sa	Ko	Fa	Ra	Sa	Pa	Tu	Ko
7.	Sa	Ki	Fi	Ra	Sa	Pu	Te	Ki
8.	Sa	Ko	Fa	Ra	Se	Pa	Te	Ki
9.	Fa	Ko	Se	Ra	Se	Pu	Tu	Ki
10.	Fi	Ko	Sa	Ra	Se	Pa	Tu	Ko
11.	Fi	Ko	Sa	Ra	Sa	Pu	Te	Ki
12.	Fi	Ki	Se	Ra	Sa	Pa	Te	Ko

13.	Fa	Ko	Sa	Ra	Se	Pu	Tu	Ko
14.	Fa	Ki	Se	Ra	Sa	Pa	Tu	Ki
15.	Fa	Ko	Sa	Ra	Sa	Pu	Te	Ko

VF-V1 sequences:

1.	Ru	Se	Pa	Ki	Ki	Sa	Pa	Te
2.	Ri	Sa	Pa	Ki	Ko	Sa	Pu	Te
3.	Ri	Se	Pu	Ki	Ko	Sa	Pa	Tu
4.	Ru	Se	Pu	Ki	Ko	Sa	Pa	Tu
5.	Ru	Se	Pu	Ki	Ko	Sa	Pa	Tu
6.	Ru	Sa	Pu	Ki	Ko	Se	Pu	Tu
7.	Ri	Se	Pa	Ki	Ko	Se	Pa	Te
8.	Ru	Sa	Pa	Ko	Ki	Sa	Pu	Te
9.	Ri	Sa	Pu	Ko	Ki	Sa	Pu	Tu
10.	Ru	Se	Pa	Ko	Ki	Sa	Pa	Tu
11.	Ri	Se	Pu	Ko	Ki	Sa	Pa	Tu
12.	Ri	Sa	Pu	Ko	Ki	Se	Pu	Tu
13.	Ri	Sa	Pa	Ko	Ko	Se	Pu	Te
14.	Ru	Sa	Pa	Ko	Ko	Se	Pu	Te
15.	Ru	Sa	Pu	Ko	Ko	Se	Pu	Tu

1.4.2. Testing set

V2 sequences:

1.	Sa	Ko	Fi	Pa	Te
2.	Sa	Ki	Fi	Pu	Tu
3.	Sa	Ko	Fu	Te	Pa
4.	Se	Ki	Fa	Tu	Pu
5.	Se	Ko	Fa	Pa	Pu
6.	Fa	Ki	Se	Pu	Tu
7.	Fa	Ko	Se	Te	Pa
8.	Fa	Ki	Se	Tu	Te
9.	Fi	Ko	Sa	Pa	Te
10.	Fi	Ki	Sa	Pu	Tu

V2-VF sequences:

1.	Se	Ko	Fa	Ra	Sa	Pa	Tu	Ki
2.	Se	Ki	Fa	Ra	Se	Pu	Pa	Ko
3.	Se	Ko	Fa	Ra	Sa	Te	Te	Ki
4.	Sa	Ki	Fi	Ra	Se	Tu	Te	Ko
5.	Sa	Ko	Fi	Ra	Sa	Pa	Tu	Ki
6.	Fi	Ki	Sa	Ra	Se	Pu	Te	Ko
7.	Fi	Ko	Sa	Ra	Sa	Te	Tu	Ki
8.	Fi	Ki	Sa	Ra	Se	Tu	Pa	Ko
9.	Fa	Ko	Se	Ra	Sa	Pa	Pu	Ki
10.	Fa	Ki	Se	Ra	Se	Pu	Pu	Ko

VF-V1 sequences:

1.	Ri	Sa	Te	Ki	Ko	Se	Pu	Te
2.	Ru	Sa	Tu	Ko	Ko	Sa	Te	Tu
3.	Ri	Sa	Pa	Ki	Ki	Se	Tu	Pa
4.	Ru	Sa	Pu	Ko	Ki	Sa	Pa	Pu
5.	Ri	Sa	Te	Ki	Ko	Se	Pu	Pa
6.	Ru	Se	Tu	Ko	Ko	Sa	Pa	Tu
7.	Ri	Se	Pa	Ki	Ki	Se	Pu	Pa
8.	Ru	Se	Pu	Ko	Ki	Sa	Te	Te
9.	Ri	Se	Te	Ki	Ko	Se	Tu	Te
10.	Ru	Se	Tu	Ko	Ko	Sa	Pa	Tu

*V1 sequences:

1.	Ko	Se	Fi	Te	Pa
2.	Ko	Sa	Fi	Tu	Pu
3.	Ko	Se	Fa	Pa	Te
4.	Ki	Sa	Fa	Pu	Tu
5.	Ki	Se	Fi	Pu	Pa

*V3 sequences:

1.	Sa	Fa	Ki	Pa	Te

2.	Se	Fa	Ki	Pu	Tu
3.	Sa	Fi	Ki	Te	Pa
4.	Se	Fi	Ko	Tu	Pu
5.	Sa	Fa	Ko	Pa	Pu

*V4 sequences:

1.	Se	Fi	Te	Ko	Pa
2.	Sa	Fi	Tu	Ko	Pu
3.	Se	Fa	Pa	Ko	Te
4.	Sa	Fa	Pu	Ki	Tu
5.	Se	Fi	Pu	Ki	Pa

*VF sequences:

1.	Sa	Fi	Pu	Te	Ki
2.	Se	Fi	Tu	Tu	Ki
3.	Sa	Fa	Pu	Tu	Ki
4.	Se	Fa	Pu	Tu	Ko
5.	Sa	Fi	Pu	Te	Ko

*VF-V2 sequences:

1.	Ri	Sa	Pu	Ko	Se	Ko	Pa	Te
2.	Ru	Se	Pa	Ko	Se	Ko	Pa	Tu
3.	Ri	Sa	Pu	Ko	Sa	Ki	Pu	Tu
4.	Ru	Se	Pa	Ki	Se	Ko	Pa	Te
5.	Ri	Sa	Pa	Ko	Sa	Ki	Pu	Te

*V1-VF sequences:

1.	Ko	Se	Fi	Ra	Sa	Pa	Te	Ki
2.	Ki	Sa	Fa	Ra	Se	Pa	Tu	Ki
3.	Ko	Se	Fi	Ra	Sa	Pu	Tu	Ko
4.	Ki	Sa	Fi	Ra	Se	Pa	Te	Ki
5.	Ko	Se	Fa	Ra	Sa	Pu	Te	Ko

2. Instructions

The following boxes correspond to slides that participants saw on the computer screen.

2.1. Experiments 1 and 2

2.1.1. Training phase

Slide 1:

In the following experiment on sentence comprehension you will listen to scrambled sentences. Your task is to pay attention to the meaning of the sentences and to judge whether the statements made are plausible or not plausible. We will begin with a short practice session that shows you how to make the plausibility judgments. Press SPACE BAR to continue.

Slide 2:

In the following practice session you will listen to four scrambled sentences. For each sentence you will be asked: Do you think this is plausible? Press / if you think the statement was plausible. Press Z if you think the statement was not plausible. If your plausibility judgment is not correct, the sentence will be repeated and you required to judge again. Press SPACE BAR to begin the practice session.

Slide 3:

You are now ready to start the experiment. You will listen to 128 scrambled sentences and asked to judge the plausibility of each sentence. However, after you have judged all 128 sentences once, a message on the screen will appear and the entire sentence set will be played again, though in a different (random) order. In total, you will thus have to make 256 plausibility judgments. Remember: Press / if you think the statement was plausible. Press Z if you think the statement was not plausible. If your plausibility judgment is not correct, the sentence will be repeated and you required to judge again. Press SPACE BAR to begin the experiment.

2.1.2. Testing phase

Experimental group version

Slide 1:

The scrambling of the previous sentence was not arbitrary but followed a complex system. In the second (and final) part of the experiment, you will listen to 64 new sentences. This time, your task is: (1) to judge whether the sentences are formed in accordance with the above-mentioned system or not, and (2) to indicate how confident you are in your decision. Rely on your intuition when judging the well-formedness of the sentences. Response times will be measured. We will begin with a short practice session. Press SPACE BAR to continue.

Slide 2:

In the following practice session you will listen to four scrambled sentences.

For each sentence you will be asked two questions:

Question 1: Was the sentence grammatical?
Press / if you think the sentence was grammatical.
Press Z if you think the sentence was ungrammatical.

Question 2: How confident are you in your decision?
Press 1 if your level of confidence is LOW.
Press 9 if your level of confidence is HIGH.

No feedback will be provided in this section: A new sentence will be played independently of whether your well-formedness judgments are accurate or not.

Press SPACE BAR to begin the practice session.

Slide 3:

You are now ready to start the last part of the experiment. It consists of 64 new scrambled sentences.

Grammaticality judgment
Press / if you think the sentence was grammatical.
Press Z if you think the sentence was ungrammatical.

Confidence judgment
Press 1 if your level of confidence is LOW.
Press 9 if your level of confidence is HIGH.

Rely on your intuition when judging the grammaticality of the sentences.

Press SPACE BAR to start.

Slide 4:

You have judged all 128 sentences once.

All sentences will be played again, though in a different (random) order.

Your task is the same as before: Judge whether every sentence is plausible or not.

Remember:
Press / if you think the statement was plausible.
Press Z if you think the statement was not plausible.

Slide 5:

Congratulations! You have reached the END OF THE EXPERIMENT. IMPORTANT: Before leaving, please take a minute to fill in the questionnaire on your right. Thank you very much for participating!

Control group version

Slide 1:

In the following experiment on sentence comprehension you will listen to scrambled sentences. Your task is to listen to the sentences and to judge whether they are grammatical or ungrammatical. We will begin with a short practice session that shows you how to make the grammaticality judgments. Press SPACE BAR to continue.

Slide 2:

In the following practice session you will listen to four scrambled sentences. For each sentence you will be asked two questions: Question 1: Was the sentence grammatical? Press / if you think the sentence was grammatical. Press Z if you think the sentence was ungrammatical. Question 2: How confident are you in your decision? Press 1 if your level of confidence is LOW. Press 9 if your level of confidence is HIGH. No feedback will be provided in this section: A new sentence will be played independently of whether your well-formedness judgments are accurate or not. Press SPACE BAR to begin the practice session.

Slide 3:

You are now ready to start the experiment. It consists of 64 new scrambled sentences. Grammaticality judgment Press / if you think the sentence was grammatical. Press Z if you think the sentence was ungrammatical. Confidence judgment Press 1 if your level of confidence is LOW. Press 9 if your level of confidence is HIGH. Rely on your intuition when judging the grammaticality of the sentences. Press SPACE BAR to start.

Slide 4:

Congratulations! You have reached the END OF THE EXPERIMENT. IMPORTANT: Before leaving, please take a minute to fill in the questionnaire on your right. Thank you very much for participating!

2.2. Experiment 3

2.2.1. Training phase

Slide 1:

In the following experiment on sentence comprehension you will listen to scrambled sentences. Your task is (1) to repeat the sentence once the prompt appears and to (2) to judge whether the statements made are plausible or not plausible. We will begin with a short practice session that shows you how to make the plausibility judgments. Press SPACE BAR to continue.

Slide 2:

In the following practice session you will listen to four scrambled sentence. For each sentence you will be asked: Do you think this is plausible? Press / if you think the statement was plausible. Press Z if you think the statement was not plausible. If your plausibility judgment is not correct, the sentence will be repeated and you required to judge again. Press SPACE BAR to begin the practice session.

Slide 3:

You are now ready to start the experiment. You will listen to 120 scrambled sentences. After every sentence, a prompt will appear on the screen and ask you to repeat it as well as you can. You will then be asked to judge how plausible the content of the sentence was. Remember: Press / if you think the statement was plausible. Press Z if you think the statement was not plausible. If your plausibility judgment is not correct, the sentence will be repeated and you required to judge again. There is no need though to repeat it verbally. Press SPACE BAR to begin the experiment.

2.2.2. Testing phase

Experimental group version

Slide 1:

The scrambling of the previous sentences was not arbitrary but followed a complex system. In the second (and final) part of the experiment, you will listen to 60 new sentences. This time, your task is: (1) to judge whether the sentences are formed in accordance with the above-mentioned system or not, and (2) to indicate how confident you are about your decision.

Rely on your intuition when judging the well-formedness of the sentences. We will begin with a short practice session. Press SPACE BAR to continue.

Slide 2:

In the following practice session you will listen to four scrambled sentences. For each sentence you will be asked three questions: Question 1: Was the sentence grammatical? In other words, did the sentence follow the same word order rules or not? Press / if you think the sentence was grammatical. Press Z you think the sentence was ungrammatical. Press SPACE BAR to continue.

Slide 3:

Question 2: How confident are you about your decision? Press 0 if you believe you have no confidence in your decision, i.e. you are guessing. Press 5 if you are somewhat confident in your decision. Press 9 if you are very confident in your decision. Press SPACE BAR to continue.

Slide 4:

Question 3: What is the basis of your grammaticality judgment? Press G if you believe your judgment was based on a guess. = You might as well have flipped a coin. Press I if you believe your judgment was based on intuition. = You are confident in your decision but don't know why it's right. Press M if you believe you relied on memory for previous sentences. = You recall part or the entire sentence from the training phase. Press R if you believe you applied a rule in judging the sentence. = You have acquired a rule during training and could verbalize it. Press SPACE BAR to begin the practice session.

Slide 5:

You are now ready to start the last part of the experiment. It consists of 60 new scrambled sentences. Grammaticality judgment Press / if you think the sentence was grammatical. Press Z if you think the sentence was ungrammatical. Confidence judgment Press 0 if you are not confident at all, i.e. you are guessing. Press 5 if you are somewhat confident in your decision. Press 9 if you are very confident in your decision. Basis for grammaticality judgment Press G if you believe your judgment was based on a guess. Press I if you believe your judgment was based on intuition. Press M if you believe you relied on memory for previous sentences. Press R if you believe you applied a rule in judging the sentence. Press SPACE BAR to start.

Slide 6:

Congratulations! You have reached the END OF THE EXPERIMENT. IMPORTANT: Before leaving, please take a minute to fill in the questionnaire on your right. Thank you very much for participating!

Control group version

Slide 1:

In the following experiment on sentence comprehension you will listen to scrambled sentences. Your task is to listen to the sentences and to judge whether they are grammatical or ungrammatical. We will begin with a short practice session that shows you how to make the grammaticality judgments. Press SPACE BAR to continue.

Slide 2:

You will listen to 60 sentences. Your task is: (1) to judge whether the sentences are formed in accordance with the above-mentioned system or not, and (2) to indicate how confident you are about your decision. Rely on your intuition when judging the well-formedness of the sentences. We will begin with a short practice session. Press SPACE BAR to continue.

Slide 3:

In the following practice session you will listen to four scrambled sentences. For each sentence you will be asked three questions: Question 1: Was the sentence grammatical? In other words, did the sentence follow the same word order rules or not? Press / if you think the sentence was grammatical. Press Z you think the sentence was ungrammatical. Press SPACE BAR to continue.

Slide 4:

Question 2: How confident are you about your decision? Press 0 if you believe you have no confidence in your decision, i.e. you are guessing. Press 5 if you are somewhat confident in your decision. Press 9 if you are very confident in your decision. Press SPACE BAR to continue.

Slide 5:

Question 3: What is the basis of your grammaticality judgment?

Press G if you believe your judgment was based on a guess.
= You might as well have flipped a coin.

Press I if you believe your judgment was based on intuition.
= You are confident in your decision but don't know why it's right.

Press M if you believe you relied on memory for previous sentences.
= You recall part or the entire sentence from the training phase.

Press R if you believe you applied a rule in judging the sentence.
= You have acquired a rule during training and could verbalize it.

Press SPACE BAR to begin the practice session.

Slide 6:

You are now ready to start the last part of the experiment. It consists of 60 new scrambled sentences.

Grammaticality judgment
Press / if you think the sentence was grammatical.
Press Z if you think the sentence was ungrammatical.

Confidence judgment
Press 0 if you are not confident at all, i.e. you are guessing.
Press 5 if you are somewhat confident in your decision.
Press 9 if you are very confident in your decision.

Basis for grammaticality judgment
Press G if you believe your judgment was based on a guess.
Press I if you believe your judgment was based on intuition.
Press M if you believe you relied on memory for previous sentences.
Press R if you believe you applied a rule in judging the sentence.

Press SPACE BAR to start.

Slide 7:

Congratulations! You have reached the END OF THE EXPERIMENT.

IMPORTANT:
Before leaving, please take a minute to fill in the questionnaire on your right.

Thank you very much for participating!

2.3. Experiment 4

2.3.1. Training phase

Slide 1:

The following experiment was designed to test your memory for sound sequences. In the experiment, you will be presented with 90 syllable sequences. Your task is to repeat each sequence once the prompt appears on the screen. Step 1: Listen attentively to the syllable sequence. Step 2: Wait for prompt „Please repeat now.' to appear on the screen. Step 3: Repeat the syllable sequence aloud. We will evaluate your performance by measuring how accurate you were in your repetitions. Press SPACE BAR to continue.

Slide 2:

We will begin with a short practice session. You will listen to four syllable sequences. Your task is to repeat each sequence after it has been played. Please wait for the prompt before you repeat the sequence. You can play the sequence again if you feel you are not able to repeat it correctly. Press SPACE BAR to begin the practice session.

Slide 3:

You are now ready to start the experiment. You will listen to 90 sequences in total. Remember: Step 1: Listen attentively to the syllable sequence. Step 2: Wait for the prompt „Please repeat now.' to appear on the screen. Step 3: Repeat the syllable sequence aloud. Press SPACE BAR to begin the experiment.

2.3.2. Testing phase

Experimental group version

Slide 1:

The arrangement of the syllables in the previous sequences was not arbitrary but determined by a complex rule-system (a ‚grammar'). For example, this grammar might state that syllable X must follow syllable Y but not be preceded by syllable Z. In the last part of the experiment, you will listen to 60 syllable sequences. 30 sequences were generated by means of the same rule-system as in the previous part of the experiment. These sequences are called ‚grammatical'. 30 sequences were generated randomly, i.e. they do NOT confirm to our rule-system. These sequences are called „ungrammatical'. Your task is to differentiate sequences that were created by means of the same grammar from those which weren't. We will begin with a short practice session. Press SPACE BAR to continue.

Slide 2:

In the following practice session, you will listen to four syllable sequences. For each sequence, you will be asked three questions: Question 1: Was the sequence grammatical? In other words, was the order of the syllables determined by the rule-system (grammar) used in the first part of the experiment? Press / for YES, the sequence is grammatical. Press z for NO, the sequence is ungrammatical. Press SPACE BAR to continue.

Slide 3:

Question 2: How confident are you about your decision? Press 0 if you believe you have no confidence in your decision, i.e. you are guessing. Press 5 if you are somewhat confident in your decision. Press 9 if you are very confident in your decision. Press SPACE BAR to continue.

Slide 4:

Question 3: What is the basis of your grammaticality judgment? Press G for GUESS. "Your judgment was based on a guess. You might as well have flipped a coin." Press I for INTUITION: "Your judgment was based on intuition. You are confident in your decision but don't know why it's right." Press M for MEMORY: "You relied on memory for previous sequences. You recall part or the entire sequence from the training phase."

Press R for RULE:
"You applied a rule in judging the sentence. You have acquired a rule during training and could verbalize it."

Press SPACE BAR to begin the practice session.

Slide 5:

You are now ready to start the last part of the experiment.

It consists of 60 syllable sequences:
30 are grammatical, 30 are ungrammatical.

Was the sequence grammatical?

Press / for YES, the sequence is grammatical.
Press Z for NO, the sequence is ungrammatical.

How confident are you in your decision?

Press 0 if you have no confidence in your decision, i.e. you are guessing.
Press 5 if you are somewhat confident in your decision.
Press 9 if you are very confident in your decision.

What is the basis of your grammaticality judgment?

Press G for GUESS
Press I for INTUITION
Press M for MEMORY
Press R for RULE

Press SPACE BAR to start.

Slide 6:

Congratulations! You have reached the END OF THE EXPERIMENT.

IMPORTANT:
Before leaving, please take a minute to fill in the questionnaire on your right.

Thank you very much for participating!

Control group version

Slide 1:

The arrangement of the syllables in the following sequences is determined by a complex rule-system (a ‚grammar').

For example, this grammar might state that syllable X must follow syllable Y but not be preceded by syllable Z.

You will listen to 60 syllable sequences.

30 sequences were generated by means of the same rule-system as in the previous part of the experiment. These sequences are called ‚grammatical'.

30 sequences were generated randomly, i.e. they do NOT confirm to our rule-system. These sequences are called „ungrammatical'.

Your task is to differentiate sequences that were created by means of the same grammar from those which weren't.

We will begin with a short practice session.

Press SPACE BAR to continue.

Slide 2:

In the following practice session, you will listen to four syllable sequences.

For each sequence, you will be asked three questions:

Question 1:

Was the sequence grammatical?

In other words, was the order of the syllables determined by the rule-system (grammar) used in the first part of the experiment?

Press / for YES, the sequence is grammatical.

Press z for NO, the sequence is ungrammatical.

Press SPACE BAR to continue.

Slide 3:

Question 2:

How confident are you about your decision?

Press 0 if you believe you have no confidence in your decision, i.e. you are guessing.

Press 5 if you are somewhat confident in your decision.

Press 9 if you are very confident in your decision.

Press SPACE BAR to continue.

Slide 4:

Question 3:

What is the basis of your grammaticality judgment?

Press G for GUESS.
"Your judgment was based on a guess. You might as well have flipped a coin."

Press I for INTUITION:
"Your judgment was based on intuition. You are confident in your decision but don't know why it's right."

Press M for MEMORY:
"You relied on memory for previous sequences. You recall part or the entire sequence from the training phase."

Press R for RULE:
"You applied a rule in judging the sentence. You have acquired a rule during training and could verbalize it."

Press SPACE BAR to begin the practice session.

Slide 5:

You are now ready to start the last part of the experiment.

It consists of 60 syllable sequences:
30 are grammatical, 30 are ungrammatical.

Was the sequence grammatical?

Press / for YES, the sequence is grammatical.
Press Z for NO, the sequence is ungrammatical.

How confident are you in your decision?

Press 0 if you have no confidence in your decision, i.e. you are guessing.
Press 5 if you are somewhat confident in your decision.
Press 9 if you are very confident in your decision.

What is the basis of your grammaticality judgment?

Press G for GUESS
Press I for INTUITION
Press M for MEMORY
Press R for RULE

Press SPACE BAR to start.

Slide 6:

Congratulations! You have reached the END OF THE EXPERIMENT.

IMPORTANT:
Before leaving, please take a minute to fill in the questionnaire on your right.

Thank you very much for participating!

2.4. Experiment 5

2.4.1. Training phase

Slide 1:

In the following experiment on sentence comprehension you will listen to 120 scrambled sentences.

Your task is:

(1) to listen to each sentence,
(2) to report the sentence type, and
(3) to judge the plausibility of the statements made.

We will begin with a short practice session that shows you how to do this.

Press SPACE BAR to continue.

Slide 2:

Sentences can consist of three types. Here are a few English examples:

Type 1:
Main Clause only
This morning John bought an apple in the supermarket.

Type 2:

Main Clause	> (Subordinate Clause)
Mike mentioned to his friends	(that he planted a tree in the garden.)
John bought a sandwich	(because he was hungry.)

Type 3:

(Subordinate Clause)	> Main Clause
(Because he was hungry,)	John bought a sandwich.
(That he planted a tree in the garden,)	Mike mentioned to his friends.

Press SPACE BAR.

Slide 3:

In the following practice session you will listen to four scrambled sentences.

Your task is

(1) to report the sentence type:

Press 1 for 'main clause only'.
Press 2 for 'main > subordinate'.
Press 3 for 'subordinate > main'.

(2) to judge the plausibility of the statement made in the sentence:

Press / if you think the statement was plausible.
Press z if you think the statement was not plausible.

Press SPACE BAR to continue.

Slide 4:

You are now ready to start the experiment.

You will listen to 120 scrambled sentences in total.

Press SPACE BAR to begin.

2.4.2. Training phase

Slide 1:

The word order of the previous sentences was not arbitrary but determined by a complex rule-system instead. In other words, we used certain rules to determine what word goes where. In the second part of the experiment, you will listen to 60 new sentences. 30 sentences were generated by means of the same rule-system as in the previous part of the experiment. These sequences are called ‚grammatical'. 30 sequences were generated randomly, i.e. they do NOT confirm to our rule-system. These sequences are called ‚ungrammatical'. This time, your task is: (1) to judge whether the new sentences were created by following the above-mentioned rule-system or not, and (2) to indicate how confident you are in your decision? We will begin with a short practice session. Press SPACE BAR to continue.

Slide 2:

In the following practice session, you will listen to four sentences. For each sentence, you will be asked three questions: Question 1: Was the sequence grammatical? In other words, was the word order determined by the same rule-system ('grammar') used in the first part of the experiment? Press / if you think the answer is YES. Press z if you think the answer is NO. Press SPACE BAR to continue.

Slide 3:

Question 2:

How confident are you in your decision?

Press 0 if you believe you have no confidence in your decision, i.e. you are guessing.

Press 5 if you are somewhat confident in your decision.

Press 9 if you are very confident in your decision.

Press SPACE BAR to continue.

Slide 4:

Question 3:

What is the basis of your grammaticality judgment?
Press G for GUESS:
"Your judgment was based on a guess. You might as well have flipped a coin."

Press I for INTUITION:
"Your judgment was based on intuition. You are confident in your decision but don't know why it's right.

Press M for MEMORY:
"You relied on memory for previous sequences in your judgment. You recall part or the entire sequence from the training phase."

Press R for RULE-KNOWLEDGE:
"You applied a rule when judging the sentence. You have acquired a rule in the first part of the experiment and could verbalize it."

Press SPACE BAR to begin the practice session.

Slide 5:

You are now ready to start the last part of the experiment. It consists of 60 new sentences:
30 sentences follow the word-order rules of our 'grammar', 30 sentences do not.
Please remember that your task involves endorsing (/) correct sequences and rejecting (z) incorrect sequences. Try to discriminate between the two types of sequences.

Grammaticality Judgment:
Was the sequence grammatical? Does the word-order feel familiar?
Press / if you think the answer is YES.
Press z if you think the answer is NO.

Confidence Judgment:
How confident are you in your decision?

Press 0 if you are not confident at all: you are guessing.
Press 5 if you are somewhat confident in your decision.
Press 9 if you are very confident in your decision.

Source Judgment:
What is the basis of your grammaticality judgment?
Press G for judgments based on a GUESS.
Press I for judgments based on INTUITION.
Press M for judgments based on MEMORY for previous sequences.
Press R for judgments based on RULE-KNOWLEDGE.

Press SPACE BAR to start.

Slide 6:

Congratulations! You have reached the END OF THE EXPERIMENT.

IMPORTANT:
Before leaving, please take a minute to fill in the questionnaire on your right.

Thank you very much for participating!

2.5. Experiment 6

2.5.1. Training phase

Slide 1:

In the following experiment you will listen to scrambled sentences.
The word order in these sentences is not arbitrary but determined by a complex rule-system instead.

In the first part of the experiment, your task is to listen to 120 sentences and to pay attention to the word order of the sentences. Half of these sentences are nonsensical. Please disregard the meaning of the sentences and try to figure out what rules determine the word order.

In the second part of the experiment, you will then be tested on your knowledge of the rules that determine the placement of words.

This is done by means of a grammaticality judgment task: you are presented with the new scrambled sentences and will be asked whether or not the word order is arranged by the same rules as the 120 sentences in part 1.

We will begin with a short practice session.

Press SPACE BAR to continue.

Slide 2:

In the following practice session you will listen to four scrambled sentences. Listen to each sentence carefully and pay attention to the word order. Remember: You will be tested on the rules that determine the word order in these sentences. Press SPACE BAR to begin the practice session.

Slide 3:

You are now ready to start the experiment. You will listen to 120 scrambled sentences, half of which are nonsensical. Please disregard the meaning of the sentences. Remember: Your task is (1) to pay attention to the word order and (2) to figure out the rules that govern the placement of words. Press SPACE BAR to begin the experiment.

2.5.2. Testing phase

Slide 1:

As has been mentioned at the beginning, the word order of the previous sentences was not arbitrary but followed a complex rule-system. In the second (and final) part of the experiment, you will listen to 60 new sentences. This time, your task is: (3) to judge whether the new sentences follow the same word order rules as the sentences in the first part of the experiment, and (4) to indicate how confident you are about your decision. We will begin with a short practice session. Press SPACE BAR to continue.

Slide 2:

In the following practice session you will listen to four scrambled sentences. For each sentence you will be asked three questions: Question 1: Was the sentence grammatical? In other words, did the sentence follow the same word order rules or not? Press / if you think the sentence was grammatical. Press z you think the sentence was ungrammatical. Press SPACE BAR to continue.

Slide 3:

Question 2: How confident are you about your decision? Press 0 if you believe you have no confidence in your decision, i.e. you are guessing. Press 5 if you are somewhat confident in your decision. Press 9 if you are very confident in your decision. Press SPACE BAR to continue.

Slide 4:

Question 3: What is the basis of your grammaticality judgment? Press G if you believe your judgment was based on a guess. = You might as well have flipped a coin. Press I if you believe your judgment was based on intuition. = You are confident in your decision but don't know why it's right. Press M if you believe you relied on memory for previous sentences. = You recall part or the entire sentence from the training phase. Press R if you believe you applied a rule in judging the sentence. = You have acquired a rule during training and could verbalize it. Press SPACE BAR to begin the practice session.

Slide 5:

You are now ready to start the last part of the experiment. It consists of 60 new scrambled sentences. Grammaticality judgment Press / if you think the sentence was grammatical. Press Z if you think the sentence was ungrammatical. Confidence judgment Press 0 if you are not confident at all, i.e. you are guessing. Press 5 if you are somewhat confident in your decision. Press 9 if you are very confident in your decision. Basis for grammaticality judgment Press G if you believe your judgment was based on a guess. Press I if you believe your judgment was based on intuition. Press M if you believe you relied on memory for previous sentences. Press R if you believe you applied a rule in judging the sentence. Press SPACE BAR to start.

Slide 6:

Congratulations! You have reached the END OF THE EXPERIMENT. IMPORTANT: Before leaving, please take a minute to fill in the questionnaire on your right. Thank you very much for participating!

3. Debriefing questionnaires

3.1. Experiment 1, 2, 3, 5 and 6

Page 1:

In the course of this experiment, you have indicated for every sentence whether you thought it was **grammatical** or **ungrammatical**.

Reflecting on your performance, please estimate (in %) your **overall level of accuracy** in these judgments.

I estimate my overall level of accuracy to be …….. %.

While performing the tasks of the experiment, did you notice any particular rule or regularity? If yes, please indicate what you believe you have noticed.

As mentioned in the experiment, the scrambling of the sentences was not arbitrary. Instead, the word order in the sentences was based on a complex system.

Reflecting now specifically on the **placement of words** within the sentences, can you recall any specific rule or regularity?

Page 2: Personal information

1. What is your name?
2. Please indicate your age ……….., sex [m / f], and nationality ………..
3. What is your occupation?

   If you are a student,
   a) type of course (e.g., BA, PhD)
   b) subject(s)
   c) year
4. What is your language background? Please complete the following:

   Native language(s)

Foreign language(s)
(include proficiency: Beginner, intermediate, upper intermediate, advanced)

3.2. Experiment 4

Page 1:
In the course of this experiment, you have indicated for every sequence whether you thought it was **grammatical** or **ungrammatical**.

Reflecting on your performance, please estimate (in %) your **overall level of accuracy** in these judgments.

I estimate my overall level of accuracy to be …….. %.

While performing the tasks of the experiment, did you notice any particular rule or regularity? If yes, please indicate what you believe you have noticed.

As mentioned in the experiment, the order of the syllables was not arbitrary but determined by a complex system.

Reflecting now specifically on the **placement of syllables** within the sequences, can you recall any specific rule or regularity?

Page 2: Personal information

1. What is your name?
2. Please indicate your age ……….., sex [m / f], and nationality ………..
3. What is your occupation?
   If you are a student, a) type of course (e.g., BA, PhD)
   b) subject(s)
   c) year

4. What is your language background? Please complete the following:

Native language(s)

Foreign language(s)

(include proficiency: Beginner, intermediate, upper intermediate, advanced)

4. Verbal reports

4.1. Experiment 1

Table 43

Verbal Reports of Experimental Group

Subject	Report
01	I think the verb was (usually) sentence-final in embedded clauses. In main clauses it was before the subject (on reflection it may have always been in second position but I didn't think of that until just now!)
02	All in the 3rd person. All in the past tense. All passive. The person's name or the time (today/yesterday) is at the beginning of the sentence, then why they did something, then what they did. The verb is before the subject, e.g. yesterday, noticed, John,...
03	I started to notice when the verbs came at the end of the sentence. I also noticed the placement of adverbs and how that contributed to/detracted from the clarity of the sentence. It seemed to me that the subject-verb-order was reversed, i.e. "Recently reflected June that her children to England moved." Instead of "Recently June reflected that her children moved to England."
04	When adverb first, verb and subject inverted. Always implausible when fruit or planet mentioned.
05	Sentences ending in verbs seemed pretty grammatical, as did those with the "usually" info at the beginning.
06	V2 with other verbs at the end. Possibly verb-final on dependent clauses with "that" and "whether" and "because" Time, reference tended toward clause-initial position, but not strictly.
07	Adverbs initially, verb final, some V2: not sure when each occurred
08	Perhaps. They seemed to end with a verb about half the time. I just couldn't remember which (if either) was the form used in the first part.
09	I think that the rule was that the verb was placed at the end of the sentence. Verbs at the end of the sentence. Repetition of "usually", "recently"
10	The time period of usually came at the start of the sentence.

The verb often came at near the end.

Objects frequently came before rather than after verbs.

Often plausible sentences were made implausible by an out of place object – such as fruits or planets (contextually), more often than by verbs or nouns.

11 The sentences started with the adverb (usually, etc.).

This was followed by the auxiliary verb.

This was followed by subject and then the object.

The main verb was at the end of the clause.

So: Adverb – auxiliary – subject – object – verb. E.g., Usually has Jane a dog beaten.

12 The grammatical sentences that seemed to follow the rules of those in the first experiment seemed to be structured as so:

(Maybe adverbs here, at least temporal adverbs) – Subject – object – finite verb

13 No. Went by instinct didn't think about rules.

14 Where it was ungrammatical the subject sometimes came later in the sentence and was therefore confused with the object.

15 I think that the final clause should have been a noun (the object of the sentence), but I don't think my results reflect this as I didn't realise for a while. Also, the time clause (usually, today) opened the sentence. The second test seemed to include more phrases like "For a while" I hadn't noticed in the first sets of sentences.

16 Verb often in second position in sentence and at end of sentence.

Rest of sentence (i.e. subject, object, etc.) in same place.

Only verbs moved.

17 -

18 Inversions in indirect statements, e.g.

Jack remarked recently his son shoes bought (interchangeable).

I found other patterns harder to pick up. Similar inversions (I think) in shorter, simpler statements, but trickier to be sure whether I was right or not.

19 Adverb first, subject, object, then verb last.

Where there is a dependent clause word order near beginning of clause was more how I'd expect it compared to proper English grammar.

20 The time event occurred was always said first, i.e. "recently", "today", etc.

Adverb – subject – verb (aux verbs)

Quietly – Sheila – to the bathroom went

Table 44

Verbal Reports of Control Group

Subject	Report
01	The start of the sentence – in the past tense. The sentence made sense and .. more likely to be indicated as grammatically correct.
02	I was more likely to say that long jumbled sentences were ungrammatical than I would for shorter phrases simply because they were a lot harder to follow. Also: When there were two people in the sentence then it was more likely to be ungrammatical because it was more unclear what was meant. A single person, however, was easier to follow.
03	Tried not to rearrange. If you could understand without decryption the grammatical. Standard pattern, switches. Short sentences were easier to understand.
04	Judgment halfway through (more for short sentences).
05	Some sentences longer – with more subjects/objects verbs others shorter. Shorter sentences appeared more grammatical subject/object/verb reversal notices. Verb/object reversal was consistent.
06	That "have" was often used incorrectly with yesterday – you cannot say "yesterday I have" The verb often came at the end of the sentence. [Strategies: incorrect have sentences would be marked as ungrammatical. Verb-final marked as grammatical because it is easier to understand.)
07	Subject of sentence mentioned before subject. [Verbs too far away from subject, e.g. at end of sentence. Verb after object instead of subject.]
08	Subject + object switched places a lot. Verbs often (or always, I can't remember) were at the end. Prepositions generally at the beginning.
09	Tendency to start with indication of time – e.g. "recently". Verbs sent to the end. Two past participles rearranged sounded funny. Subject + main verb split up. Lots of were ... (maybe dative) clauses put in between.
10	When there was an intransitive verb, the displacement of the syntax made it appear far more ungrammatical. [When sentence finished with preposition it seemed more

	ungrammatical.] When the word order was Verb + Subject + Object it appeared much more ungrammatical than if it was Object + Subject + Verb. The position of adverb had no effect on grammaticality.
11	I felt it was most ungrammatical when words such as "has" were at the end of the sentence.
12	Verb at the end of the sentence at times. Subject, object + verb put in variety of ways/orders. Did the verb keep moving down the sentence?
13	Most sentences contained one major inversion; some contained two. Dating [?] the later sentences, the word "has" was often displaced to the end of the sentence. The main variable is the position of the verb.
14	Inversion of verb and its object, so it was subject + object + verb as opposed to subject + verb + object. Starting sentences with a time marker.
15	Grammatical sentences often started with a time (e.g., usually) but by no means was this always so. Grammatical sentences tended to have a natural break in the middle of the sentence – each sub-clause had a similar "feel"

4.2. Experiment 2

Table 45

Verbal Reports of Experimental Group

Subject	Report
01	2nd part of each sentence scrambled – verb often left until after the object. Word order more like foreign languages.
02	Around 1/3 of the way through, I noticed what I thought was the following rule: Description of time or causation must begin the sentence and a verb must

not end the sentence. It must end with a noun or adjectival phrase.

03 The sentences I assumed to be correct began with a temporal adverb, followed by a participle if present, then by the subject, then the object, with the main verb of the sentence at the end. I used this syntax to identify whether a sentence was grammatical or not.

Because I was not aware that the first section of the test, to an extent, represented a "new language". I was not sure whether the subject came first or not but decided to identify these sentences where it did as correct.

04 V NP X order

and

NP1 NP2 V X order

05 Verb at the end?

Times and subject nearer the beginning of the sentence?

06 It seemed like it was sections of the sentence that were switched around, rather than just words.

Not really I'm afraid.

07 Temporal adjective at the beginning of the sentence.

Subject – object – verb order.

08 Whether the verb was placed at the end or not.

Is it possible that subordinate clauses came first (i.e. beginning with since, recently, etc.) and, in some cases, the verbs were placed at the end of the sentence.

09 The pattern seemed to be time-verb-subject-object-other verb or something like that. I only thought of this in the 2nd (grammatical) part of the experiment.

I also noticed that phrases like "has gone" were split into "has ... gone".

10 Often in the first part the irregularity was an extra detail at the end of the (otherwise ok) sentence.

The part participle was often at the end of the sentence, meaning the rest of the sentence couldn't be judged as accurate or not until the end.

Nouns or past participles at end of sentence.

Subject nearly always at beginning of sentence.

Detail/extra description of verb action came just before the verb (i.e. provenance, direction).

11	Verbs were sometimes at the end of the sentence, or very close to the beginning. And also "frequency", e.g. usually, sometimes, recently placed at the beginning of the sentence.
12	Sentence beginnings sometimes corresponded (e.g. with "yesterday")
13	Verb placement, objects before noun, (adv – subj – verb) Has/have at end were obviously wrong. Subject after verb sometimes wrong. Adv and subj before verb (Recently Susan bought).
14	Some sentences were grammatical. Others were random. No strategy, but thought "I have heard a type of sentence like this before"
15	Focused on the position of the verb. Thought the verb should be at the end. Has/have should not be at the end. This sounded wrong.

4.3. Experiment 3

Table 46

Verbal Reports of Experimental Group

Subject	Report
01	Clause structure: Time/reason – subject – object – verb – adverb Some sentences had more than one clause.
02	It seemed that if the sentence normally would have ended in a verb + direct object, they were reversed. Clauses indicating time seemed to be oddly placed in the sentences. Reversing of the subject and verbs. Reversing the order of verbs and direct-objects at the ends of sentences.
03	The verb tended to placed after its direct object. The subject of the verb tended to precede the verb.
04	Complex systems invariably begin with a word like "since", "after", or "during", denoting time-periods. Followed by an event followed by a person's reaction to that event.
05	Seemed link lots of the grammatical sentences started with timing e.g. "recently", "yesterday".

06 During the second exercise it seemed that opposite grammatical constructions were used to contrast. There were also grammatically correct English sentences.

Time words came first, then verb, then subject, then object. "Recently, participated Julia in an experiment."

In a sentence with clauses the verb is relegated to the final position in the clause e.g. "Since her roommate fond of pears is, intends Julia to purchase some."

07 Yes, I think: Prepositional phrase at start of sentence. Main clause: subject never first (verb often before subject, but I don't know if this is a rule.)

Regularities: astronomical reference = implausible.

Repetition of names

Fruit of ten implausible.

Chris is suspicious of the government.

08 I judged the sentences that began with a verb to be ungrammatical.

Sentences didn't begin with verb? Sentences sometimes ended with a verb. Subject verb order always disrupted? Sentences often began with a proper noun.

09 There seemed to be an inversion in the placing of the verb and the object of the sentence; usually the verb was being placed after the subject.

Verb generally placed after the subject of the verb.

Conjunctions + clauses of time very often placed at the beginning of sentences.

10 Yes – the direct object was often placed before the verb but its correctness depended on other phrases.

It seemed ok if the time statement was in position one before the verb.

Certain prepositional phrases between the verb and direct object ruined the grammar.

11 Sentences were implausible if they referred to things such as fruit in certain contexts, e.g. laminating the recipe card with peaches. References to planets and aliens meant the sentences were likely to be implausible.

Sentence order changed so it was more like other languages: subject – object – verb rather than English: subject – verb – object.

12	The verb in the grammatical one was at the end of the sentence – subject object verb.
13	In 1 clause sentences, verbs came before nouns or direct objects, and nouns before direct object, prepositions or qualifiers first. "In the morning, ate Joe a large breakfast." If 2nd clause, verb came at the end. Preposition/qualifier, verb, noun, direct object (1st clause), direct object, prep/qual, noun, verb (2nd clause).
14	Normal rules of grammar seemed to be working in reverse – verbs/nouns were not in the places they usually are. Verbs come after nouns, e.g. his son a goal scored, a car bought, to the moon went.
15	Yes, frequently the past tense verb and direct object of phrase were switched, phrases noting time were placed in the centre of sentences rather than the beginning, subjects and verb order often reversed.

Table 47

Verbal Reports of Control Group

Subject	Report
01	Most sentences were ungrammatical. This was often due to wrong order of nouns and verbs. Also some particles missed out.
02	Often the order of verbs was inverted, or the action and the subject would be split up (by modifying details, etc.)
03	If the sentence was one main clause, the verb was before the subject, "bought David his dinner", and the subject before the object. Often a qualifying phrase "for a long time" was delayed to the end of the sentence – confusingly. But if there was a subordinate clause, the verb was delayed to the end of that clause "Because the children after dinner sweets liked".
04	Often the verb came before the subject and object. Just as often order was subject – object – verb (like in Latin).
05	Thought verb-subject constructions (Discussed Jennifer yesterday...) were wrong. Thought odd sentence structures such as "Jennifer in a blue car

went...” were correct.

06 In general, sentences seemed to be of the form “Today, Sam the test took.” rather than “Today Sam took the test.”

Only that above: but I also remember some instances where that didn’t seem to happen, so it might not count as a rule.

07 The rule I noticed whether the beginning of the sentence bound the sentence in entirety or not e.g. “When ... decided.” – somehow it “sounded” better when a time stated in the beginning finds an action in the end.

Also, a pile up of nouns or verbs were often marked ungrammatical by me. However, both the rules have been not applied/misapplied too in certain cases.

The only difference I noticed was that the sentence was divided into 3 parts – a time, a character and an action, divided by a comma or a pause. While most sentence started with the time factor “After dinner, when Paul...” – the action and character were shifted around so that in the re-placement, some sentences finished with an action, or a character.

08 Sometimes nouns/subjects and verb were reversed.

09 Misplacing of the verb causes a paradox where nouns become verbs etc.

He for she and she for he.

A verb before a name: considered John etc.

A noun before a verb where we would expect it another way round.

A verb to finish or start a sentence which is not an infinitive.

10 An absence of definite articles.

Reversal of the verb and the subject.

11 There were a lot of sentences based on different “time-settings” e.g. “recently”, “some time ago” or just plain past tense sentences.

Mostly word order of verb/subject was reversed, although occasionally it reversed to standard subject/verb syntax.

There were a lot of subclauses at the beginning of sentences.

12 -

13 -

14 -

15 -

4.4. Experiment 4

Table 48

Verbal Reports of Experimental Group

Subject	Report
01	I thought ko couldn't occur at the end papu and pupa seemed to be okay sequences at the end. Tete pupu seemed wrong. Te could occur at the end. Ki sa ge seemed fine at the beginning.
02	Possibly that "fi" cannot precede "ki"; that "soh" cannot precede "koh" No word was longer than three syllables; two tri-syllabic words did not seem to follow each other.
03	Rules/regularities noticed: the nr of syllables was consistent among longer sentences and among shorter sentences? Certain sounds just "sounded right" or wrong. Not sure it is complete recollection but I think that certain words didn't form the end of the sentence otherwise didn't sound right. To expand on this, certain words sound "correct" in certain places in the sentences, whereas others didn't. Couldn't tell you which words though.
04	The sequence which sounds like "corsair" never appears in the first part of the sentence. 4 syllable sequences always at the start of the sentence. Many sequences 3,2,2 and 3,2,3 and 3,2. Some syllable pairs only in certain phrase positions.
05	"Pa" often followed by "tuki" or "teko" "Racair" or "kissair" often fell in the middle of phrases. Last word of phrase often "pappu" or "pateh".
06	"t" and "p" sounds were often together, though in no specific order when the "oo" vowel sound was present When there was a different vowel, "p" was usually first "k" sound followed by "ee" sound Words can end on any vowel.
07	"Sa ko fi" is a beginning sequence. The final syllable is often a "k" sound with a vowel.

	The last few syllables use “p”, “t” and “k” sounds.
08	None that I could explain.
	By repeating the phrases in the first part of the experiment.
	I had an intuitive feeling about what sounded “right”, especially when I noticed a mistake in the way that I repeated the syllable sequences.
09	Not particularly – any ideas of rules got scrambled in my head by the 2nd part.
10	When remembering for repeating, I broke them into 2 or 3 parts.
	Sa kik fi
	Ra sa / rah seh etc.
	Pu pa / pa te
	In the latter part of test, I tried to remember these [1-3]/apply them.
11	Few that I could write down, though some patterns familiar.
	Consonantal patterns easier to remember, e.g. (f-)(c-)(s-)(p-)(t-) seemed to come up a lot.
12	Variation between short and long sentences – easier sentences were the short cues.
	In long sentences “ra se” was in the middle.
	Rep. of one sound followed by a different sound, e.g. cosa, cose.
	Rep. of “ru si pa ki” at the start of sentence.
	Patuki/patuko – certain sequences repeated to form what sounded like words.
13	None.
14	Beginning: Fi Keeser, See Keefer
	Middle: ras – er or ras – a
	End: tuko, tuk-ee
15	No T sounds together in the first part.
	No “ee” sounds twice in the first part of a sequence.

4.5. Experiment 5

Table 49

Verbal Reports of Experimental Group

Subject	Report
01	The order was often a “time indicator” i.e. recently followed by verb then noun.
	The object and subject often came before the verb.

02 In the first half the same kinds of sentences (1, 2 or 3) tended to come in groups.

Time words come at the beginning of sentences, subordinate clauses come first in main clauses (2nd) word order is verb, subject, object.

03 Verb often placed at the end of main clauses: subject – object – verb

04 Verbs were not placed directly after subject.

Verbs were placed at the end of a clause.

2 verbs were not normally placed next to each other.

05 It seems that main clauses have the sequence verb, subject, object and subordinate clauses follow the sequence subject, object, verb.

06 Type 1 sentences are of the form:

adverb – verb – subject – object

Other rules were harder to work out.

07 Syntactical "irregularities"

Verb before subject

Object before verb

08 Since (name) did this / in the afternoon / (name) / this / did.

I think subordinate and main clauses (I think) were easier to spot because they followed a more obvious pattern.

09 Adv –S – O – (adv) – V

10 The structure SOV recurred, passive verbs instead of active ones, big delay before the verb.

11 During training very rarely (I can only remember once) was a proper noun used, but during the formal testing proper nouns were used frequently.

No. "By ear" it was able to hear scramblings that sounded or seemed familiar but that familiarity is difficult to articulate.

12 I think I noticed that subject and object were always inverted.

I also noticed that subordinate clauses beginning in "that" were used in sentence-type "2", while subordinate clauses telling time or reason (since, when, after) were used in sentence-type "3".

I think, as mentioned before, that subject and verb were inverted; and adverbs tended to be at the start of sentences.

13 In main clauses, the sequence seems to be: temporal preposition > main

	verb > subject > dir object > individual objects. Subordinate clauses have the verb last, with the subject first, directly after any causal propositions, with all other elements between the subject and the verb.
14	In sentences that started with the subordinate clause followed by the main clause, the verbs always followed each other. The “grammatical” sentences did not start with the verb. “Since” indicated subordinate clause followed by main clause; “after” performed the same function or denoted time in the main clause sentences. Verbs were generally placed in the middle or at the end of the sentence. Sentences with only the main clause usually began with a reference to time.
15	Simple clause: (prep, sub) S, O, V Compound sentence: [ ] S, dir O, V, [ ] V, ind O

4.6.Experiment 6

Table 50

Verbal Reports of Experimental Group

Subject	Report
01	Subject + verb usually reversed - "bought Joanna" Verb at the end of a phrase - placed before the subject
02	1. Subordinate clauses about time occurred first in a lot of cases e.g. “when...” 2. Sometimes these clauses ended with a verb, and the main clause then started with a verb. 3. The sentences would often end with verbs. 4. The sentences would often begin with “Because...” 5. A lot of the sentences began with time e.g. “Recently”, etc.
03	Often the order of the verb and name were reversed e.g. “Said Emma” instead of “Emma said”. Verbs often delayed until the end of the sentence/phrase. Object often delayed until end of phrase.

04 Since/because - subj - obj - verb, verb - subj - obj

Usually/today -verb -subj - obj -with/etc.

Subj - verb - today (that ... verb)

05 Instead of the usual subject – verb – object order of English (e.g., I killed him) the order is reversed to ... either: subject – object – verb (e.g., Joanna melons chopped) or verb – subject – object (e.g., Chopped Joanna melons).

06 Object - before - subject

Verb -after - subject

07 Adverbial clauses tended to be placed at the beginnings or ends of sentences

Dependent clauses were often found at the beginning, opening with some detail about time or spatial concerns

Verb and object seemed to be inverted, although when transitive or in the past, a final verb would often be found at the very end of the sentence

The sentence often seemed to demonstrate an overarching balance, with the cluster of proper nouns and linked verb in the middle.

08 I think that if the temporal statement – e.g., "In June", "Yesterday", etc., came first, then verb governing the subject (and the object if it had one) came before the subject or object, e.g. "Ate Susan" rather than "Susan ate"? If not, the object's verb could still be if the "wrong" (in terms of English grammar) case, but not the verb governing the subject. I remained somewhat unclear as to the relevance of prepositions, and whether the placing of conjunctions like "because" had a role. I think verbs governing the subject did <u>not</u> ever go to the end of the sentence – only either directly before or after the subject.

09 If the sentence begins with an adverbial, e.g. Yesterday – then word order is: Yesterday – verb – subject – object

If the adverbial is ..., then order changes to subject – verb – object – yesterday

In a because or that sentence the main clause order is verb, subject, object and the because or that clause is: because subject – object – verb

10 Rules

1. If there was one verb in sentence, then the order could be:

Subject - verb - time (interchangeable), direct object - indirect object or modifier

2. If sentence had 2 verbs and indicated consequence then the order was:

Because/since/after – subject – object – verb, verb 2 – subject 2- object 2

3. If sentence had 2 verbs then order was:

Subject – verb – object (interchangeable), direct obj – indirect or modifier

General helpful rule: no verbs begin sentences

11 I thought I had a rule in which subject/verb could be either way round as long as the verb occurred next to its time clause but I didn't feel very confident with this ()

The "rule" above... That following a relative (that/which) an object usually occurred

12 Regular passive construction

Instead of subject verb object usually a verb subject object sentence structure, although a verb was sometimes noticed at the end of a sentence in part A.

The sentence always had the subject passively doing the verb: i.e., David jumped → jumped David: placing the subject after the verb.

The last element of each sentence usually referred to an adverbial clause describing the previous verb element, place or noun. The adverbial element at the end could also be an adverb, even if the previous elements was not a verb.

13 Yes – some sentences were simpler: i.e., the verb command work was switched around/"back to front", e.g. instead of "Jane wanted...", it said "...wanted...Jane" without mixing around the rest of the sentence. Other scrambles did this, but to greater degree.

Verb switches around.

4 blocks, phrases switch around.

When arbitrary, it was more noticeable that the answer to 2 above, the two words "inverted" + "Jane" would have been separated further apart in the sentence (i.e., not occur together).

Whereas, when rules/system scrambled sentences, you could understand that sentence blocks occurred together, just were switched around. But didn't change "meaning" of sentence so much.

14	Of course it is very hard to recognise rules when sentences are all jumbled up! But I think that based on the 1^{st} experiment it was possible to look for sentences where the 2 verbs were “the right way round” e.g. He phoned his wife after seeing the accident, or the wrong way round e.g. “he swam after he got cold.” If the order was the right way round, I marked the sentence as grammatically correct. The sentence where there was just one verb were much more difficult to judge, especially if the verb was placed at the end of the sentence. I ... ticked the “box” for rules based decision because I’m not a linguist, apart from the rule enunciated above, I couldn’t discern any particular subject (object) verb order.
15	Verbs don’t come at the beginning of sentences. After “when” or “that” the verb is sent to the end of the clause. After time/names (such as “after dinner”, etc.) the verb follows straight afterwards.

References

Abrams, M., & Reber, A. S. (1988). Implicit learning: Robustness in the face of psychiatric disorders. *Journal of Psycholinguistic Research, 17*, 425-439.

Adams, J. K. (1957). Laboratory studies of behaviour without awareness. *Psychological Bulletin, 54*, 383-405.

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259-302). Honolulu, HI: University of Hawai'i Press.

Allen, R., & Reber, A. S. (1980). Very long term memory for tacit knowledge. *Cognition, 8*, 175-185.

Allwood, C. M., Granhag, P. A., & Johansson, H. (2000). Realism in confidence judgments of performance based on implicit learning. *European Journal of Cognitive Psychology, 12*, 165-188.

Altmann, G., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology-Learning Memory and Cognition, 21*(4), 899-912.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by human infants. *Psychological Science, 8*, 321-324.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1999). Statistical learning in linguistic and non-linguistic domains. In B. MacWhinney (Ed.), *Emergentist approaches to language*. Hillsdale, NJ: Lawrence Erlbaum.

Berry, D. C. (1984). *Implicit and explicit knowledge in the control of complex systems.* Unpublished DPhil thesis, University of Oxford.

Berry, D. C. (1991). The role of action in implicit learning. *Quarterly Journal of Experimental Psychology, 43*, 881-906.

Berry, D. C. (1996). How implicit is implicit learning? In G. Underwood (Ed.), *Implicit cognition* (pp. 203-225). Oxford: Oxford University Press.

Berry, D. C., & Broadbent, D. (1984). On the relationship between task performance and associated verbalisable knowledge. *Quarterly Journal of Experimental Psychology, 36*, 209-231.

Berry, D. C., & Broadbent, D. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology, 79*, 251-272.

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

Lawrence Erlbaum.

Björkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception & Psychophysics, 54*(1), 75-81.

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. Sharwood Smith (Eds.), *Grammar and second language teaching: A book of readings* (pp. 19-30). Rowley, MA: Newbury House.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-69). Cambridge: Cambridge University Press.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bloom, P. (2001). Précis of *How Children Learn the Meanings of Words*. *Behavioral and Brain Sciences, 24*, 1095-1103.

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences, 2*(2), 67-73.

Bohnacker, U. (2006). When Swedes begin to learn German: from V2 to V2. *Second Language Research, 22*, 93-119.

Bonatti, L. L., Peña, M., Nesport, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*, 451–459.

Bowles, M. A. & Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology: Expanding the scope of investigation. *Studies in Second Language Acquisition, 27*, 415-440.

Braine, M. D. S. (1963). On learning the grammatical order of words. *Psychological Review, 70*(4), 323-348.

Braine, M. D. S. (1965). The insufficiency of a finite state model for verbal reconstructive memory. *Psychonomic Science, 2*, 291-292.

Braine, M. D. S. (1966). Learning the positions of words relative to a marker element. *Journal of Experimental Psychology, 72*, 532-540.

Broadbent, D. (1977). Levels, hierarchies and the locus of control. *Quarterly Journal of Experimental Psychology, 29*, 181-201.

Broadbent, D., & Aston, B. (1978). Human control of a simulated economic system. *Ergonomics, 21*, 1035-1043.

Broadbent, D., Fitzgerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology, 77*, 33-50.

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

Brooks, R. L., & Vokey, R. J. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 120*, 316-323.

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40*, 227-259.

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

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

Channon, S., Shanks, D. R., Johnstone, T., Vakili, K., Chin, J., & Sinclair, E. (2002). Is implicit learning spared in amnesia? Rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia, 40*, 2185-2197.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics, 36*(4), 387-395.

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

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory, IT-2*, 113-124.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N., & Miller, G. A. (1958). Finite state languages. *Information and Control, 1*, 91-112.

Clahsen, H., & Muysken, P. (1986). The availability of Universal Grammar to adult and child learners. *Second Language Research, 2*, 93-119.

Cleeremans, A., & McClelland, J. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235-253.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2*(10), 406-416.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology :*

Learning, Memory, and Cognition, 30, 1119-1130.

Curran, T. (2001). Implicit learning revealed by the method of opposition. *Trends in Cognitive Sciences, 5*, 503-504.

de Graaff, R. (1997). The *eXperanto* experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19*, 249-276.

DeKeyser, R. (1994). Implicit and explicit learning of L2 grammar: A pilot study. *TESOL Quarterly, 28*(1), 188-194.

DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17*, 379-410.

DeKeyser, R. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*, 195-221.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 313-348). Oxford: Blackwell.

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the Process Dissociation Procedure. *Psychonomic Bulletin & Review, 8*(2), 343-350.

Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies, 11*(9), 25-45.

Dienes, Z. (2008). Subjective measures of unconscious knowledge. *Progress in Brain Research, 168*, 49-64.

Dienes, Z. (in press). Artificial grammar learning. In P. Wilken, T. Bayne & A. Cleeremans (Eds.), *Oxford companion to consciousness*. Oxford: Oxford University Press.

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

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

Dienes, Z., & Fahey, R. (1995). The role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 848-862.

Dienes, Z., & Longuet-Higgins, C. (2004). Can musical transformations be implicitly learned? *Cognitive Science, 28*(4), 531-558.

Dienes, Z., & Perner, J. (1996). Implicit knowledge in people and connectionist networks. In G. Underwood (Ed.), *Implicit cognition* (pp. 227-256). Oxford: Oxford University

Press.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22*, 735-755.

Dienes, Z., & Perner, J. (2003). Unifying consciousness with explicit knowledge. In A. Cleeremans (Ed.), *The unity of consciousness* (pp. 214-232). Oxford: Oxford University Press.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research-Psychologische Forschung, 69*(5-6), 338-351.

Dienes, Z., Altmann, G., Gao, S. J., & Goode, A. (1995). The transfer of implicit knowledge across domains. *Language and Cognitive Processes, 10*(3-4), 363-367.

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*(5), 1322-1338.

Dienes, Z., Broadbent, D. E., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*, 875-882.

Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition, 13*, 431-469.

Doughty, C. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 256-310). Oxford: Blackwell.

Dulany, D. E. (1962). The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness* (pp. 102-129). Durham, NC: Duke University Press.

Dulany, D. E., Carlson, R., & Dewey, G. (1984). A case of syntactical learning and judgement: How concrete and how abstract? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 113*, 541-555.

Eisenberg, P. (1999). *Grundriss der deutschen Grammatik: Der Satz*. Stuttgart: Metzler.

Ellis, N. C. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology, 5*, 289-318.

Ellis, N. C. (1994). Implicit and explicit language learning: An overview. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1-31). London: Academic Press.

Ellis, N. C. (2007). Implicit and explicit knowledge about language. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 6: Knowledge about language, pp. 119-132): Springer.

Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, N. C., & Schmidt, R. (1998). Rules or association in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes, 13*(2-3), 307-336.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (2001). Introduction: Investigating form-focused instruction. In R. Ellis (Ed.), *Form-focused instruction and second language learning* (pp. 1-46). Oxford: Blackwell.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol: Multilingual Matters.

Erdelyi, M., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology, 6*, 159-171.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Eriksen, C. W. (1960). Discrimination and learning without awareness: a methodological survey and evaluation. *Psychological Review, 67*, 279-300.

Evans, S., & Azzopardi, P. (2007). Evaluation of a 'bias-free' measure of awareness. *Spatial Vision, 20*(1-2), 61-77.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 458-467.

Fotos, S. (1993). Consciousness-raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics, 14*, 385-407.

Fotos, S., & Ellis, R. (1991). Communicating about grammar: A task-based approach. *TESOL Quarterly, 25*, 87-112.

Fox, A. (2005). *The structure of German* (2nd ed.). Oxford: Oxford University Press.

Frensch, P. A. (1998). One concept, multiple meanings: On how to define the concept of implicit learning. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit* *learning* (pp. 47-104). Thousand Oaks, CA: Sage.

Frensch, P. A. (1998). One concept, multiple meanings: On how to define the concept of implicit learning. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 47-104). Thousand Oaks, CA: Sage.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science, 12*(1), 13-18.

Fu, Q. F., Fu, X. L., & Dienes, Z. (2008). Implicit sequence learning and conscious awareness. *Consciousness and Cognition, 17*, 185-202.

Gaillard, V., Vandenberghe, M., Destrebecqz, A., & Cleeremans, A. (2006). First- and third-person approaches in implicit learning research. *Consciousness and Cognition, 15*, 709-722.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course* (2nd ed.). Hillsdale, NJ.

Gómez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology, 33*(2), 154-207.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431-436.

Gómez, R. L. (2006). Dynamically guided learning. In Y. Munakata & M. Johnson (Eds.), *Attention & Performance XXI: Processes of change in brain and cognitive development* (pp. 87-100). Oxford: Oxford University Press.

Gómez, R. L. (2007). Statistical learning in infant language development. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*(5), 178-186.

Gómez, R. L., & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple associative knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(2), 396-410.

Gómez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition, 28*(2), 253-263.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Green, P., & Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied*

Linguistics, 13, 168-184.

Greenwald, A. G. (1992). New look 3: Unconscious cognition reclaimed. *American Psychologist, 47*, 766-779.

Greenwald, A. G., Klinger, M. R., & Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory & Cognition, 17*, 35-47.

Gregg, K. (1984). Krashen's Monitor and Occam's Razor. *Applied Linguistics, 5*, 79-100.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research, 21*, 1-23.

Hirsh-Pasek, K. A., & Golinkoff, R. M. (Eds.). (2006). *Action meets word: How children learn verbs*. Oxford: Oxford University Press.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: a survey and appraisal. *Behavioral and Brain Sciences, 9*, 1-23.

Huddleston, R., & Pullum, G. K. (2005). *A student's introduction to English grammar*. Cambridge: Cambridge University Press.

Hulstijn, J., & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning, 34*, 23-43.

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

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

Jansen, L. (2008). Acquisition of German word order in tutored learners: A cross-sectional study in a wider theoretical context. *Language Learning, 58*(1), 185-231.

Jansen, L. (2008). Acquisition of German word order in tutored learners: A cross-sectional study in a wider theoretical context. *Language Learning, 58(1)*, 185-231.

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 948-969.

Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology, 42*, 61-112.

Jusczyk, P., & Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review, 107*(2), 384-396.

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237*, 1445-1452.

Kihlstrom, J. F. (2004). Availability, accessibility, and subliminal perception. *Consciousness and Cognition, 13*, 92-100.

Kihlstrom, J. F., Barnhardt, T. M., & Tataryn, D. J. (1992). Implicit perception. In (pp. 17-54). New York: Guilford. In R. F. Bornstein & T. S. Pittman (Eds.), *Perception without awareness: Cognitive, clinical, and social perspectives* (pp. 17-54). New York: Guilford.

Kinder, A., & Assmann, A. (2000). Learning artificial grammars: No evidence for the acquisition of rules. *Memory & Cognition, 28*(8), 1321-1332.

Kinder, A., Shanks, D. R., Cock, J., & Tunney, R. J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology: General, 132*(4), 551-565.

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*(1), 169-181.

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

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Krashen, S. D. (1985). *The Input Hypothesis*. London: Longman.

Krashen, S. D. (1994). The Input Hypothesis and its rivals. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45-77). London: Academic Press.

Krashen, S. D., & Terrell, T. (1983). *The Natural Approach: Language acquisition in the classroom*. New York: Prentice Hall.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of non-local musical rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1417-1432.

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition, 10*, 294-340.

Leow, R. P. & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition, 26*, 35-57.

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning, 47*, 467-506.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware vs. unaware learners. *Studies in Second Language Acquisition, 22*, 557-584.

Leow, R. P. (2001). Attention, awareness and foreign language behavior. *Language Learning, 51*, 113-155.

Leung, J., & Williams, J. N. (2006). Implicit learning of form-meaning connections. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 465-470). Mahwah, NJ: Lawrence Erlbaum.

Lewicki, P. (1986). *Nonconscious social information processing*. New York: Academic Press.

Lewicki, P., Czyzewska, M., & Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 13*, 523-530.

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology, 20*, 24-37.

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of information. *American Psychologist, 47*, 796-801.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory for serial order and item information. *Journal of Experimental Child Psychology, 93*, 95-119.

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: Oxford University Press.

Marcel, A. J. (1983). Conscious and unconscious perception: experiments on visual masking and word recognition. *Cognitive Psychology, 15*, 197-237.

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

Mathews, R. C. (1990). Abstractness of implicit grammar knowledge: Comments on Perruchet and Pacteau's analysis of synthetic grammar learning. *Journal of Experimental Psychology: General, 119*, 412-416.

Mathews, R. C. (1991). The forgetting algorithm: How fragmentary knowledge of exemplars can abstract knowledge. *Journal of Experimental Psychology: General, 120*, 117-119.

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

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us? In A. H.-J. Do, L. Dominguez & A. Johansen (Eds.), *Proceedings of the 25th annual Boston University Conference on Language Development* (pp. 480–490). Somerville, MA: Cascadilla Press.

Maye, J., & Gómez, R. L. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183-206.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), 101-111.

McGeorge, P., & Burton, M. (1989). The effects of concurrent verbalisation on performance in a dynamic system. *British Journal of Psychology, 80*, 455-465.

McLaughlin, B. (1978). The Monitor model: Some methodological considerations. *Language Learning, 28*, 309-332.

Merikle, P. M., & Daneman, M. (1998). Psychological investigations of unconscious perception. *Journal of Consciousness Studies, 5*(1), 5-18.

Merikle, P. M., & Daneman, M. (2000). Conscious vs. unconscious perception. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 1295-1303). Cambridge, MA: MIT Press.

Merikle, P. M., & Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Consciousness and Cognition, 6*, 219-236.

Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition, 79*, 115-134.

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology, 56*, 484-491.

Moeser, S. D., & Bregman, A. S. (1972). The role of references in the acquisition of a miniature artificial language. *Journal of Verbal Learning and Verbal Behavior, 11*, 759-769.

Moeser, S. D., & Bregman, A. S. (1973). Imagery and language acquisition. *Journal of Verbal Learning and Verbal Behavior, 12*, 91-98.

Morgan, J. L. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior, 20*, 67-85.

Mori, K., & Moeser, S. D. (1983). The role of syntax markers and semantic referents in learning an artificial language. *Journal of Verbal Learning and Verbal Behavior, 22*, 701-718.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1-32.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417-528.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language, 53*, 225–237.

Parodi, T. (2000). Finiteness and verb placement in second language acquisition. *Second Language Research, 16*, 355-361.

Payne, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science, 5*, 241- 248.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604–607.

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

Perruchet, P. (2008). Implicit learning. In J. Byrne (Ed.), *Learning and memory: A comprehensive reference* (Vol. 2: Cognitive psychology of memory, pp. 597-621). Oxford: Elsevier.

Perruchet, P., & Amorim, M. A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 785-800.

Perruchet, P., & Pacteau, C. (1990). Synthetic Grammar Learning - Implicit Rule Abstraction or Explicit Fragmentary Knowledge. *Journal of Experimental Psychology-General, 119*(3), 264-275.

Perruchet, P., & Pacteau, C. (1991). Implicit Acquisition of Abstract Knowledge About

Artificial Grammar - Some Methodological and Conceptual Issues. *Journal of Experimental Psychology-General, 120*(1), 112-116.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233-238.

Perruchet, P., & Pacton, S. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 34*(1), 80-96.

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics, 17*, 97-119.

Perruchet, P., & Vinter, A. (1998). PARSER : A model for word segmentation. *Journal of Memory and Language, 39*, 246-263.

Perruchet, P., Gallego, J., & Savy, I. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognitive Psychology, 22*, 493-516.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psvchology: General, 133*, 573–583.

Pittner, K., & Berman, J. (2004). *Deutsche Syntax: Ein Arbeitsbuch*. Tübingen: Narr.

Pothos, E. M. (2007). Theories of Artificial Grammar Learning. *Psychological Bulletin, 133*(2), 227-244.

Reber, A. S. (1965). *Implicit learning of artificial grammars.* Unpublished MA thesis, Brown University.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 317-327.

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 of tacit knowledge. *Journal of Experimental Psychology: General, 118*, 219-235.

Reber, A. S. (1990). On the primacy of the implicit: Comment on Perruchet and Pacteau. *Journal of Experimental Psychology: General, 119*(3), 340-342.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.

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

Reber, A. S., & Lewis, S. (1977). Toward a theory of implicit learning: The analysis of the form and structure of a body of tacit knowledge. *Cognition, 5*, 333-361.

Reber, A. S., & Millward, R. B. (1968). Event observation in probability learning. *Journal of Experimental Psychology, 77*, 317-327.

Reber, A. S., & Millward, R. B. (1971). Even tracking in probability learning. *American Journal of Psychology, 84*, 85-99.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 17*(5), 888-896.

Rebuschat, P., & Williams, J. N. (2006). Dissociating implicit and explicit learning of natural language syntax. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2594). Mahwah, NJ: Lawrence Erlbaum.

Redington, M., Friend, M., & Chater, N. (1996). Confidence judgments, performance, and practice, in Artificial Grammar Learning.

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

Reingold, E. M., & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language, 5*(1), 10-28.

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

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition, 18*, 27-67.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*, 223-247.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, & Hernstadt (1991) and Knowlton & Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition, 27*, 235-268.

Roehr, K. (2008). Linguistic and metalinguistic categories in second language learning. *Cognitive Linguistics, 19*(1), 67-106.

Rohrmeier, M., Rebuschat, P., & Cross, I. (submitted). Incidental learning of melodic structure.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies, 49*(3), 329-359.

Rosenthal, D. M. (2000). Consciousness, content, and metacognitive judgments. *Consciousness and Cognition, 9*, 203-214.

Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Oxford University Press.

Rosenthal, D. M. (2008). Consciousness and its function. *Neuropsychologia, 46*(3), 829-840.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493–515.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172-196.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, Current Directions in Psychological Science*(12), 110-114.

Saffran, J. R., & Thiessen, E. D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Handbook of language development* (pp. 68-86). Cambridge, MA: Blackwell.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science, 274*, 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101-105.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition, 105*, 669-680.

Schachter, J. (1988). Second language acquisition and its relation to Universal Grammar. *Applied Linguistics, 9*, 219-235.

Schmidt, J. R., Crump, M. J. C., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition, 16*(2), 421-435.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165-209). London: Academic Press.

Schmidt, R. (Ed.). (1995). *Attention and awareness in foreign language learning*. Honolulu, HI: University of Hawai'i Press.

Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Scott, R., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34*, 1264-1288.

Scott, V. M. (1989). An empirical study of explicit and implicit teaching strategies in French. *Modern Language Journal, 73*, 14-22.

Scott, V. M. (1990). Explicit and implicit grammar teaching strategies: New empirical data. *French Review, 63*, 779-789.

Segal, E. M., & Halwes, T. G. (1965). Learning of letter pairs as a prototype of first language learning. *Psychonomic Science, 3*, 451-452.

Segal, E. M., & Halwes, T. G. (1966). The influence of frequency of exposure on the learning of a phrase structural grammar. *Psychonomic Science, 4*, 157-158.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115*(2), 163-196.

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

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: Relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences, 12*(8), 314-321.

Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge: Cambridge University Press.

Shanks, D. R. (2005). Implicit learning. In K. Lambert & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202-220). London: Sage.

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

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

Smith, K. H. (1966). Grammatical intrusions in the free recall of structured letter pairs. *Journal of Verbal Learning and Verbal Behavior, 5*, 447-454.

Squire, L., & Frambach, M. (1990). Cognitive skill learning in amnesia. *Psychobiology, 18*, 109-117.

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

Stadler, M. A., & Roediger, H. L., III. (1998). The question of awareness in research on implicit learning. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 105-132). Thousand Oaks, CA: Sage.

Stanley, W. B., Mathews, R. C., Buss, R. R., & Kotler-Cope, S. (1989). Insight without awareness: On the interaction of verbalisation, and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology, 41*, 553-577.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-81). Cambridge: Cambridge University Press.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*(1), 1-42.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27(2),* 173-204.

Townsend, D. J., & Bever, T. G. (2001). *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press.

Tunney, R. J., & Altmann, G. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1-12.

Tunney, R. J., & Altmann, G. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 614-639.

Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition, 31*(7), 1060-1071.

VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition, 18*, 495-510.

Vokey, J. R., & Higham, P. A. (2005). Abstract analogies and positive transfer in artificial grammar learning. *Canadian Journal of Experimental Psychology, 59*, 54-61.

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

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

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

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

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.

Wilkinson, L., & Shanks, D. R. (2004). Intentional control and implicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 354-369.

Williams, J. N. (2004). Implicit learning of form-meaning connections. In J. Williams, B. VanPatten, S. Rott & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 203-218). Mahwah, NJ: Lawrence Erlbaum.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*(2), 269-304.

Williams, J. N. (in prep). Implicit and explicit learning. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition*. San Diego, CA: Academic Press.

Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua, 118*(4), 522-553.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 15*, 1047-1060.

Winter, B., & Reber, A. S. (1994). Implicit learning and the acquisition of natural languages. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 115-145). London: Academic Press.

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

Yang, L. R., & Givón, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition. *Studies in Second Language Acquisition, 19*, 173-193.

Ziori, E., & Dienes, Z. (2008). How does prior knowledge affect implicit and explicit concept learning? *Quarterly Journal of Experimental Psychology, 61*, 601-624.