INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Willful control and the learning of complex systems

LeCompte, Denny Charles, Ph.D.
Rice University, 1992
RICE UNIVERSITY

WILLFUL CONTROL AND THE LEARNING OF COMPLEX SYSTEMS

by

DENNY C. LECOMPTÉ

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHILOSOPHY

APPROVED, DISSERTATION COMMITTEE

Michael J. Watkins, Chairman
Professor of Psychology

Henry L. Roediger, III
Professor of Psychology

Nancy J. Cooke
Assistant Professor of Psychology

Richard E. Grandy
Professor of Philosophy

Houston, Texas
April, 1992
ABSTRACT

Willful Control and the Learning of Complex Systems

by

Denny C. LeCompte

A series of experiments explores the role of willful control over the learning of complex, rule-governed systems such as language. Willful control is operationalized as the enhancement of learning by the deliberate application of cognitive strategies. Subjects studied strings of symbols generated according to the rules of a system known as an artificial grammar. They were then tested on their knowledge of the grammar's rules. In some conditions, symbols drawn from the vocabulary used in the grammar were inserted into each string, rendering the string somewhat ungrammatical. In the informed condition, the inserted symbols were identified; in the uninformed condition, the inserted symbols were not identified. If subjects were able to exert willful control by ignoring the inserted symbols, performance in the informed condition should exceed performance in the uninformed condition. If they were not able to exert willful control, there should no difference between the informed and uninformed conditions.

Evidence that the subjects had at least some degree of willful control was obtained in each of 10 experiments. Experiments 1, 2, 3, and 4 included a condition in which the subjects saw strings with no inserted symbols. Performance in this condition was consistently superior to performance in the informed condition, implying that the extent of willful control was less than complete.

Experiments 5, 6, 7, and 8 tested the hypothesis that the extent of willful
control would depend on the number of symbols inserted. Experiment 9 tested the hypothesis that exposure to strings with no ungrammatical symbols would enhance subjects' ability to ignore such symbols in subsequent strings. Finally, Experiment 10 tested the hypothesis that willful control would be increased if the inserted symbols did not come from the same vocabulary as the grammar. None of these hypotheses was supported.

The general conclusion from these experiments is that subjects can exert some degree of willful control over the learning of complex systems, but the extent of that control is substantially limited.
ACKNOWLEDGEMENTS

I would like to thank Nancy Cooke, Richard Grandy, and Roddy Roediger for their advice and guidance. I would also like to thank Stephanie LeCompte for her comments and for her support throughout this project. Finally, I would like to thank my mentor Mike Watkins for sharing his wisdom and his friendship, although no thanks could adequately repay him for all that he has done for me.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Complex systems</td>
<td>1</td>
</tr>
<tr>
<td>Approaches to the study of complex systems</td>
<td>2</td>
</tr>
<tr>
<td>Willful control</td>
<td>8</td>
</tr>
<tr>
<td>Willful Control in the Learning of Complex Systems</td>
<td>9</td>
</tr>
<tr>
<td>Willful Control and the Effect of Irrelevant Information</td>
<td>21</td>
</tr>
<tr>
<td>The Present Study</td>
<td>24</td>
</tr>
<tr>
<td>The Extent of Willful Control in Learning Artificial Grammars</td>
<td>27</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>27</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>31</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>35</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>42</td>
</tr>
<tr>
<td>Willful Control and the Amount of Irrelevant Information</td>
<td>48</td>
</tr>
<tr>
<td>Experiment 5</td>
<td>48</td>
</tr>
<tr>
<td>Experiment 6</td>
<td>53</td>
</tr>
<tr>
<td>Experiment 7</td>
<td>62</td>
</tr>
<tr>
<td>Experiment 8</td>
<td>64</td>
</tr>
<tr>
<td>Further Explorations of Effects on Willful Control</td>
<td>67</td>
</tr>
<tr>
<td>Experiment 9</td>
<td>67</td>
</tr>
<tr>
<td>Experiment 10</td>
<td>76</td>
</tr>
<tr>
<td>Summary and Conclusions</td>
<td>86</td>
</tr>
<tr>
<td>References</td>
<td>93</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Means for Experiments 1-4 ........................................... 32
Table 2: Statistical Analyses for Each of Five Response Criteria in Experiment 3 41
Table 3: Statistical Analyses for Each of Five Response Criteria in Experiment 4 47
Table 4: Mean Probability Correct (and Average Confidence Scores) for
Experiments 5, 6, 7, and 8 ................................................... 52
Table 5: Statistical Analyses for Each of Five Response Criteria in Experiment 5 55
Table 6: Statistical Analyses for Each of Five Response Criteria in Experiment 6 61
Table 7: Statistical Analyses for Each of Five Response Criteria in Experiment 8 69
Table 8: Mean Probability Correct (and Average Confidence Scores) for
Experiments 9 and 10 ..................................................... 73
Table 9: Statistical Analyses for Each of Five Response Criteria in Experiment 9 75
Table 10: Statistical Analyses for Each of Five Response Criteria in Experiment 10 81
Willful Control and the Learning of Complex Systems

Learning has long been at the center of inquiry in psychological research. For much of this century, experimental psychologists addressed themselves to this problem almost exclusively (Hilgard, 1987). Furthermore, many of the studies on learning focused on the acquisition of relatively simple information such as the direct association between a stimulus and a response (Hilgard, 1987). Not all phenomena can be explained in simple stimulus-response terms, however. Research has shown that the learning of certain complex tasks or tasks involving complex information can be qualitatively different from simple associative learning. One powerful example is the learning of language. Chomsky (1959) argued forcefully that a child’s acquisition of language could not be explained in terms of the build-up of simple associations. The information necessary to comprehend and produce spoken language is simply too rich and too complex.

Language is not the only kind of information that is too complex to be learned exclusively through the build-up of simple associations. A variety of other complex systems exist in which intricate rules specify the relationships among a limited set of units. What will be argued is that the learning of such complex systems deserves special attention because it seems to have unique properties which are not found in simple associative learning. First, the concept of a complex system will be elaborated. Next, examples of relevant research will be briefly discussed and several key issues outlined. One of these issues will then be singled out for more detailed study. Finally, a theoretical framework and an empirical methodology for addressing this issue will be delineated.

Complex Systems

As defined here, a complex system is a collection of units that relate to each other according to a set of rules which are relatively unintelligible from the system’s
output. The vagueness of this definition is necessary to capture the variety of things that might be classified as complex systems. The critical components of the definition are that there must exist a set of rules, and these rules must not be easily perceived.

For the purposes of this research, a further constraint will be placed on the discussion of complex systems. Complex systems can be learned in at least two ways. One way is through formal instruction. The other way is by observing the system at work or observing outputs of the system. Learning by observation is the focus of the present research.

One example of a complex system learned largely by observation is social interaction. An intricate set of rules underlies most of human social interaction, and because these rules are often culture-specific, it would be difficult to argue that they are in any way innate. People must therefore learn how to behave appropriately in social situations. Although some basic rules are taught explicitly by parents or schools, the vast majority of social amenities must be inferred from experience.

Other complex systems learned by observation include traffic patterns and computer systems. People may have a "feel" for traffic within their city, knowing "intuitively" that certain routes are likely to be heavier or lighter depending on conditions in other parts of the traffic system. Computer systems, or even specific computer software programs, are sometimes learned without reference to instruction manuals; instead, people learn contingency relations among various components of a computer program simply by interacting with the program (Carroll & Rosson, 1987).

These examples illustrate the importance of complex systems in daily life. They also illustrate that despite the complexity of these systems, people do learn to understand them. The question of interest in this study is the manner in which they achieve this understanding.

Approaches to the study of complex systems. Psychologists have, in recent years, devoted appreciable effort to studying how people learn complex systems. This
effort has involved a number of experimental methods, each of which approaches the issue in a slightly different manner. In one method, subjects learn to make rapid responses to a sequence of signals. The successive locations of these signals are determined by an inconspicuous set of rules. In a second method, subjects learn to control a computer simulation. The operation of the simulation is determined by relatively imperceptible rules which combine the subject's current input with previous states of the system. In a third method, subjects learn to recognize lawful patterns of symbols generated by a complicated rule-system. Learning in this instance is based solely on prior exposure to the output of the system.

The first method uses reaction time to measure subjects' rule-learning. In one study (Lewicki, Czyzewska, & Hoffman, 1987; see also Lewicki, Hill, & Bizot, 1988), each subject sat before a computer screen divided into four quadrants. Whenever the letter "X" appeared in one of these quadrants, the subject pressed a key corresponding to that quadrant. In each block of five trials, the choice of quadrant was random for the first two trials, but the choice for the remaining trials was completely determined, although not in an obvious way, by the locations of the first two trials. The subjects' reaction times decreased more rapidly for the three rule-governed trials than for the two random trials. For this method, the difference in reaction times between the random and rule-governed trials reveals the degree of learning.

The second method used to study the learning of complex systems has sometimes been referred to as the dynamic control task. The surface features of this task have taken forms as diverse as a national economy, a sugar factory, and a bus system, but the basic method remains the same: The subject provides an input (e.g., a number) for a complex system which has several levels (e.g., a system representing a sugar factory might have 10 different levels of production). This input is combined mathematically with the system's previous level to yield a new level. The subject's task
is to make the system reach and maintain a specific level. The method is illustrated by Berry and Broadbent (1984). In their study, each subject tried to control a computer simulation of a personality. The simulation exhibited 12 levels of friendliness, which ranged from “very rude” to “loving”, and the subject could input any one of these same 12 levels. Each level was represented by an integer between 1 and 12, and the simulation’s behavior followed the equation: \( \text{Output level} = (\text{Input level} \times 2) - \text{Prior level} + \text{Random element} \), where the random element either lowered the output by one level, did nothing, or raised it by one level. The subject tried to maintain the simulation’s behavior at the “friendly” level (which was near the middle of the scale). The subjects’ ability to control the simulation improved substantially over repeated trials.

The third method for studying the learning of complex systems has received considerable attention. It evolved from research on language learning. Because language is so complex and takes so long to learn, a simpler, more tractable method was developed that still retains some of the characteristics of human languages (Miller, 1967; Reber, 1967). This method requires subjects to learn an artificial grammar by observing stimuli generated by the grammar. The artificial grammar (see Figure 1) consists of a finite number of states (represented by black dots in the diagram) and a finite number of allowable transitions between those states, as well as allowable starting and ending points. At each transition, a symbol is generated (e.g., a letter of the alphabet). By traversing the grammar from beginning to end, a string of symbols is generated. Because there are recursions in the grammar, the number of possible strings is infinite; however, the particular strings that can be generated are limited by the rules of the grammar.

In a typical grammar learning experiment (e.g., Reber, 1976), subjects learn to reproduce a series of letter strings generated by a grammar. Subjects may or may not be informed in advance that there are rules governing the order of the letters within the
Figure 1. The artificial grammars used to generate the stimuli in these experiments. Note that in each experiment the letters A-F used in this figure were replaced by other letters, digits, or non-alphabetic symbols.
strings. In either case, after seeing the strings, the subjects are informed of the strings' rule-based nature. Note, however, that the particular rules of the grammar are not revealed. Subjects are then shown a new set of strings, only some of which conform to the rules of the grammar, and they classify each of these test strings as grammatical or ungrammatical.

These are the most commonly used methods for studying the learning of complex systems. Using these experimental procedures, researchers have focused on three main issues. Each issue has been addressed using at least two different methods, demonstrating that the questions are not bound to specific experimental procedures.

The first, and perhaps most thoroughly studied issue, concerns the role of consciousness in the learning of complex systems. The critical finding concerning this issue has been that even when subjects' performance indicates considerable knowledge, they often have little or no ability to verbalize this knowledge. For instance, in the Lewicki et al. (1987) study mentioned earlier, the subjects' reaction times to rule-governed signals improved substantially, yet most subjects were unaware that the signals obeyed any rules at all. Similarly, in the dynamic control task, experience with the computer simulation did not increase subjects' verbal knowledge of its underlying rules, despite the fact that the same experience did substantially improve their ability to control the system (Berry & Broadbent, 1984).

Research using the grammar learning procedure has also focused on the role of consciousness. After being exposed to some of the strings in a grammar, the subjects in Reber and Lewis's (1977) experiment tried over several sessions to re-order sets of scrambled letters into grammatical, yet previously unseen, strings. In each session, Reber and Lewis also asked the subjects to introspect and verbally report everything they could concerning their knowledge of the grammar. Although by the end of the experiment the subjects were able to form grammatical strings with 68% of the
disordered letter-sets, their verbal knowledge was inadequate to account for their performance. Hence, learning the complex system does not seem to have been a completely conscious process. Experiments of this sort have led some researchers to suggest that learning of complex systems is "implicit" (Reber, 1967, 1989) because subjects' performance implies knowledge of which they are not conscious.

The second issue in the learning of complex systems concerns the nature of subjects' knowledge. Is it situation-specific or more abstract? The main technique for studying the level of abstractness involves transfer of training experiments in which subjects learn about a complex system with one set of surface features and are then transferred to another system with different surface features but the same underlying rule structure. Reber (1969) had subjects learn a series of strings to a criterion of two successive error-free recalls. The number of trials necessary to reach this criterion decreased as a function of the number of previous strings recalled. After recalling 18 strings successfully, Reber changed the strings: For one group of subjects, the strings were generated by the same grammar but were instantiated by different letters. For a second group, the strings were instantiated by the same letters but were generated by a different grammar. For a third group, the strings were instantiated by different letters and were generated by a different grammar. For a final group of subjects, the strings were instantiated by the same letters and were generated from the same grammar; in other words, there was no change. For those subjects learning strings from a new grammar, regardless of whether the letters were also changed, performance was seriously impaired relative to the no-change group. On the other hand, for those subjects learning strings from the same grammar with different letters, performance was not discernibly worse than in the no-change condition.

Transfer results such as these have been replicated many times using different methods (e.g., Berry & Broadbent, 1988; Mathews, Buss, Stanley, Blanchard-Fields, Cho, & Druhan, 1989). They indicate that subjects' knowledge is not necessarily tied
to the specific exemplars they studied; rather, it is general to the rule-system underlying the exemplars.

The third issue, and the one of most interest for present purposes, concerns the degree of control that subjects have over the learning of complex systems. The question concerns whether or not subjects can willfully determine what they learn and how they learn. Two opposing positions can be outlined. One position states that learning of this kind is beyond willful control and even beyond consciousness. The other position attributes the learning to processes that can be, and often are, controlled willfully by the learner. Before describing the evidence relevant to this debate, the issue of willful control will be discussed more fully.

**Willful Control**

The issue of willful control of learning has long been central to experimental psychology. During the middle part of this century, when behaviorist theories prevailed, many researchers held the assumption that learning was predominantly nonwillful, controlled more by the environment than by the individual (Bower & Hilgard, 1981; Skinner, 1974). Although this view of learning was not universal (see Leahey, 1987), the tendency of many behaviorists to avoid any consideration of mind biased theorists away from considering the conscious efforts of the learner. In contrast, a common characteristic of current information-processing approaches to learning has been the assumption, usually implicit, that learning is largely controlled by the individual. In many information-processing theories, the individual willfully determines what is remembered through strategies such as rehearsal, chunking, or imagery. For example, such “control processes” were essential to Atkinson and Shiffrin’s (1968) three-store model of memory. More recently, Baddeley’s (1986) model of working memory adds the articulatory loop and the visual-spatial scratchpad
to the array of processes under the command of the individual. Conversely, the extent
to which nonwillful processes control learning has been downplayed.

Experimental psychologists often avoid addressing the issue of control directly
(see Watkins, 1989). Rather, they tend to consider a phenomenon or task in primarily
willful or nonwillful terms. A full understanding of some phenomena, however, will
surely include both willful and nonwillful control. Moreover, the extent to which
control of learning is willful will probably prove to depend on such factors as the task
demands during learning, the task demands during testing, the kind of material being
learning, and the individuals doing the learning (cf. Jenkins, 1979). To understand
willful control, therefore, will require some degree of direct analysis. In other words,
willful control must be measured and the conditions under which it occurs identified.

Willful Control Over the Learning of Complex Systems. Although willful
control is seldom examined directly, the area of complex system learning serves as a
principal exception. Indeed, whether individuals can willfully control how they learn is
one of the major debates in complex system learning literature. Reber and his
colleagues (e.g., Allen & Reber, 1980; Reber, 1989; Reber, Kassin, Lewis, & Cantor,
1980; Reber & Lewis, 1977; Reber, Walkenfeld, & Hernstadt, 1991; see also Lewicki,
1986 and Servan-Schreiber & Anderson, 1990) have taken the strong position that little
or no willful control is involved in this kind of learning:

Prima facie, some actions of mind are causally controlled by conscious
action . . . . On the other hand, some actions of mind are prima facie
causally controlled by aspects that lie outside of consciousness.
Essentially every complex knowledge acquisition task is accomplished
largely in the absence of conscious control (Reber, Allen, & Regan,
1985, p. 22).
The issue of control is often tied to one of the other major themes in the learning of complex systems; namely, the issue of consciousness. Because Reber maintains that this kind of knowledge is learned without conscious processes, he is led to the position that control over learning is nonwillful. It would be difficult to argue that learning is both unconscious and willfully controlled. Instead, Reber (1989) claims that knowledge of complex systems, or more generally, complex knowledge, is acquired "automatically."

Because the idea of automaticity is central to the nonwillful control argument, it deserves detailed consideration. The term "automatic" is a loaded one in cognitive psychology. Although definitions of what constitutes an automatic process vary considerably and are sometimes contradictory, they can be sorted into two broad categories. One category derives primarily from work on attention (e.g., Shiffrin & Schneider, 1977). These definitions state that a process is automatic if it a) requires no attentional capacity so that it does not interfere with other tasks, b) is not available to conscious awareness, and c) cannot be consciously inhibited. Another aspect of this class of definitions is that automaticity is usually assumed to result from extensive practice (an assumption which eliminates most physiological processes from consideration). Hence, automaticity is learned. Before becoming automatic, a process is controlled consciously. Skills such as typing are illustrative of the phenomena captured by this kind of automaticity.

The other category of automaticity definitions derives not from attention research, but from memory research. Hasher and Zacks (1979) proposed a theory of automaticity which stated that there were two kinds of automatic processes. The first roughly corresponds to the skill-learning automaticity outlined above. The other kind of automatic process is similar in that it does not drain attentional capacity, does not require awareness or intention, and cannot be willfully inhibited. What distinguishes
this second category of automatic process is that it is assumed to be "hard-wired" in the brain; in other words, it is innate. Consequently, automatic processes should not be affected by developmental processes: The very young and the very old should resemble young adults. Likewise, automatic processes should show little effect of individual differences. The acquisition of knowledge concerning frequency and spatial location illustrate the kinds of processes Hasher and Zacks had in mind. In contrast to the extensive practice required by automatic processes of the skill-learning variety, these "hard-wired" or innate processes should not be affected by practice at all (Hasher and Zacks, 1979).

When Reber refers to "automatic" processes, he seems to refer to Hasher and Zacks's innate automaticity. For instance, Reber, Walkenfeld, and Hernstadt (1991) explored the role of individual differences in the learning of artificial grammars, arguing that none should be present. Reber suggests that complex systems require a kind of learning that is evolutionarily prior to the kind of learning required for the verbal materials commonly found in psychology experiments (Reber, 1989, 1990; see also Abrams & Reber, 1988). The skill-learning view of automaticity is generally inconsistent with arguments about evolution and a lack of individual differences.

It should be noted that the automatic vs. control processes distinction is roughly analogous to the willful vs. nonwillful control distinction; nevertheless, they are not the same. Although willful control generally corresponds to control processes, nonwillful control goes beyond automaticity. Both automatic and control processes are characteristics associated with the individual subject, differing only in their respective susceptibility to conscious governance. Nonwillful control includes not only "internal" processes; it also includes those characteristics of the stimulus environment that constrain the nature of cognitive processing. Hence, in nonwillfully controlled tasks,
manipulating subjects’ effort, practice,\textsuperscript{1} or instructions will have little effect; only changes in the stimulus (or, perhaps, in the physiology of the subject) will influence performance. The distinction between willful and nonwillful control is, therefore, not redundant with the distinction between controlled and automatic processes.

Reber (1989, 1990) is not alone in advocating a lack of willful control in the acquisition of complex knowledge. Lewicki and his colleagues (1986; Lewicki, Cyzewska, & Hoffman, 1987; Lewicki, Hill, & Bizot, 1988) have also argued that complex information is learned via "internal processing algorithms" which are unconscious and beyond the control of the learner. In a similar vein, Broadbent and his colleagues (e.g., Berry & Broadbent, 1988; Broadbent, FitzGerald, & Broadbent, 1986; Hayes & Broadbent, 1988) have suggested that there are two modes for learning: a selective mode which consists of various willfully controlled strategic processes, and an unselective mode which consists primarily of the automatic acquisition and aggregation of frequency information. Although the specific theories differ to some extent, all of these researchers share the belief that the willful efforts of the subject are relatively unimportant to the learning of complex systems.

The claim that people cannot exert willful control over the learning of complex systems is based mainly on experiments comparing intentional learning and incidental learning. Reber (1976) argued that if grammar learning is not controlled consciously, then conscious processes should not improve learning. To test this hypothesis, Reber manipulated the instructions that subjects were given. All subjects were told to memorize the letter strings. In addition, some subjects were told that the letter strings were generated according to a set of rules and that discovering these rules would make the strings easier to learn. Hence, learning of the 	extit{strings} was intentional for all subjects; however, learning of the 	extit{rules} governing the strings was intentional only for

\textsuperscript{1}Note that even with automaticity acquired through practice, once performance reaches the level of automaticity, further practice has little effect.
half of the subjects. Note that this is somewhat of a departure from the usual usage of intentional and incidental learning in that “incidental learning” usually refers to learning that occurs when subjects are not even trying to encode the stimuli for later recollection. What Reber found was that the subjects intentionally trying to learn rules took longer to learn the strings than the subjects who learned the rules incidentally. Furthermore, when the subjects were asked to discriminate between grammatical and ungrammatical strings, the incidental subjects actually outperformed the intentional subjects (see Brooks, 1978, for a replication). Thus, not only did efforts to willfully encode the rules not help subjects’ performance, but they actually hurt performance. Reber (1976) took these results to mean that the learning that did occur in each group was not a result of conscious strategies; the efforts of the intentional subjects merely led them to generate conscious but erroneous rules, thereby muddling their automatically acquired, “intuitive” understanding of the artificial grammar.

Similar results have come from experiments using the dynamic control task. As discussed earlier, in this paradigm, the subject puts information into a system (e.g., a simulated person) which changes its level (of friendliness) depending on a combination of the subject’s input (also a level of friendliness) and its own previous level. The subjects’ ability to control the system’s level improves far more than their ability to explain verbally the rules that govern the system (Berry & Broadbent, 1984). Berry and Broadbent (1984) explored whether subjects could, with proper instruction, improve their degree of learning. They gave some subjects training (including verbal instructions and examples) on what types of inputs would be most effective in controlling the system. Training did not result in any improvement in performance.

Along the same lines, Berry and Broadbent (1988) trained subjects on the simulated person task and then transferred them to a “new” simulated person which was governed by the same underlying equation. It differed only in the arrangement of
the text and graphics on the computer screen. Some subjects were informed that the
two simulations were based on the same rules; others were not informed. The
informed subjects performed no better or slightly worse on the second simulation than
on the first simulation. The uninformed subjects, on the other hand, performed
substantially better on the second simulation. In fact, they performed almost as well as
a control group that did not switch to a new simulation. As in the intentional condition
of the grammar learning task, revealing the existence of an underlying structure to the
complex system impaired rather than aided performance.

An apparent implication of these studies is that subjects are unable to willingly
control their learning of complex systems. When informed of the presence of rules or
when told explicitly how best to approach the task, performance showed either no
benefit or a slight decrement. If subjects were able to control their learning, then this
additional information should have helped them.

Still, this incidental learning advantage reported by Reber (1976) has not always
been replicated. A number of experiments using procedures more or less similar to
those used by Reber (1976) have failed to reveal any statistically reliable advantage for
incidental instructions (Dienes, Broadbent, and Berry, 1991; Dulany, Carlson, and
In addition, Abrams (cited in Reber, 1989) failed to replicate Reber’s (1976) finding,
even though he tried to directly replicate Reber’s procedures (with the exception that the
experiment was conducted with each subject sitting at a computer).

Mathews, Buss, Stanley, Blanchard-Fields, Cho, and Druhan (1989) argued
that only a weak manipulation of intentionality has been used in most experiments.
Rather than simply giving vague instructions to discover rules, Mathews et al. tried to
induce active hypothesis testing in their intentional subjects. Specifically, in their
intentional, or edit, condition, the subjects were shown a series of strings, each
containing one to four letters that violated the rules of the grammar. They tried to
identify all of the violations in the string and were then shown the correct string. In their incidental, or match, condition, subjects saw a “good” string, held it in mind for a few seconds, and then chose the string that exactly matched it from a set of five alternatives. At test, their subjects were given 100 multiple-choice trials. Each trial comprised five alternatives which violated the grammar 0, 1, 2, 3, or 4 times. The subjects were to choose the string with no violations. The interesting finding is that there was no appreciable difference between the two conditions. The edit task did not improve performance in any way, further strengthening the argument against an important role for willful control in learning complex systems.

Although the evidence that intentional learning hurts performance is equivocal at best, it seems fairly clear that intentional instructions do not help performance. This general lack of an intentional-learning advantage supports the nonwillful control position. It has generally been assumed that if subjects were able to willfully control their learning of complex systems, then being informed of those rules would help. The data suggest that information concerning the presence of rules is of little use.

Not all of the studies that address willful control frame the issue in terms of intentional learning of the grammar’s rules vs. incidental learning of the grammar’s rules. For example, in a study by Abrams and Reber (1988) subjects suffering from major psychiatric disorders (e.g., depression, schizophrenia) or from alcoholism were compared with normal college students on an incidental grammar learning task. The interesting finding was that no differences emerged between the patient groups and the college students in terms of grammar learning. Abrams and Reber did find that on a simpler task the college students outperformed the patients, showing that simple, rather than complex, rules had to be learned, the patients’ mental impairments had a detrimental effect.

Another study supporting the nonwillful control position was reported by Servan-Schreiber and Anderson (1990). They presented evidence that subjects in a
grammar learning experiment spontaneously learned strings in terms of "chunks" of letters that were meaningful units within the grammar. Furthermore, when the strings were presented so as to discourage chunking into meaningful units, performance was substantially impaired. These data imply that characteristics of the stimulus determine subjects' performance, at least in part.

Finally, Hayes and Broadbent (1988) showed that in a dynamic control task, learning was unaffected by a concurrent, secondary task. Their subjects learned to control one of two simulated persons, one governed by a relatively obvious rule, the other governed by a more obscure rule. Previous experiments had shown that subjects typically discovered and could verbalize the obvious rule but not the obscure rule, although performance was otherwise equivalent for the two rules. Hayes and Broadbent assumed that this implied different "modes" of learning. The kind of learning used for the obvious rule corresponds to ordinary hypothesis testing, whereas the kind of learning used for the obscure rule corresponds to the "implicit" knowledge acquisition discussed by Reber (1989, 1990) and Lewicki (1986). Hayes and Broadbent trained their subjects on one of the two simulations and then added a random letter generation task (see Baddeley, 1966) in which the subjects tried to produce random letters while interacting with the simulation. Performance was equal for the two conditions, and neither was impaired by the introduction of the secondary task. Hayes and Broadbent report the mean of only the last 10 of the 50 trials in each phase; consequently, any differences present in the earlier trials are not known. In the final phase of the experiment, the experimenters reversed the rules (although maintaining the obscure/obvious distinction) and continued the letter generation task. The obvious-rule subjects were substantially impaired on the last 10 trials. The obscure-rule subjects, in contrast, performed at the same level on the last 10 trials of the final phase as on the last 10 trials of the two earlier phases. Hayes and Broadbent concluded that the subjects in
the obscure-rule condition unconsciously learned the reversal whereas the subjects in
the obvious-rule condition were unable to do so consciously because they were
preoccupied with the secondary task. That dividing subjects’ attention had no effect on
the obscure-rule subjects supports the notion that nonwillful control position.

Not all researchers subscribe to the idea that the learning of complex systems is
beyond willful control. Dulany, Carlson, and Dewey (1984, 1985) have strongly
advocated the position that complex knowledge is acquired via willfully controlled
processes. Others, such as Brooks (1978), Vokey and Brooks (1992; see also Brooks
& Vokey, 1991), and Mathews (1991) have also argued for a greater degree of willful
control than proposed by Reber (1989), Lewicki (1986), or Broadbent (1989). Each of
these researchers advocates a somewhat different explanation of how people master
complex systems, but all believe that willful control is essential.

Dulany, Carlson, and Dewey (1984, 1985) suggest that subjects do not learn
the rules that actually comprise the complex system. Referring specifically to artificial
grammars, they claim that each subject consciously infers a set of rules that are
correlated with the actual rules of the grammar. When asked to state the rules, the
subjects are therefore unable to specify the actual rules. Subjects in Dulany et al.’s
(1984) experiment were able to specify which parts of a test string they believed
violated the grammar, implying that they were judging string grammaticality on the
basis of conscious knowledge. Dulany et al. (1984, 1985) do not offer an explanation
for why incidental learning is sometimes better, and seldom worse, than intentional
learning.

An analogy strategy for grammaticality judgments has been suggested by
Brooks (1978). He proposes that when subjects are asked to decide whether a test
string is grammatical, they give positive answers whenever the test string resembles at
least one of the strings they saw at study. Hence, the subjects do not automatically
abstract rules; instead, they consciously (i.e., willfully) try to recollect similar instances
from the past. The knowledge of the artificial grammar is little more than the ability to remember a variety of instances: No general rules are learned. To test this analogy hypothesis, Vokey and Brooks (1992) exposed subjects to strings from a grammar. The test items either did or did not fit the grammar, and the similarity of a test item to any of the previously studied items was varied independently of its grammaticality. Vokey and Brooks found that both grammaticality and similarity accounted for statistically significant proportions of the variability in subjects' performance, thereby supporting the hypothesis that subjects based some of their grammaticality judgments on analogies to earlier strings. In fact, the statistical effect size for grammaticality and similarity were essentially equal (see McAndrews & Moscovitch, 1985, for a similar finding).

Brooks' (1978) analogy hypothesis can also explain why intentional instructions do not improve subjects' performance: If performance is based on remembering specific strings, then looking for rules is unlikely to help. In fact, Brooks (1978) suggested that the reason he and Reber (1976) found an advantage for incidental instructions was that the subjects in the intentional conditions were wasting so much effort looking for rules that they were not learning the strings as well as the incidental subjects. This hypothesis does not explain the failure of other studies to find an intentional-learning advantage.

Mathews and his colleagues (Mathews, 1990; Mathews, 1991; Mathews et al., 1989; Stanley, Mathews, Buss, & Kotler-Cope, 1989) do allow for some nonwillfully controlled processes, but they also advocate an important role for willfully controlled processes. In their view, knowledge of a complex system is based on two sources: a "mental model" of the task and memory for past experiences with the task. By mental model they mean that subjects have consciously inferred rules that govern the system. The memory-based process is similar to that advocated by Vokey and Brooks (1992). Specifically, Mathews (1991) suggests that subjects are able to recollect fragments of
previously seen strings, and these recollected fragments are used as "rules." That is, subjects base their decisions on whether or not there is a match between the test string and at least one recollected fragment of a previously studied string. Using a combination of these two processes, subjects learn to understand a complex system.

The considerable evidence supporting the nonwillful control position notwithstanding, a variety of evidence has been accumulated that supports the opposing position—that complex systems are learned through willful control. This evidence does not, however, uniquely support any of the aforementioned theories. Although most of the data indicates that intentional-learning instructions are not beneficial in this kind of learning, some experiments do show that, at least under some circumstances, intention to learn can help. Reber, Kassin, Lewis, and Cantor (1980) found that when strings were presented to subjects in a simultaneous display that made the structure of the strings salient, then intentional instructions led to superior performance. Millward (1981) also found an advantage for intentional learning. Reber et al. (1980) suggested that Millward’s results were due to his use of relatively long strings which highlighted the structure of the strings. These studies indicate that subjects can control learning if the rules underlying the complex system are made relatively obvious.

In an experiment by Reber, Kassin, Lewis, and Cantor (1980), subjects were shown a schematic rendering of an artificial grammar (as in Figure 1) and a demonstration of how to generate strings from the grammar. After generating several strings themselves, the subjects were exposed three times to a set of 21 strings from the grammar. This was followed by a discrimination task in which they tried to classify a different set of strings as grammatical or ungrammatical. The performance of these subjects was compared with that of subjects who simply observed the strings several times without any training on, or even knowledge of, the rules. The performance of the subjects who were shown the grammar before observing the strings was much
better than that of the subjects who only observed the strings. This is, of course, an extreme example of training: The subjects were essentially given the rules. Nevertheless, it illustrates that under some conditions, subjects can willfully control what they learn.

Using a dynamic control task, Stanley, Mathews, Buss, and Kotler-Cope (1989) examined the effect of several kinds of intentional-learning instructions. One group of subjects was given general hints as to the rules underlying the simulated personality's behavior. For instance, the subjects were told that the simulation tended to be "overreactive"; that is, given an input behavior, the output behavior is likely to be in the same direction (i.e., more positive or more negative than the previous level) but more extreme. The experimenters provided specific examples of this behavior along with possible solutions. Another group was given a specific heuristic which, if followed exactly, would lead to the maximum possible performance on the task. A control group given no special instructions was also included. The subjects' performance indicated that the two methods of training were essentially equivalent and that both were better than the control condition. In other words, the subjects were able to use the knowledge acquired during training to benefit their performance.

A recent experiment by Dienes, Broadbent, and Berry (1991) suggests the presence of willfully controlled processes. They found that a concurrent, secondary task (generating random numbers) impaired grammar learning. Random number generation requires considerable conscious effort, leaving little for any other task. If grammar learning were completely automatic (i.e., nonwillful), performance should not have been impaired by the number-generation task. Note that this conclusion is in direct contrast to that reported by Hayes and Broadbent (1988); using the dynamic control task, they found that learning was unaffected by a secondary task. The results of Dienes et al. suggest that task-specific factors may constrain general conclusions about complex system learning.
The conclusion from this review is that, even though the preponderance of evidence is consistent with the nonwillful control position, an appreciable contribution of willfully controlled processes cannot be ruled out. This equivocal conclusion stems from the fact that existing evidence is rather narrow in its focus. Although Reber, Allen, and Regan (1985) referred to the issue of control as "central," it has been addressed primarily in terms of differences between incidental rule-learning and intentional rule-learning. Converging evidence is scarce. Moreover, the intentional vs. incidental learning paradigm is a weak test: Even if subjects have some ability to control their learning, vague instructions concerning rules may mislead them into searching for inappropriate rules. In general, subjects given intentional-learning instructions may be at a loss for how to approach the task. Mathews et al. (1989) tried to address this problem by more explicitly directing the subjects' behavior in the intentional-learning paradigm, but they still found no difference between incidental and intentional learning. There are, of course, the studies by Reber et al. (1980) and Stanley et al. (1989) that showed that when subjects were explicitly given knowledge of the underlying rules, they performed better. These studies, although suggestive, seem to speak to the potential for formal instruction to improve the learning of complex systems. The focus of the present research, in contrast, is how people learn complex systems from examples, not from formal instruction.

**Willful Control and the Effect of Irrelevant Information.** The methods used to study the learning of complex systems have not provided any clear insight into the issue of willful control. How, then, should willful control be examined? One answer is look at how this issue has been addressed in other domains.

Of all of the areas in cognitive psychology, the issue of control has been addressed most often by researchers interested in attention. For instance, the concept of automatic processing--i.e., processing beyond the willful control of the individual--was developed in the context of attentional studies. Shiffrin and Schneider (1977)
trained subjects to respond to the presence of certain target items and to ignore certain distractor items. Items were consistently used as either targets or distractors. After thousands of trials, subjects were able to detect the presence of a very briefly displayed target just as often when it was accompanied by three distractor items as when it was presented alone. Shiffrin and Schneider argued that subjects’ attention was automatically drawn to the target items. To test this hypothesis, Shiffrin and Schneider (1977, Experiment 4d) had subjects search for targets in a 2 x 2 display containing three distractor items, but the subjects knew that the target would appear only on one of the diagonals. The items in the other, irrelevant, diagonal would never be target items; however, some of these distractors were items to which subjects had been previously trained to respond automatically. In other words, in a previous task, the subjects were trained to automatically respond positively to these items, but in this task, they were to try and ignore these items. The subjects’ detection rate was seriously impaired on those trials in which former targets were present on the irrelevant diagonal. A related experiment (Shiffrin & Schneider, 1977, 4a) ruled out the possibility that subjects were simply unable to ignore an irrelevant diagonal. Instead, they were specifically unable to ignore information to which they had been trained to respond automatically.

This example illustrated a method of investigating willful control: Measure the extent to which people can ignore nominally irrelevant information that is presumed to be processed automatically. If subjects are able to ignore it, then some aspects of its processing must be under willful control. The “nominal irrelevance” technique, as it might be called, can be found in other studies of attention. These studies would include the Stroop effect (Stroop, 1935). The classic Stroop effect occurs when subjects are asked to name the color of the ink in which the name of a color is written. Thus, when blue is written is green ink, subjects are slower to say “green” than when an unrelated word is written in green ink. The color name cannot be ignored. This finding is
consistent with the generally accepted theory that reading, at least for educated adults, is an automatic process (Rayner & Pollatsek, 1989).

Along the same lines as the Stroop effect, attention researchers have identified an effect of irrelevant information known as “Garner interference” (Pomerantz, 1983). Garner interference is found in tasks that require subjects to selectively attend to one stimulus while ignoring another stimulus. For example, the subjects might see a pair of parentheses such as (( , )), ( ), or ). They are told to hit one key when the left parenthesis “faces” right, (( or ), and another when the left parenthesis faces left, )) or ). The right parenthesis is irrelevant to the task, at least nominally so. The experimenter then varies the right parenthesis so that it is either constant or varies orthogonally to the direction of the left parenthesis. Garner interference is indicated when orthogonal variation of the irrelevant stimulus impairs performance relative to no variation of the irrelevant stimulus. An absence of Garner interference indicates that subjects are able to willfully control their attention. If they cannot control their attention, then the irrelevant variation impairs performance.

The nominal irrelevance technique has also been used in memory research. Perhaps the clearest instance of this technique being used is with research on the suffix effect. The suffix effect is an impairment of immediate recall of auditorily presented items resulting from the presence of a nominally irrelevant auditory item at the end of the list (Dallet, 1965). For example, a group of subjects might hear lists of letters, their task being to recall the items in each list in their order of appearance. If some of the lists are followed by an auditory distractor such as the word “zero,” recall of those lists is worse than recall of lists with a non-speech recall signal. In fact, performance in the suffix conditions is nearly as bad as if subjects had been asked to remember an additional item, suggesting that they had little, if any, ability to ignore the suffix (Crowder, 1967). In contrast to the auditory suffix effect, presenting a nominally
irrelevant visual stimulus following a visually presented list has little or no effect on immediate recall of visual items.

This brief review of attention and memory phenomena illustrates the usefulness of the nominal irrelevance technique. Each of the paradigms discussed--automatic visual search, the Stroop effect, Garner interference, and the suffix effect--allows a range of phenomena to be classified as willfully or nonwillfully controlled. Moreover, all of these paradigms allow for a more fine-grained analysis than a simple classification as willful or nonwillful. In the Stroop paradigm, for example, the degree to which subjects’ color-naming responses are slowed by incongruent words serves as a continuous measure of the loss of willful control. A continuous measure allows for a much richer consideration of the willful control concept.

The Present Study

The central methodological problem for this research was how to apply the nominal irrelevance technique to the study of the learning of complex systems. A decision was made to limit the research to a single method. Although any of the methods discussed to this point--the reaction time task, the dynamic control task, or the grammar learning task--would have served adequately, the present series of experiments employed the grammar learning method. This choice was based on the fact that the grammar learning task has been employed most frequently in previous research.

Although many earlier studies compared intentional learning with incidental learning, all but one of the present series of experiments involved only intentional learning. Presumably, subjects need to be aware that rules are present in order to exert willful control over how these rules are learned.

The nominal irrelevance technique was applied to the grammar learning task. Before describing the procedure that was used for the present series of experiments,
mention should be made of an alternative procedure which was considered and rejected. The rejected procedure was perhaps a more obvious approach than the one used. In fact, it was used in several pilot studies. Relatively few subjects were tested in any of these preliminary experiments; nevertheless, the data revealed inherent problems with the procedure.

To implement the nominal irrelevance technique, irrelevant information must be introduced into the stimuli. In the rejected alternative, the subjects learned strings that fit the grammar (grammatical strings) and strings that did not fit the grammar (ungrammatical strings). Some subjects were informed of which strings were grammatical. Other subjects remained uninformed. Willful control was indicated by the advantage of the informed over the uninformed condition. The logic was that if the informed subjects were able to ignore the irrelevant information, they should perform better than the uninformed subjects. A "baseline" condition was also included in which no ungrammatical strings were presented. This condition was necessary to determine whether there were any limitations on subjects' willful control. If subjects had complete willful control, then performance in the informed condition should be equal to that in the baseline condition. That is, the subjects should behave as though the irrelevant information had not been presented.

This procedure proved problematic. If the ungrammatical strings deviated from the grammar too dramatically, subjects were able to discriminate between the two kinds of strings. Unfortunately, the only way to make the ungrammatical strings more difficult to discriminate from the grammatical strings was to make them more grammatical. The ungrammatical strings therefore contained a non-trivial amount of useful information about the grammar. Consequently, in addition to seeing grammatical strings, subjects in the uninformed condition also saw ungrammatical strings which contained some grammatical information. In the baseline condition, on the other hand, subjects saw only grammatical strings. In the preliminary experiments,
performance in the uninformed condition tended to equal or exceed that in the baseline condition, implying that the irrelevant information did not impair performance. The logic of the nominal irrelevance technique requires that the irrelevant information have a detrimental effect; otherwise, there is little point in measuring subjects’ ability to ignore it. Comparing grammatical and ungrammatical strings was not an effective implementation of the nominal irrelevance technique.

A more effective procedure was to add irrelevant information to each string. Altering each string is roughly analogous to presenting subjects with examples of a “noisy” system. This is perhaps more reflective of real-life complex systems, which are seldom without flaws. In this procedure, additional letters were inserted into the strings produced by the grammar, rendering the strings ungrammatical. These letters were drawn from the same vocabulary as the grammar; hence, the grammatical and ungrammatical letters were not distinguishable without knowledge of the grammar. For the subjects to ignore the ungrammatical letters, they had to be informed of the locations of these letters. This was done by marking each ungrammatical letter with a black dot, positioned directly above it. Willful control was indicated by a subject’s ability to ignore the dotted letters. Note that in order to measure this ability a control condition was needed in which the subjects were uninformed of the locations of the ungrammatical letters. The difference between the informed and the uninformed conditions served as the measure of willful control.

A third condition was also included in some experiments. In this condition, the ungrammatical letters were not present in the grammar’s strings. By comparing the informed condition (strings with dotted ungrammatical letters) to this baseline condition, it was possible to measure the limits of willful control: Any disadvantage of the informed condition indicates that subjects were unable to completely block the
effects of ungrammatical information. If the subjects had complete willful control, learning in the informed condition should equal that in the baseline condition.

THE EXTENT OF WILLFUL CONTROL IN LEARNING ARTIFICIAL GRAMMARS

The purpose of the first four experiments was to directly measure the extent of and limitations on subjects' willful control over the learning of artificial grammars. The remaining experiments explored variables that seemed likely to increase or decrease willful control. An identification of such variables would be important to any comprehensive theory of how people learn complex systems.

Experiment 1

The purpose of this experiment was to measure the extent of willful control in the grammar learning paradigm. The subjects saw strings of letters generated from an artificial grammar (see Figure 1). Immediately afterward, they were tested on their knowledge of this grammar. This study-test cycle was repeated twice more, each time with a new grammar. The presentation of each grammar differed depending on the experimental condition: In the baseline condition, the subjects saw strings with no ungrammatical letters added. In the informed condition, the subjects saw strings which, although otherwise grammatical, included two ungrammatical letters, and these ungrammatical letters were clearly marked. In the uninformed condition, the subjects also saw strings with two ungrammatical letters, but these letters were unmarked.

Unsurprisingly, performance in the uninformed condition was impaired relative to performance in the baseline condition. The main question, however, was whether identifying the ungrammatical letters would allow the subjects to block their effects. If artificial grammars are learned consciously and willfully, then performance in the informed condition should be better than performance in the uninformed condition. On the other hand, if, as Reber (1989) claims, subjects cannot willfully control grammar
learning, then there should be no difference in performance between the informed and uninformed conditions. In either case, if there are any limitations on willful control whatsoever, these should be manifested in a disadvantage of the informed condition as compared to the baseline condition.

**Method**

**Design.** Each subject participated in the baseline condition, the informed condition, and the uninformed condition. In each condition, the subject saw 20 strings generated by an artificial grammar; a different grammar was used for each condition. In the baseline condition, none of the strings contained ungrammatical letters. In the informed condition, each string included two marked letters that violated the grammar. In the uninformed condition, each string included two unmarked letters that violated the grammar.

The three grammars appeared in the same order for all of the subjects. The condition in which each grammar appeared was rotated so that, across all subjects, each grammar was used in each condition equally often. Additionally, the order in which the strings from each grammar was presented was the same for all subjects.

**Materials.** The strings were generated from three artificial grammars (see Figure 1). Grammar 1 and Grammar 2 have been used previously by Reber (1969; Reber & Allen, 1978). Grammar 3 was generated by the experimenter. It had the same number of single letter and multi-letter recursions, and each of the five states produced two letters via transfer to a new (or to the same) state. Thus, on these characteristics, at least, the three grammars were equivalent.

The letters for each grammar were chosen from the alphabet in a pseudo-random fashion: Vowels and the letter Y were excluded, and letters were replaced when too many visually similar letters were assigned to the same grammar. For Grammar 1, the letters were: C, L, K, W, B. For Grammar 2, the letters were G, N, Z, X, R. For Grammar 3, the letters were M, V, H, D, J. The assignment of letters to
particular positions within the grammar was arbitrary. Each of the letters listed here corresponds to one of the letters in Figure 1.

From each grammar, 45 unique strings were generated. There were nine strings at each of five letter-lengths: 7, 8, 9, 10, and 11. From each set of nine strings, four were randomly selected to be viewed during the study phase. The remaining five strings were used during the test phase. The 20 study strings were randomly ordered, as were the 25 test strings. The same ordering was used for all subjects.

An additional version of each of the study strings was created by randomly inserting two letters within the existing string; however, no letter was ever inserted before the first letter and none of the inserted letters appeared next to each other. For example, if the original string were CKKCBWL, the altered version might be CWKKCKBWL or CKKCWBWLK, with an extra W and an extra K. The original string could not be changed to read WCKKCKBWL or CKKWKCBWL (Note that bold type is used here for purely expository purposes; subjects saw all letters in the same typeface). The choice of which letter to insert was random with the constraint that it was one of the five letters used in the grammar but could not have been produced by the grammar. For example, a G could not be inserted in the middle of the string XGGGZZ because the grammar permits the letter G to repeat an infinite number of times. The original version of each string was used for the baseline condition, and the altered version was used for the informed and uninformed conditions.

Each of the strings used during the test phase was also altered. Specifically, a single letter, chosen at random, was removed. At test, the subjects saw the string with the missing letter replaced by a dash. One of the remaining four letters was chosen at random to serve as the lure in a two-alternative forced-choice test. The position of the
correct answer and of the lure was determined randomly for each subject on each test trial.

Procedure. The subjects were tested individually or in groups of up to five people. They were told that they would see strings of letters arranged according to a complex set of rules. The experimenter informed the subjects that their task was to learn the rules. He also told the subjects that the best strategy for learning these rules was simply to study the patterns of letters that occurred in the strings. They were specifically warned not to generate any simplistic rules.

Each string was presented for 20 seconds. In the informed and uninformed conditions, the subjects were told that each string contained two ungrammatical letters. In the informed condition, each of the two ungrammatical letters was marked with a black dot ("•") directly above it, and the subjects were told to ignore these letters. After seeing all 20 strings, the subjects were presented with 25 new strings, each of which had a letter missing. On the computer screen appeared the question: "Which of these letters fits the blank?" Directly below this question were two buttons (rectangular areas on the computer screen), each labeled with a letter. The subjects responded by moving the computer cursor into the button representing the answer they preferred and depressing the key on the computer mouse. Immediately following this response, a new test string replaced the previous one, and two new response alternatives appeared.

After the subjects had responded to all 25 test strings, they read instructions on the computer screen explaining the next condition and then pressed a button on the computer to begin that condition.

Subjects. The subjects were 21 Rice university undergraduates. They participated as part of a course requirement or in return for extra credit in a psychology course.
Results and Discussion

The dependent measure was the probability of a choosing the correct answer on a given test trial. The means of these probabilities are reported in the first row of Table 1. The first question is whether adding ungrammatical letters impaired performance when they were unmarked. The uninformed condition was reliably worse than the baseline condition, $t(20) = 5.09, p < .001$. Given that the addition of ungrammatical letters hurt performance, the next question is whether the effect of the ungrammatical letters was attenuated by informing the subjects of their location. Statistical analysis shows that the informed condition exceeded the uninformed condition to a reliable extent, $t(20) = 2.32, p = .02$. This indicates that subjects were able to ignore the ungrammatical letters at least to some degree. A final question concerns the limitation on the subjects' ability to ignore the ungrammatical letters. If any such limitation existed, then the informed condition should have been reliably worse than the baseline condition, which, in fact, it was, $t(20) = 2.82, p = .005$. Thus, Experiment 1 demonstrates that subjects have at least a limited ability to willfully control how they learn artificial grammars.2

Experiment 2

The results of the first experiment indicated that subjects could willfully control their learning, although this ability was not without limitations. Experiment 2 was an attempt to replicate this finding, although with a slightly different test procedure. Instead of being given two alternatives to choose from at test, the subjects were given five alternatives which they ranked from most likely to least likely to fit in the test string. Another change in method between Experiments 1 and 2 was that each subject in Experiment 2 participated in only one experimental condition, rather than in all three.

2In this experiment, and in all of the other within-subjects experiments, the pattern of means are generally the same for each of the grammars considered alone as they are for the data considered as a whole. The few exceptions are unsystematic, presumably the result of the relatively sparse data used to compute each mean.
Table 1

Means for Experiments 1-4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Measure</th>
<th>Informed</th>
<th>Uninformed</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>proportion correct</td>
<td>.68</td>
<td>.59</td>
<td>.77</td>
</tr>
<tr>
<td>2</td>
<td>proportion correct</td>
<td>.32</td>
<td>.23</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>average rank (1-5)</td>
<td>2.70</td>
<td>2.96</td>
<td>2.30</td>
</tr>
<tr>
<td>3</td>
<td>proportion correct</td>
<td>.66</td>
<td>.62</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>average confidence score</td>
<td>4.06</td>
<td>3.85</td>
<td>4.40</td>
</tr>
<tr>
<td>4</td>
<td>proportion correct</td>
<td>.70</td>
<td>.60</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>average confidence score</td>
<td>4.15</td>
<td>3.84</td>
<td>4.41</td>
</tr>
</tbody>
</table>
Method

Subjects and Design. The subjects were 51 Rice university undergraduates who participated as part of a course requirement or in return for extra credit in a psychology course. They were evenly and randomly distributed among the three experimental conditions, which were the same as in Experiment 1: baseline, uninformed, and informed.

Materials. There were 20 study strings, with an unaltered version of each string for the baseline condition and an altered version for the informed and uninformed conditions. There were also 25 test strings. The strings—unaltered, altered, and test versions—were the same as those used for Grammar 1 of Experiment 1.

Procedure. The study procedure was exactly as in Experiment 1, with subjects studying each of 20 letter strings for 20 seconds in an attempt to discover the underlying rules. The test procedure was slightly different, however. The subjects saw the 25 test strings, each with a single letter missing, but instead of choosing between two alternatives, they ranked the five letters that comprised the vocabulary of the grammar. Specifically, they chose their most favored response among all five letters which were presented in a column below the test string and entered their response into the computer. The chosen response disappeared from the column of letters, and the subjects then selected their most favored response from the remaining four letters. This elimination process continued until the subjects had chosen between the final two letters. Immediately after this final choice was made, the next test string appeared.

Results and Discussion

The ranking procedure used in this experiment allows for the computation of more than one dependent measure. Perhaps the most obvious measure is the probability that a subject’s first choice was correct. The second row of Table 1
presents the means for this measure. A comparison of the subjects in the baseline condition with the subjects in the uninformed condition reveals that the baseline subjects substantially outperformed the uninformed subjects, demonstrating that the presence of ungrammatical letters impaired learning, \( t(32) = 3.79, p < .001 \). Likewise, the advantage of the informed subjects over the uninformed subjects indicates that informing the subjects of which letters were ungrammatical alleviated some of the impairment caused by these letters, \( t(32) = 2.14, p = .02 \). Finally, although informing subjects about the ungrammatical letters allowed subjects to block their effect, the data implied that the subjects were not completely successful because the subjects in the informed condition performed at a reliably lower level than did the subjects in the baseline condition, \( t(32) = 1.88, p = .03 \). If the informed subjects had been completely successful in ignoring the ungrammatical letters, there would have been no appreciable difference between these two groups.

Another way that performance was measured in this experiment was by computing the average ranking of the correct answer, assuming that choosing the correct answer first yielded a rank of 1 and never choosing it yielded a rank of 5. The third line of Table 1 summarizes the results of such an analysis. The same pattern of results emerged: The ungrammatical letters impaired performance; informing subjects of which letters were ungrammatical allowed them to ignore these letters; but their ability to ignore the ungrammatical letters was less than complete. These conclusions follow, respectively, from the fact that the baseline condition was reliably superior to the uninformed condition, \( t(32) = 4.95, p < .001 \); the informed condition was reliably superior to the uninformed condition, \( t(32) = 2.01, p = .03 \); and the informed condition was reliably inferior to the baseline condition, \( t(32) = 2.68, p = .006 \). The results of Experiment 2 generally replicated those of Experiment 1.
Experiment 3

In the first two experiments, the subjects studied the strings by simple observation. Most of the experiments in the grammar learning literature (e.g., Reber, 1967) have had subjects learn the strings by learning to recall them. Requiring the recall of each string during study, including both grammatical and ungrammatical letters, might be expected to diminish willful control because the subjects would have to process the ungrammatical letters deeply enough to recall them. Hence, a demonstration of willful control in spite of a recall requirement would further strengthen the evidence that subjects can exert willful control over the learning of artificial grammars. The aim of Experiment 3 was to determine whether the general pattern of results found in Experiments 1 and 2 would be replicated when the subjects were required to study via repeated recall.

Method

Design. The design was identical to that of Experiment 1.

Materials. New grammars were used in this experiment because some subjects had previously participated in either Experiment 1 or 2. Strings were generated from Grammars 5, 6, and 8 (see Figure 1). Each of these grammars had six vocabulary items rather than five. The assignment of letters to grammars was essentially random, although some adjustments were made to limit the number of visually similar letters assigned to the same grammar. For Grammar 5, the letters were: S, Q, C, X, L, V. For Grammar 6, the letters were I, Z, K, D, N, P. For Grammar 8, the letters were A, J, H, R, F, T. From each grammar, 30 strings were generated. They ranged from 4 to 9 letters, with two of 4 letters, four of 5 letters, four of 6 letters, eight of 7 letters, seven of 8 letters, and five of 9 letters. At letter lengths 4-7, half of the strings were assigned to the study list and half were assigned to the test list. Additionally, one of the
8-letter strings was assigned to the study list. The other six 8-letter strings and all five 9-letter strings were assigned to the test condition.

Two versions of each study string were necessary: the original version and an altered version. The altered version was created by adding two letters to each string according to the same rules as described in Method section of Experiment 1.

Each test string was also altered. A single letter was randomly removed and replaced by a dash. For example, the string CLLCBLWL was replaced by C–LCBWL. The original letter and a lure chosen randomly from the remaining five letters were used in the two-alternative forced-choice test procedure.

Procedure. The study procedure differed from that of Experiments 1 and 2. The subjects were told that a set of complex rules determined the make-up of each set of strings, but they were told that in addition to learning the strings for the later test of the rules, they would study the strings by learning to recall them. Each string was presented for 4 s, followed immediately by a recall prompt. The subjects then typed in what they could remember of the string and hit the return key. Depending on their response, the computer then signaled "CORRECT" or "INCORRECT." The same string then reappeared for another 4 s, again followed by a recall prompt. This cycle continued with the same string until it had been recalled perfectly twice, although the successful recalls did not have to be consecutive. Following the second successful recall, a new string appeared. As in Experiments 1 and 2, subjects in the informed and uninformed conditions were told that each string contained two ungrammatical letters, but the subjects in the informed condition were also told that each ungrammatical letter would have a black dot directly above it.

The test procedure was much the same as in Experiment 1, except that the subjects had to rate their confidence in their choice. Specifically, the subjects saw a test string with one letter missing, and they chose between two alternatives as to which fit
into the string. After this initial choice, the computer screen said, “Your answer was ___.
Rate your confidence.” The subjects then clicked one of three on-screen buttons labeled either “unsure,” “sure,” or “very sure.” After the subjects gave a confidence rating, the old test string was replaced by a new test string.

Subjects. The subjects were 24 Rice university undergraduates who participated as part of a course requirement or in return for extra credit in a psychology course.

Results and Discussion

Learning Data. The average number of trials required to reach the criterion of two flawless recalls was recorded for each subject. The baseline condition showed the fastest rate of learning, requiring an average of only 2.29 trials to reach criterion. The uninformed condition required an average of 3.05 trials, and the informed condition required an average of 3.63 trials. The difference between the baseline and uninformed conditions was reliable, $t(23) = 8.29, p < .001$. This result shows that the presence of irrelevant information made the strings harder to learn, perhaps only because they were two characters longer. The difference between the informed and uninformed conditions was also reliable, $t(23) = 3.29, p = .003$, suggesting that subjects were actively trying to ignore irrelevant information in the informed condition, thereby distracting them from the immediate recall task. Finally, the difference between the informed and baseline condition was also reliable, $t(23) = 7.12, p < .001$.

Test Data. The average probability of choosing correctly on a given trial was computed for the informed, uninformed, and baseline conditions. These means are reported in the fourth line of Table 1. Casual observation suggests that, at least in the uninformed and baseline conditions, performance was roughly comparable to that found in Experiment 1. Probability of choosing correctly in the baseline condition reliably exceeded the probability of choosing correctly in the uninformed condition, $t(23) = 5.33, p < .001$, showing that the ungrammatical letters impaired performance.
The more important question was whether the subjects were able to ignore the ungrammatical letters when informed of their locations. The results suggest that the subjects had limited success in this task: The advantage of the informed condition over the uninformed condition was small and not statistically reliable, $t(23) = 1.19, p = .12$. In contrast, performance in the baseline condition was reliably greater than performance in the informed condition, $t(23) = 4.01, p < .001$, indicating that in the informed condition the subjects' ability to ignore the ungrammatical letters was substantially limited.

Although these analyses did not indicate that subjects were able to willfully control their learning in this experiment, a more in-depth analysis did show some evidence of such effects. Because the subjects gave confidence ratings for each of their responses, these responses were used for additional analyses. Each response was first converted to a 6-point scale by assigning a 6, 5, or 4, to a correct answer rated "very sure," "sure," or "unsure," respectively, and a 1, 2, or 3 to an incorrect answer rated "very sure," "sure," or "unsure," respectively.

The average confidence scores for each condition are presented in the fifth line of Table 1. The basic pattern of results was unchanged, although certain comparisons reached conventional levels of statistical reliability where they did not before. Adding ungrammatical letters impaired performance, as indicated by the advantage of the baseline condition over the uninformed condition, $t(23) = 5.58, p < .001$. The critical comparison between the informed and uninformed conditions revealed a statistically reliable advantage for the informed condition, $t(23) = 2.40, p = .01$, implying that the subjects in this experiment were indeed able to exhibit at least some degree of willful control. The fact that performance in the informed condition was reliably worse than performance in the baseline condition, $t(23) = 3.36, p = .001$, indicates that the extent of willful control was limited.
A final analysis was conducted in which the confidence scores were analyzed repeatedly, each time according to a different criterion for a correct answer. In the first analysis, only a response of 6 was scored as correct. In the second analysis, either a 5 or 6 was scored correct. In the third analysis, a 4 or greater was scored as correct (which is equivalent to the correct/incorrect data reported previously). For the sake of completeness, even more lenient criteria were included: Thus, in the fourth analysis, a 3 or greater was scored as correct, and in the fifth and final analysis, anything other than a 1 was scored as correct. The mean probability of a correct response for each analysis was computed for each condition, and these means are plotted cumulatively in Figure 2. The darker the shading, the more strict the criterion. Note that, although not of interest in this report, each band of shading can be read as the proportion of responses falling into each of the six levels of the confidence scale.

In addition to the means, the three essential comparisons for this paradigm were made at each of the five criterion levels. These results of these analyses are shown in Table 2. The first comparison tested whether the ungrammatical letters impaired performance. In fact, at every criterion level, performance in the uninformed condition was worse than in the uninformed condition. The second comparison tested whether subjects could willfully block the effects of the ungrammatical letters. If so, there should be an advantage for the informed over the uninformed condition. Although the informed condition did show at least a numerical advantage at every level, this advantage was reliable only for the “≥5” level, the “≥3” level and the “≥2” level. The third comparison tested whether there were any discernible limits on subjects’ ability to ignore the ungrammatical letters. Limits on this ability are indicated by an advantage of the baseline condition over the informed condition. The advantage for the baseline condition was reliable only for the “=6,” “≥5,” and “≥4” criteria. What conclusions can be drawn from the results using these different criteria? Generally speaking, the pattern across criteria is stable, only becoming less pronounced at the more lenient
Figure 2. Probability of recall as a function of study condition for five response criteria that range from most stringent, "≥ 6," to least stringent, "≥ 2" for Experiment 3.
Table 2  
\textbf{Statistical Analyses for Each of Five Response Criteria in Experiment 3.}

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comparison</th>
<th>Statistics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$t(23)$</td>
<td>$r$</td>
<td>$p$</td>
</tr>
<tr>
<td>$\geq 6$</td>
<td>baseline $&gt;$ uninformed</td>
<td>3.76</td>
<td>.61</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>informed $&gt;$ uninformed</td>
<td>0.62</td>
<td>.13</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td>baseline $&gt;$ informed</td>
<td>2.18</td>
<td>.17</td>
<td>.02</td>
</tr>
<tr>
<td>$\geq 5$</td>
<td>baseline $&gt;$ uninformed</td>
<td>5.76</td>
<td>.77</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>informed $&gt;$ uninformed</td>
<td>1.83</td>
<td>.20</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>baseline $&gt;$ informed</td>
<td>2.30</td>
<td>.43</td>
<td>.02</td>
</tr>
<tr>
<td>$\geq 4$</td>
<td>baseline $&gt;$ uninformed</td>
<td>5.33</td>
<td>.55</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>informed $&gt;$ uninformed</td>
<td>1.19</td>
<td>.24</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>baseline $&gt;$ informed</td>
<td>4.01</td>
<td>.64</td>
<td>.00</td>
</tr>
<tr>
<td>$\geq 3$</td>
<td>baseline $&gt;$ uninformed</td>
<td>3.16</td>
<td>.30</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>informed $&gt;$ uninformed</td>
<td>1.72</td>
<td>.34</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>baseline $&gt;$ informed</td>
<td>0.90</td>
<td>.18</td>
<td>.19</td>
</tr>
<tr>
<td>$\geq 2$</td>
<td>baseline $&gt;$ uninformed</td>
<td>2.63</td>
<td>.48</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>informed $&gt;$ uninformed</td>
<td>2.13</td>
<td>.41</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>baseline $&gt;$ informed</td>
<td>0.85</td>
<td>.17</td>
<td>.20</td>
</tr>
</tbody>
</table>

\textbf{Note.} $p$-values are rounded to two decimal places.
criteria. Perhaps most noteworthy is that there is some evidence, albeit inconsistent, that identifying the ungrammatical letters did benefit subjects' performance.

An examination of the results using the various dependent measures suggests that the subjects probably had some willful control over their learning, even though they were required to recall each string, including the ungrammatical letters. This finding reinforces the claim that there is willful control in the learning of artificial grammars.

Experiment 4

The preceding experiments strongly indicate that subjects have some ability to ignore irrelevant information presented during grammar learning. One possible criticism of these experiments is that a true test of the nonwillful control position requires incidental learning of the grammar's rules, and in all of these experiments, the subjects were aware that the strings were rule-governed. Specifically, Reber (1989, 1990), Broadbent (1989), and others (Lewicki, 1986) might try to discount the results of the first three experiments by saying that the unconscious processes responsible for learning complex systems only come into play when subjects learn the rules of the grammar incidentally, although it seems reasonable that if rule-learning processes operate unconsciously, they should process whatever irrelevant information is encountered by subjects. The nonwillful control theories predict that even if there is control when subjects are intentionally learning rules, that control should be eliminated when the rules are learned incidentally.

What predictions are made by those theories that assume learning processes to be willfully controlled? Although the details of the theories differ, they could all be fairly described as memory-based. These theories suggest that, at test, subjects recall previously seen strings, or parts of strings, and use their recollections to make decisions about the test strings. Consequently, willful control theories predict that
subjects should be able to ignore ungrammatical letters, regardless of whether they are intentionally trying to learn rules at study. In fact, these theories do not give any reason to expect that subjects should have any limitations on their ability to ignore ungrammatical letters, although it is doubtful that any theorist would want to predict complete willful control.

The aim of the present experiment was to determine the extent of willful control under conditions of incidental learning. Theories advocating nonwillful control predict that subjects should have no control when the grammar is learned incidentally because they should not be able to direct unconscious processes to ignore the ungrammatical letters. In contrast, theories advocating willful control predict that because performance is solely a function of recollection, subjects should have considerable, if not complete, willful control regardless of intentional or incidental learning instructions. Thus, this experiment is an even more stringent test of the proposition that artificial grammar learning is beyond willful control.

Method

Subjects and Design. There were three between-subjects conditions: baseline, informed, and uninformed. The subjects were 45 Rice university undergraduates who participated as part of a course requirement or in return for extra credit in a psychology course. They were randomly assigned to the three conditions. Fifteen subjects served in each condition.

Materials. The strings used here were variations of some of the 45 strings used for Grammar 1 in Experiment 1. Specifically, each of the letters used in Experiment 1 was replaced with a randomly chosen (without replacement) digit from 1 to 5. Thus, CLLKCBWL became 23342153. Only 40 strings were used. One string at each of the five string lengths (7-11 letters) was randomly removed from the 25 test strings used in Experiment 1.
Procedure. The subjects were told that they would see a series of large numbers and that their task was simply to remember these numbers. The experimenter warned them that there would be a test at the end, but the nature of the test was not specified. In addition, the subjects in the informed condition were told that in each long number they saw, there would be two extra digits which “did not belong.” Each of the extra digits would be indicated by the presence of a black dot (“•”) directly above it. The subjects were told to learn the number “as though the dotted digits did not exist.” No mention of the ungrammatical digits was made to the uninformed subjects.

The digit strings were presented in 24-point Geneva font at a rate of 15 s per string. As soon as the last study string had been presented, the experimenter explained that the digits were arranged in strings according to a complex set of rules and that their knowledge of those rules would now be tested. As in previous experiments, the subjects saw each test string with a single digit replaced by a dash. They then chose which of two digits they thought fit in the string and rated their confidence in their answer as “unsure,” “sure,” or “very sure.” When queried, none of the subjects reported expecting a test on underlying rules.

Results and Discussion

The probability of choosing the correct answer was computed for each of the three conditions. The mean probabilities are reported in the sixth row of Table 1. As expected, the presence of ungrammatical digits did impair performance; the performance in the baseline condition clearly exceeded that in the uninformed condition, *t*(28) = 3.82, *p* < .001. Informing subjects of the location of the ungrammatical digits attenuated the effect of the ungrammatical letters, as shown by the reliable advantage of the informed over the uninformed condition, *t*(28) = 2.26, *p* = .01. Nevertheless, informing subjects did not eliminate the effect of the ungrammatical letters because
performance in the baseline condition reliably exceeded that in the informed condition, 
\( t(28) = 2.056, p = .03 \).

Subjects' responses were converted to the 6-point scale described in the the Results and Discussion section of Experiment 3. A 6 represented a correct answer rated "very sure" and a 1 represented an incorrect answer rated "very sure." The average confidence score was computed for each condition. These means are shown in the seventh row of Table 1. Including confidence scores did not change the pattern of results. The ungrammatical digits impaired learning, as shown by the fact that confidence scores in the baseline condition reliably exceeded those in the uninformed condition, 
\( t(28) = 3.78, p < .001 \). Also, the subjects were able to ignore the ungrammatical digits when informed of their locations, as shown by the fact that scores in the informed condition were reliably higher than those in the uninformed condition, 
\( t(28) = 2.27, p = .02 \). Finally, subjects' ability to ignore the ungrammatical digits was somewhat limited, as shown by the fact that scores in the baseline condition were appreciably higher than scores in the informed condition, 
\( t(28) = 2.03, p = .03 \).

The subjects' confidence ratings were also used to analyze the data according to different criteria for a correct answer. As in Experiment 3, the results were analyzed repeatedly, with an increasingly lenient criterion for a correct answer. The means for each condition at each response criterion are shown in Figure 3. The corresponding statistical analyses are presented in Table 3. At all but the most lenient criterion, there was an advantage for the baseline condition over the uninformed condition, indicating that the ungrammatical digits impaired performance. The advantage of the informed over the uninformed condition was not as consistent: Only at the "≥4" and "≥2" criteria did the evidence clearly indicate that the subjects were able to block the effects of the ungrammatical digits. Similarly, the superiority of the baseline condition as compared to the informed condition, which indicates a limitation on willful control, was found only at the "≥4" criterion.
Figure 3. Probability of recall as a function of study condition for five response criteria that range from most stringent, "≥ 6," to least stringent, "≥ 2" for Experiment 4.
Table 3

Statistical Analyses for Each of Five Response Criteria in Experiment 4

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comparison</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t(28)</td>
</tr>
<tr>
<td>=6</td>
<td>baseline &gt; uninformed</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td>informed &gt; uninformed</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>baseline &gt; informed</td>
<td>0.69</td>
</tr>
<tr>
<td>≥5</td>
<td>baseline &gt; uninformed</td>
<td>2.74</td>
</tr>
<tr>
<td></td>
<td>informed &gt; uninformed</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>baseline &gt; informed</td>
<td>1.64</td>
</tr>
<tr>
<td>≥4</td>
<td>baseline &gt; uninformed</td>
<td>3.82</td>
</tr>
<tr>
<td></td>
<td>informed &gt; uninformed</td>
<td>2.26</td>
</tr>
<tr>
<td></td>
<td>baseline &gt; informed</td>
<td>2.05</td>
</tr>
<tr>
<td>≥3</td>
<td>baseline &gt; uninformed</td>
<td>2.36</td>
</tr>
<tr>
<td></td>
<td>informed &gt; uninformed</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>baseline &gt; informed</td>
<td>1.56</td>
</tr>
<tr>
<td>≥2</td>
<td>baseline &gt; uninformed</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>informed &gt; uninformed</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>baseline &gt; informed</td>
<td>-0.75</td>
</tr>
</tbody>
</table>

Note. p-values are rounded to two decimal places.
The pattern of results found here was much the same as that found in Experiments 1, 2, and 3. In this experiment, however, learning of rules was incidental. If some “unconscious mechanism” controlled grammar learning, then subjects’ conscious efforts to ignore the ungrammatical digits should not have affected their performance. That these efforts did result in a large advantage of the informed over the uninformed condition argues strongly against the position that the learning of artificial grammars is beyond willful control.

**WILLFUL CONTROL AND THE AMOUNT OF IRRELEVANT INFORMATION**

The next four experiments examine, in one form or another, the effect of varying the amount of irrelevant information. Intuition would suggest that the more irrelevant information is added to the stimuli, the more difficult it will be to block that information.

**Experiment 5**

The purpose of Experiment 5 was to test the intuitively appealing notion that reducing the amount of to-be-ignored information would increase the extent of willful control. One way to reduce the amount of irrelevant information is to arrange it so that it forms a perceptual group (cf. Kahneman & Henik, 1981). For instance, research on the suffix effect shows that presenting several suffixes (i.e., irrelevant information) in succession does no more damage to memory than presenting only a single suffix (Crowder, 1978; Morton, 1976; Watkins & Watkins, 1982). One explanation of this finding is that the multiple suffixes are perceived as a group, and the group can be ignored just as easily as a single suffix.

Analogously, in this experiment, the subjects saw ungrammatical letters which were either distributed throughout each string, thereby impeding perceptual grouping, or clustered together, thereby facilitating perceptual grouping. To increase the probability that clustering would have an effect, four rather than two ungrammatical
letters were added to each string. The subjects were either informed of the locations of the ungrammatical letters or they remained uninformed. Thus, there were four conditions. The baseline condition used in Experiments 1-4 was dropped.

If clustering the ungrammatical letters leads to perceptual grouping, there should, presumably, be more willful control when the ungrammatical letters are clustered than when they are distributed.

Method

Design. Each subject participated in four conditions: clustered-informed, clustered-uninformed, distributed-informed, and distributed-uninformed. For each condition, the subjects saw strings from a different grammar. The four grammars appeared in the same order for all subjects; however, the condition to which each grammar was assigned was counterbalanced across subjects. Specifically, half of the subjects saw the two distributed conditions first, and the other half of the subjects saw the two clustered conditions first. Within each of these two groups, half of the subjects saw the informed condition followed by the uninformed condition for both the distributed and clustered conditions; the other half of the subjects saw the uninformed condition followed by the informed condition.

Materials. Four grammars were used for this experiment (Grammars 5-8, see Figure 1). These materials include all of those materials used in Experiment 3 plus additional, similar materials. For each grammar, there were 40 strings. These ranged in length from 4 to 9 letters, with 2 of four letters, 4 of five letters, 4 of six letters, 8 of seven letters, 12 of eight letters, and 10 of nine letters. At each letter length, half of the strings were assigned to the study list and half were assigned to the test list.

Two additional versions of each of the study strings were created by adding four ungrammatical letters. The choice of letters to insert was random except that an inserted letter could not have been produced by the grammar. In the distributed version the additional letters were inserted randomly with the constraint that no letter was ever
inserted before the first letter and none of the inserted letters appeared next to each other. In the clustered version, the ungrammatical letters were added as a block after one of the letters, chosen at random. The left-to-right order of the ungrammatical letters in the distributed version was duplicated in the clustered version. The unaltered version of each string was never presented to the subjects.

The test strings were prepared in the same manner as in Experiment 1. A single letter was randomly removed from the test string and replaced with a dash. The subjects saw the test string along with two choices, one correct and the other randomly chosen from the remaining four letters.

Within each grammar, the order of the study strings, as well as the order of the test strings, was the same for every subject.

Procedure. The procedure was generally the same as in Experiment 1. The experimenter described the nature of the letter strings and the subjects’ task. The subjects saw each of 20 study strings for 15 s. For the uninformed conditions, the subjects knew some letters were ungrammatical, but the locations of these letters were not indicated. For the informed conditions, each ungrammatical letter had a black dot directly above it during presentation.

During the test, the subjects saw 20 test strings. The subjects chose one of the two response alternatives provided with each test string and then provided a confidence rating as in Experiment 3. That is, the subjects rated each of their responses as “unsure,” “sure,” or “very sure.” The old test string was then replaced by a new test string.

Subjects. The subjects were 24 Rice university undergraduates who participated as part of a course requirement or in return for extra credit in a psychology course.
Results and Discussion

The probability of a subject choosing the correct alternative was computed for each of the four conditions. These probabilities are reported in the first panel of Table 4. General analysis of the data revealed that presenting the ungrammatical letters as a cluster resulted in slightly higher performance; however, the main effect of clustering was only marginally reliable, \(F(1, 23) = 2.86, \text{MSe} = .014, p = .10\). Averaging over the distributed and clustered conditions, informing subjects of which letters were ungrammatical resulted in an appreciable increase in performance, \(F(1, 23) = 13.16, \text{MSe} = .010, p = .001\).

More importantly, performance in the informed condition was reliably greater than performance in the uninformed condition, both when the ungrammatical letters were distributed throughout the string, \(t(23) = 2.46, p = .02\), and when the ungrammatical letters were clustered into a single unit, \(t(23) = 2.75, p = .01\), indicating that the subjects were able to block the effects of ungrammatical letters, regardless of whether or not they were clustered. Thus, there was willful control in both conditions. Nevertheless, the central question for this experiment was whether clustering the ungrammatical letters would lead to an increase in the extent of willful control. If so, there should be an interaction between clustering and informing. Contrary to expectations, there was no such interaction, \(F(1, 23) = .03, \text{MSe} = .009\). There was no evidence that perceptual grouping influenced the extent of willful control.

Because confidence ratings were collected, the data were re-analyzed in terms of their average confidence score on the 6-point scale described in Experiment 3. The means for each condition are shown in parentheses in the top panel of Table 4. Statistical analyses revealed a main effect for clustering, \(F(1, 23) = 5.13, \text{MSe} = .159, p = .03\): Confidence scores in the clustered conditions were somewhat higher than confidence scores in the distributed conditions. More important was the fact that
Table 4

Mean Probability Correct (and Average Confidence Scores) for Experiments 5, 6, 7, and 8.

<table>
<thead>
<tr>
<th>Experiment 5</th>
<th>Informed</th>
<th>Uninformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>clustered</td>
<td>.74 (4.38)</td>
<td>.67 (4.04)</td>
</tr>
<tr>
<td>distributed</td>
<td>.70 (4.19)</td>
<td>.63 (3.86)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment 6</th>
<th>Informed</th>
<th>Uninformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ungrammatical letter</td>
<td>.68 (4.18)</td>
<td>.65 (4.03)</td>
</tr>
<tr>
<td>2 ungrammatical letters</td>
<td>.67 (4.12)</td>
<td>.61 (3.85)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment 7</th>
<th>Informed</th>
<th>Uninformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ungrammatical letters</td>
<td>.64</td>
<td>.59</td>
</tr>
<tr>
<td>4 ungrammatical letters</td>
<td>.59</td>
<td>.51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment 8</th>
<th>Informed</th>
<th>Uninformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ungrammatical letters</td>
<td>.70 (4.13)</td>
<td>.60 (3.84)</td>
</tr>
<tr>
<td>4 ungrammatical letters</td>
<td>.68 (4.14)</td>
<td>.60 (3.78)</td>
</tr>
</tbody>
</table>
informing subjects as to which letters were ungrammatical led to an increase in confidence scores, $F(1, 23) = 23.05$, MSe = .116, $p < .001$. The advantage of the informed over the uninformed condition was reliable in the distributed, $t(23) = 4.32$, $p < .001$, and in the clustered conditions, $t(23) = 3.11$, $p = .003$, demonstrating that the subjects were able to ignore the ungrammatical letters. The extent of willful control was not, however, influenced by whether the ungrammatical letters were distributed or clustered: As in the previous analyses, the interaction between informing and clustering was not reliable, $F(1, 23) = .01$, MSe = .096.

Finally, an even more detailed analysis of subjects’ responses was conducted. As in Experiments 3 and 4, the results were analyzed repeatedly, with an increasingly lenient criterion for a correct answer. Such detailed analyses might reveal more subtle influences of the variables. The means at each criterion level are presented in Figure 4. The corresponding statistical analyses are presented in Table 5. The pattern of results was essentially the same at all criterion levels: Informing subjects of which letters were ungrammatical allowed them to block some of their effects, but the extent of this ability did not interact with whether the ungrammatical letters were clustered.

The data from this experiment are clear. Subjects were able to exert willful control in both the distributed and clustered conditions, and the clustering did not affect the extent of this control.

Experiment 6

The results of Experiment 5 were surprising. There are a number of plausible explanations for why willful control was not affected by clustering. For instance, the clustering manipulation may not have been sufficient to facilitate perceptual grouping; in which case, the subjects would have perceived the clustered conditions as containing four ungrammatical letters rather than just one. This experiment was intended to
Figure 4. Probability of recall as a function of study condition for five response criteria that range from most stringent, "≥ 6," to least stringent, "≥ 2" for Experiment 5.
<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comparison</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t(23)</td>
</tr>
<tr>
<td>≥6</td>
<td>distributed: informed &gt; uninformed</td>
<td>3.53</td>
</tr>
<tr>
<td></td>
<td>clustered: informed &gt; uninformed</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.09</td>
</tr>
<tr>
<td>≥5</td>
<td>distributed: informed &gt; uninformed</td>
<td>4.45</td>
</tr>
<tr>
<td></td>
<td>clustered: informed &gt; uninformed</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.48</td>
</tr>
<tr>
<td>≥4</td>
<td>distributed: informed &gt; uninformed</td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>clustered: informed &gt; uninformed</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.16</td>
</tr>
<tr>
<td>≥3</td>
<td>distributed: informed &gt; uninformed</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>clustered: informed &gt; uninformed</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.24</td>
</tr>
<tr>
<td>≥2</td>
<td>distributed: informed &gt; uninformed</td>
<td>-0.33</td>
</tr>
<tr>
<td></td>
<td>clustered: informed &gt; uninformed</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Note. p-values are rounded to two decimal places.
provide a more sensitive test of the hypothesis that the amount of irrelevant information has an effect on willful control.

In some conditions, subjects tried to ignore one ungrammatical letter; in others, they tried to ignore two ungrammatical letters. If increasing the amount of irrelevant information reduces the extent of willful control, then the advantage of the informed condition over the uninformed condition should be relatively smaller in the two-letter condition than in the one-letter condition.

**Method**

**Design.** Each subject participated in four experimental conditions, two informed and two uninformed. One of the informed and one of the uninformed conditions was presented with one ungrammatical letter, and the remaining two conditions were each presented with two ungrammatical letters.

Each condition was presented using a different grammar. All of the subjects saw the four grammars in the same order, but the conditions were rotated so that each grammar appeared equally often in each condition. Specifically, for half of the subjects, the first two grammars were presented with one ungrammatical letter and the second two grammars were presented with two ungrammatical letters. For the other half of the subjects, this order was reversed. In addition, within each of the two groups of subjects, half of the subjects saw the informed conditions first while the other half saw the uninformed conditions first. The strings of each grammar appeared in the same order for all subjects.

**Materials.** The strings were generated from Grammars 1, 2, 3, and 4 (see Figure 1). The 20 study strings and 20 of the 25 test strings created for the baseline condition of Experiment 1 for Grammars 1, 2, and 3 were simply used again. The number of test strings used for each grammar was reduced by randomly removing one 7-letter string, one 8-letter string, one 9-letter string, one 10-letter string, and one 11-
letter string. New strings were generated for Grammar 4 in the manner described in the Method section of Experiment 1.

The subjects never saw an original, unaltered string. Rather, for each study string, two alternate versions were created by adding ungrammatical letters in the same manner as in Experiment 1. One version consisted of one ungrammatical letter added to each string; the other version consisted of two ungrammatical letters added to each string. In both cases, each ungrammatical letter was inserted after a randomly chosen letter within the string. The only restrictions were that no two ungrammatical letters could be inserted after the same letter within the string and that no ungrammatical letter could fit the rules of the grammar.

Procedure. The procedure was generally the same as in Experiment 1. The subjects studied 20 strings. They were then shown 20 new strings, each of which had a letter missing. The subjects decided which of two alternatives fitted in the blank. The only differences between the present procedure and that of Experiment 1 were that each string was presented for 15 s rather than 20 s and, as in Experiment 3, after choosing an answer for each test item, the subjects rated their confidence in their response as "unsure," "sure," or "very sure."

Subjects. The subjects were 32 Rice university undergraduates who participated as part of a course requirement or in return for extra credit in a psychology course.

Results and Discussion

The data are summarized in the third panel of Table 4. Manipulating the number of letters had no discernible effect on performance, $F(1, 31) = 2.59$, MSe = .010, $p = .12$. On the other hand, informing the subjects of which letters were ungrammatical improved performance reliably, $F(1, 31) = 4.32$, MSe = .015, $p = .05$. The advantage of the informed condition over the uninformed condition was reliable when there were two ungrammatical letters, $t(31) = 2.20$, $p = .02$, indicating willful control. When
there was only one ungrammatical letter, however, the advantage of the informed condition was not reliable, \( t(31) = 0.82, p = .21 \). Hence, there is little evidence that the subjects were able to exert willful control when there was only one ungrammatical letter.

The fact that one comparison was reliable and the other was not seems to suggest that varying the number of ungrammatical letters influenced willful control, albeit in an unexpected direction. Before drawing such a conclusion, however, the data must be analyzed to determine whether the advantage for the informed condition was reliably larger in the two-letter than in the one-letter conditions. This interaction was not reliable, \( F(1, 31) = 0.74, MSe = .015 \). Thus, despite the fact that there was no evidence of willful control when only one ungrammatical letter was added, the advantage of the informed over the uninformed condition in the one-letter condition was not statistically distinguishable from that in the two-letter condition.

The confidence ratings were analyzed with the hope that they would clarify some of these equivocal results. The confidence ratings were converted to a 6-point scale, and the average rating for each condition was computed. These means are shown in parentheses in the second panel of Table 4. The results for these confidence scores generally resembled those for the probability correct data. There was a main effect of varying the number of letters, \( F(1, 31) = 4.25, MSe = .109, p = .05 \). The level of performance did not differ for the two informed conditions, \( t(31) = 0.69, p = .49 \), but for the two uninformed conditions, performance was better when there was only one ungrammatical letter, \( t(31) = 2.45, p = .02 \). There was also a main effect of informing subjects as to which letters were ungrammatical, \( F(1, 31) = 6.11, MSe = .226, p = .02 \). As with the earlier data, the informed advantage was reliable when there were two ungrammatical letters, \( t(31) = 3.41, p = .001 \), but not when there was only one ungrammatical letter, \( t(31) = 1.25, p = .11 \), although this effect was not far from
conventional levels of statistical reliability. There was clear evidence for willful control only when there were two ungrammatical letters. Nevertheless, as with the earlier data, the extent of willful control did not differ as a function of the number of ungrammatical letters, \( E(1, 31) = 1.37, MSe = .091, p = .25 \).

The confidence ratings were also used to establish different criteria for what was considered as a correct response. For each of the five criteria, means were computed and statistical analyses were conducted. The means for each mean at each response criterion are presented in Figure 5. The corresponding analyses are presented in Table 6. The analyses indicate that when two ungrammatical letters were present, the informed condition reliably exceeded the uninformed condition at all but one criterion, namely, the “\( \geq 3 \)” level. This single exception is not particularly informative. For the analysis of the one ungrammatical letter condition, the informed condition reliably exceeded the uninformed condition only for the “\( \geq 5 \)” and the “\( =6 \)” levels; in other words, given that the subjects were “sure” or “very sure” of their answers, informing them of the location of the ungrammatical letters did improve their performance. These analyses provide at least some evidence that subjects were able to exert willful control even when there was only one letter.

The central purpose of this experiment was to determine whether the number of ungrammatical letters would influence the extent of willful control. The critical test of this question was whether there was a statistically reliable interaction between the number of ungrammatical letters and informing subjects of the locations of these letters. Regardless of the response criterion, there was no reliable interaction. Hence, the null hypothesis cannot be rejected: Varying the number of irrelevant letters does not influence willful control.
Figure 5. Probability of recall as a function of study condition for five response criteria that range from most stringent, "= 6," to least stringent, "≥ 2" for Experiment 7
Table 6

Statistical Analyses for Each of Five Response Criteria in Experiment 6.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comparison</th>
<th>Statistics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t(31)</td>
<td>r</td>
<td>p</td>
</tr>
<tr>
<td>=6</td>
<td>one: informed &gt; uninformed</td>
<td>2.25</td>
<td>.37</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>two: informed &gt; uninformed</td>
<td>4.27</td>
<td>.61</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.42</td>
<td>.08</td>
<td>.68</td>
</tr>
<tr>
<td>≥5</td>
<td>one: informed &gt; uninformed</td>
<td>2.13</td>
<td>.36</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>two: informed &gt; uninformed</td>
<td>3.84</td>
<td>.57</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>1.16</td>
<td>.20</td>
<td>.26</td>
</tr>
<tr>
<td>≥4</td>
<td>one: informed &gt; uninformed</td>
<td>0.82</td>
<td>.15</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>two: informed &gt; uninformed</td>
<td>2.20</td>
<td>.37</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.86</td>
<td>.15</td>
<td>.40</td>
</tr>
<tr>
<td>≥3</td>
<td>one: informed &gt; uninformed</td>
<td>-0.56</td>
<td>-.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>two: informed &gt; uninformed</td>
<td>0.68</td>
<td>.12</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.91</td>
<td>.16</td>
<td>.37</td>
</tr>
<tr>
<td>≥2</td>
<td>one: informed &gt; uninformed</td>
<td>-1.56</td>
<td>-.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>two: informed &gt; uninformed</td>
<td>-1.73</td>
<td>-.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.42</td>
<td>.08</td>
<td>.68</td>
</tr>
</tbody>
</table>

Note. p-values are rounded to two decimal places.
Experiment 7

Experiment 6 failed to reveal any effect of varying the number of ungrammatical letters on willful control. Perhaps the relationship between willful control and the number of ungrammatical letters is not linear. Specifically, it might be that there two letters is no more difficult than one letter, but as the number of letters increases, ignoring the additional letters takes a toll. As a rough analogy, a person may be able to lift either 25 or 50 pounds easily but be completely unable to lift 100 pounds. The hypothesis that increasing the number of ungrammatical letters will diminish willful control was therefore tested again, this time using either two or four ungrammatical letters.

Method

Design. The design of this experiment was nearly identical to that of Experiment 6. The only difference was that instead of one or two ungrammatical letters in each string, there were either two or four ungrammatical letters in each condition.

Materials. The strings were generated from Grammars 1-4, as in Experiment 6. The only difference between this experiment and Experiment 6 was that there were 25 test strings for each grammar rather than 20. For Grammars 1, 2, and 3, the five additional strings were the unused strings from Experiment 1. For Grammar 4, five new strings were generated.

For each study string, two alternate versions were created by adding ungrammatical letters in the same manner as in Experiment 1. One version consisted of two ungrammatical letters added to each string; the other version consisted of four ungrammatical letters added to each string. The ungrammatical letters were chosen randomly, and each one was inserted after a randomly chosen letter within the string. The only restrictions were that no two ungrammatical letters could be inserted after the same letter within the string and that no ungrammatical letter could fit the rules of the
grammar. Hence, any two ungrammatical letters were always separated by at least one grammatical letter.

Procedure. The procedure was generally the same as in Experiment 1. The subjects studied 20 strings. They were then shown 25 new strings, each of which had a letter missing. The subjects decided which of two alternatives fitted in the blank. The only differences between the present procedure and that of Experiment 1 were that there were four rather than three conditions in this experiment and each string was presented for 15 s rather than 20 s.

Subjects. The subjects were 40 Rice university undergraduates who participated as part of a course requirement or in return for extra credit in a psychology course.

Results and Discussion

The data are summarized in the third panel of Table 4. Increasing the number of ungrammatical letters impaired performance, \( F(1, 39) = 12.59, \text{MSe} = .013, p = .001 \). The difference in performance between two and four letters was reliable both for the informed condition, \( t(39) = 2.41, p = .02 \), and for the uninformed condition, \( t(39) = 2.76, p = .009 \). Thus, adding more ungrammatical letters hurts learning which shows that this manipulation has at least some effect. As in previous experiments, informing subjects of which letters were ungrammatical did improve learning, \( F(1, 39) = 12.05, \text{MSe} = .014, p = .001 \). The advantage of the informed condition was evident both with two ungrammatical letters, \( t(39) = 2.32, p = .03 \), and with four ungrammatical letters, \( t(39) = 2.86, p = .007 \). The data indicated that subjects were able to ignore either two or four ungrammatical letters. The question of most importance, however, was whether the extent of willful control would differ as a function of the number of ungrammatical letters present in each string. It did not: No reliable interaction emerged between the number of ungrammatical letters and whether or not the subjects were informed, \( F(1, 39) = 0.80, \text{MSe} = .011, p = .38 \). The data did not support the hypothesis that the amount of irrelevant information influences willful control.
Experiment 8

Attempts to show that varying the number of ungrammatical letters affects the extent of willful control have been largely unsuccessful. Perhaps in the informed conditions, subjects use a learning strategy that somehow bypasses, at least in part, the ungrammatical letters. For instance, they might be repeating the grammatical letters of a string over and over, omitting all of the dotted letters. Omitting four letters might be no more difficult than omitting two letters. Presumably, then, omitting two letters hurts performance relative to a baseline condition with no ungrammatical letters (see Experiments 1-4) because subjects are able to use a more efficient strategy when no ungrammatical letters are present.

If the subjects usually avoid learning the ungrammatical letters by repeatedly articulating the grammatical letters, varying the number of ungrammatical letters is unlikely to have much effect on learning. If, however, they are prevented from articulating the grammatical letters, then varying the number of ungrammatical letters should, presumably, have an effect. Experiment 8 was designed to check this possibility. The use of a letter-repetition strategy was prevented by requiring subjects to articulate a nonsense syllable over and over while observing the study strings (cf. Levy, 1971). Specifically, they were required to silently repeat the syllable "blah" throughout the study phase. It was repeated silently so that several subjects could be tested together in a single room. In a given condition, subjects saw either two or four ungrammatical letters in each string, and they were either informed or uninformed of which letters were ungrammatical. If a repetition strategy were responsible for earlier failures to find an effect of varying the number of ungrammatical letters, then, because a repetition strategy is largely precluded in this experiment, such an effect should emerge. That is, the extent of willful control (i.e., the advantage of the informed over the uninformed condition) should be greater when there are two ungrammatical letters than when there are four ungrammatical letters.
Method

Design. Each subject participated in four conditions: informed with two ungrammatical letters, uninformed with two ungrammatical letters, informed with four ungrammatical letters, and uninformed with four ungrammatical letters. The counterbalancing scheme was the same as in Experiments 5, 6, and 7.

Materials. The original, unaltered versions of the strings generated for Experiment 5 (from Grammars 5-8) were also used in this experiment. There were 20 study strings and 20 test strings. The original version was never presented to the subjects; instead, two alternate versions of each study string were created: a version with two ungrammatical letters added, and a version with four ungrammatical letters added. The choice of letters to add to each string was random except that an additional letter could not have been produced by the grammar. The additional letters were inserted randomly with the constraint that no letter was ever inserted before the first letter and none of the inserted letters appeared next to each other. The test strings were the same as those used in Experiment 5.

Procedure. The procedure was similar to previous experiments. The subjects saw each string for 15 s. After seeing a set of 20 strings, they took a two-alternative forced-choice test in which they tried to decide which letter fit in the test string. After each choice, the subjects gave a confidence rating. The only major departure from previous procedures is that throughout each study period, the subjects were required to silently mouth the syllable “blah.” They mouthed at a rate of approximately two syllables per second. The experimenter monitored their compliance.

Results and Discussion

The mean probability of choosing a correct answer was computed for each condition. These means are presented in the bottom panel of Table 4. The number of ungrammatical letters present in each string had no reliable effect on performance, $F(1,$
27) = 0.11, MSe = .016. In contrast, informing subjects of the location of the ungrammatical letters improved performance appreciably, $F(1, 27) = 10.58, MSe = .021, p = .003$. Furthermore, the advantage of the informed over the uninformed condition was reliable both with two ungrammatical letters, $t(27) = 3.17, p = .002$, and with four ungrammatical letters, $t(27) = 2.48, p = .01$. The subjects were able to willfully ignore either two or four ungrammatical letters.

The critical question was whether the extent of willful control would be influenced by the number of ungrammatical letters. It had no effect in Experiment 6 or 7, but in this experiment, the subjects were discouraged from learning the strings by simply repeating the grammatical letters over and over. Nevertheless, the advantage of the informed over the uninformed condition was not discernibly influenced by the number of ungrammatical letters, $F(1, 27) = 0.26, MSe = .007$. Thus, even with articulation of the letter strings suppressed, the number of ungrammatical letters did not seem to affect willful control.

For the sake of completeness, the data were re-analyzed in terms of subjects' confidence scores. Confidence scores were averaged for each subject in each condition. These means are reported in parentheses in the bottom panel of Table 4. The results for these confidence scores simply repeated what was found for the previous data. There was no main effect of the number of ungrammatical letters, $F(1, 27) = 0.20, MSe = .108$. There was a main effect of informing subjects of the location of the ungrammatical letters, $F(1, 27) = 14.20, MSe = .214, p < .001$. The informed advantage was reliable for the conditions with two ungrammatical letters, $t(27) = 3.33, p = .003$, and for the condition with four ungrammatical letters, $t(27) = 3.37, p = .002$. The interaction of the two independent variables was not statistically reliable, $F(1, 27) = 0.32, MSe = .058$. In other words, willful control was evident both with two and four ungrammatical letters, and the extent of control was not reliably influenced by the number of ungrammatical letters.
The confidence scores were also analyzed according to response criterion, as they were in previous experiments. The means for each criterion are pictured in Figure 6. Statistical analyses of these data are summarized in Table 7. The pattern of results does not change much with different criteria. For the three strictest criteria ("≥6", "≥5", "≥4"), an informed advantage was evident both for the conditions with two ungrammatical letters and for the conditions with four ungrammatical letters. At the more lenient criteria, the advantage of the informed over the uninformed conditions was diminished or completely absent. The most meaningful finding to emerge from this analysis is that the difference between the informed and uninformed conditions did not differ as a function of the number of ungrammatical letters for any of the five response criteria.

**FURTHER EXPLORATIONS OF EFFECTS ON WILLFUL CONTROL**

Variations in the number of ungrammatical letters added to studied strings seemed to have had little effect on subjects' level of willful control. The final two experiments explore other factors that might be expected to influence the extent of willful control.

**Experiment 9**

When a teacher or a textbook presents examples of a problem or concept, the initial examples are often simplified. Complicating factors are removed because they might impede understanding. This technique might also apply to learning artificial grammars. Rather than immediately exposing subjects to strings with ungrammatical letters ("bad" strings), they could be initially exposed to strings without ungrammatical letters ("good" strings). Perhaps the early exposure to good strings would make it easier for them to ignore the ungrammatical letters when they later saw them in bad strings because they would know more about what constitutes a good string. It might therefore be easier to conceptually segregate the ungrammatical letters from the grammatical letters. In other words, if they initially saw only good strings, they might have more willful control.
Figure 6. Probability of recall as a function of study condition for five response criteria that range from most stringent, "≥ 6," to least stringent, "≥ 2" for Experiment 8.
Table 7

Statistical Analyses for Each of Five Response Criteria in Experiment 8.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comparison</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( t(27) )</td>
</tr>
<tr>
<td>=6</td>
<td>two: informed &gt; uninformed</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>four: informed &gt; uninformed</td>
<td>2.58</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.33</td>
</tr>
<tr>
<td>≥5</td>
<td>two: informed &gt; uninformed</td>
<td>2.68</td>
</tr>
<tr>
<td></td>
<td>four: informed &gt; uninformed</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>1.50</td>
</tr>
<tr>
<td>≥4</td>
<td>two: informed &gt; uninformed</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>four: informed &gt; uninformed</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.51</td>
</tr>
<tr>
<td>≥3</td>
<td>two: informed &gt; uninformed</td>
<td>2.06</td>
</tr>
<tr>
<td></td>
<td>four: informed &gt; uninformed</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.49</td>
</tr>
<tr>
<td>≥2</td>
<td>two: informed &gt; uninformed</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>four: informed &gt; uninformed</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Note. \( p \)-values are rounded to two decimal places.
Experiment 9 tested this hypothesis. The subjects either saw 10 good strings followed by 10 bad strings or they saw the 10 good and 10 bad strings mixed together randomly. The subjects were either informed or uninformed of the locations of the ungrammatical letters in the bad strings. In the informed condition, if no ungrammatical letters were indicated, the subjects knew the string was good. If limiting initial exposure to good strings enhances willful control, then the advantage of the informed over the uninformed condition should be greater when the good and bad strings are blocked then when they are randomly mixed.

Method

Materials and design. Each subject participated in four experimental conditions, two informed and two uninformed. For one of the informed and one of the uninformed conditions, the subjects saw 10 good strings followed by 10 bad strings. For the remaining two conditions, the subjects saw 10 good and 10 bad strings mixed randomly.

For each condition, the strings were drawn from a different grammar. The four grammars appeared in the same order for all of the subjects, but each grammar was assigned to each condition equally often. Specifically, for half of the subjects, the first two grammars were presented in the blocked condition and the second two grammars were presented in the mixed condition. For the other half of the subjects, this order was reversed. In addition, within each of the two groups of subjects, half of the subjects saw the informed conditions first while the other half saw the uninformed conditions first.

The strings used in Experiments 5 and 8 (generated from Grammars 5-8) were also used here. Two versions of each string were necessary: the original version (a "good" string) and a version in which two ungrammatical letters were added (a "bad" string). The ungrammatical letters were chosen randomly, and each one was inserted
after a randomly chosen grammatical letter within the string. The only restrictions were that no two ungrammatical letters could be inserted after the same letter within the string and that no ungrammatical letter could fit the rules of the grammar.

In each condition, the subjects saw 10 good strings and 10 bad strings. The experimental manipulation concerned the order in which the two kinds of strings were presented. In the blocked conditions, the first 10 strings were always free of ungrammatical letters and the last 10 strings always had ungrammatical letters. In the mixed conditions, the determination of which strings were presented with and without ungrammatical letters was random and separate for each subject. Furthermore, and unlike previous experiments, the order in which the 20 study strings occurred was randomized separately for each subject. The blocked vs. mixed manipulation was varied orthogonally within subjects with informed vs. uninformed conditions.

Procedure. The procedure was similar to that of the previous experiments. The main difference was that for every condition, some of the strings had no ungrammatical letters. For the uninformed condition, the subjects knew some letters in some strings were ungrammatical, but they did not know which letters were ungrammatical or even which strings contained ungrammatical letters. In contrast, for the informed condition, each occurrence of an ungrammatical letter was indicated by the presence of a black dot directly above it.

For each condition, there were 20 study strings, each of which was presented for 15 s, and 20 test strings. The test was the same two-alternative forced-choice test used in previous experiments. The subjects also gave confidence ratings for each response.

Subjects. The subjects were 24 Rice university undergraduates who participated as part of a course requirement or in return for extra credit in a psychology course.
Results and Discussion

The mean probability of choosing the correct alternative is reported for each of the four conditions in the top panel of Table 8. The effect of blocking was not statistically significant, $F(1, 23) = 0.19, MSe = .007$. Informing subjects of which letters were ungrammatical did have an effect, $F(1, 23) = 11.89, MSe = .008$. $p = .002$. Performance in the informed condition reliably exceeded that in the uninformed condition both for the blocked condition, $t(23) = 2.40, p = .01$, and for the mixed condition, $t(23) = 2.48, p = .01$. In other words, the subjects were able to exert willful control in both the blocked and mixed conditions. The key question in this experiment was whether the extent of willful control was greater for the blocked than for the mixed condition. No such interaction was found, $F(1, 23) = 0.48, MSe = .009$.

The subjects' confidence ratings were converted to a 6-point confidence score and averaged for each subject in each condition. These data are summarized in parentheses in the top panel of Table 8. The pattern of results is unchanged. There was no main effect of blocking, $F(1, 23) = 2.71, MSe = .054$. $p = .11$, but there was a main effect of informing subjects of which letters were ungrammatical, $F(1, 23) = 14.02, MSe = .143$. $p = .001$. The informed advantage was true both of the blocked conditions, $t(23) = 2.31, p = .02$, and of the mixed conditions, $t(23) = 3.42, p = .001$. Again, blocking the good and bad strings did not affect the extent of willful control, $F(1, 23) = .002, MSe = .128$.

As in previous experiments, the confidence data were analyzed several times, each time using a different confidence score as the criterion for a correct answer. The means for each condition at each response criterion are shown in Figure 7. Statistical analyses of the data at each criterion are presented in Table 9. For the three strictest criteria ("=6", "≥5", "≥4"), an informed advantage was evident both for the blocked conditions and mixed conditions. For the more lenient criteria, the informed advantage
Table 8

Mean Probability Correct (and Average Confidence Scores)

for Experiments 9 and 10.

<table>
<thead>
<tr>
<th></th>
<th>Experiment 9</th>
<th>Experiment 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Informed</td>
<td>Uninformed</td>
</tr>
<tr>
<td>mixed</td>
<td>.80 (4.64)</td>
<td>.73 (4.35)</td>
</tr>
<tr>
<td>blocked</td>
<td>.78 (4.57)</td>
<td>.73 (4.27)</td>
</tr>
</tbody>
</table>
Figure 7. Probability of recall as a function of study condition for five response criteria that range from most stringent, "≥ 6," to least stringent, "≥ 2" for Experiment 9.
### Table 9

**Statistical Analyses for Each of Five Response Criteria in Experiment 9**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comparison</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$t(23)$</td>
</tr>
<tr>
<td>=6</td>
<td>mixed: informed &gt; uninformed</td>
<td>2.64</td>
</tr>
<tr>
<td></td>
<td>blocked: informed &gt; uninformed</td>
<td>3.12</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.52</td>
</tr>
<tr>
<td>≥5</td>
<td>mixed: informed &gt; uninformed</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td>blocked: informed &gt; uninformed</td>
<td>2.78</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.00</td>
</tr>
<tr>
<td>≥4</td>
<td>mixed: informed &gt; uninformed</td>
<td>2.40</td>
</tr>
<tr>
<td></td>
<td>blocked: informed &gt; uninformed</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.69</td>
</tr>
<tr>
<td>≥3</td>
<td>mixed: informed &gt; uninformed</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>blocked: informed &gt; uninformed</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.19</td>
</tr>
<tr>
<td>≥2</td>
<td>mixed: informed &gt; uninformed</td>
<td>-0.69</td>
</tr>
<tr>
<td></td>
<td>blocked: informed &gt; uninformed</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.63</td>
</tr>
</tbody>
</table>

**Note.** $p$-values are rounded to two decimal places.
was either not statistically reliable or else there was no advantage whatsoever.

Critically, the effect of informing subjects did not interact with whether or not the good and bad strings were blocked, regardless of the criterion.

In summary, there was evidence for willful control with mixed and blocked presentation of the good and bad strings; however, there was no support for the hypothesis that limiting initial exposure to good strings enhances willful control.

Experiment 10

In the previous experiments, the subjects showed some ability to ignore ungrammatical letters when they were identified as such. Experiments 1-4 also showed that this ability was far from complete. In each of these experiments, there was clear evidence that there were limits on subjects' willful control. This experiment explored one way which might increase the extent of willful control. Specifically, it was an attempt to determine whether selecting the ungrammatical letters from a vocabulary different from the grammar would make it easier for the subjects to ignore the ungrammatical letters. Hence, in this experiment, the ungrammatical letters were marked with a black dot, but they were also different from the grammar in other ways.

The subjects were either informed or uninformed of which letters were ungrammatical, and the ungrammatical letters were either drawn from the vocabulary of the grammar or they were drawn from a different vocabulary. There were not enough letters in the alphabet for each of the four grammars to have a unique set of five letters plus a unique set of five letters to be used in the different-vocabulary condition. The different-vocabulary letters could not be re-used because the subjects might learn that these letters were always ungrammatical. Consequently, rather than using letters of the alphabet, various symbols available on the Apple Macintosh keyboard were used to represent each grammar. Thus, each grammar had five unique grammatical symbols and five unique ungrammatical symbols.
Method

**Design.** Each subject participated in four conditions, two informed and two uninformed. For one of the informed and for one of the uninformed conditions, the ungrammatical symbols inserted into each string were drawn from the same vocabulary as the grammar. For the other two conditions, the ungrammatical symbols were drawn from a different vocabulary from the grammar. The counterbalancing scheme was the same as in Experiments 5-8.

**Materials.** Grammars 1-4 were used for this experiment. The 20 study and 20 test strings for each grammar were identical to those used in Experiment 7 except that each letter in each string was replaced with a symbol. For Grammar 1, the symbols were Ω, β, Δ, ξ, and φ; for Grammar 2, the symbols were =, Σ, ζ, and ©; for Grammar 3, the symbols were ø, μ, f, δ, and §; and for Grammar 4, the symbols were γ, I, π, ©, and ¶. The original version of each string was never presented to subjects. Instead, they saw one of two altered versions. One altered version was created for the same-vocabulary conditions by adding two ungrammatical symbols. The symbols were added following the same rules used in previous experiments (e.g., see Method of Experiment 1). A second altered version was created for the different-vocabulary conditions by substituting a new symbol for each of the two ungrammatical letters in the same-vocabulary version. The new symbols were not drawn from the vocabulary of the grammar. Each symbol from the vocabulary of the grammar was always replaced by the same new symbol, and for each grammar the new symbols were different; hence, there was a one-to-one mapping between same and different symbols. For Grammar 1, the additional symbol set comprised ), %, ∧, !, and †; for Grammar 2, the additional symbol set comprised $, &, ∼, >, and .jms; for Grammar 3, the additional symbol set comprised /, +, #, ¥, and Ø; and for Grammar 4 the additional symbol set comprised å, @, ø, ≠, and ?.
As in previous experiments, in the informed conditions, the locations of the ungrammatical symbols were indicated by the presence of a black dot directly above the symbol.

Procedure. The procedure was generally the same as in previous experiments. The subjects were instructed that each set of strings was created by a complex set of rules and that during study they should try to look for patterns in the strings. Each string was presented for 15 s. After seeing a set of 20 strings, the subjects saw 20 new strings, each with a single symbol missing. The subjects chose one of the two alternatives with which they were presented as fitting in the test string. Immediately after making each choice, the subjects rated their confidence in their response as “very sure,” “sure,” or “unsure.”

Subjects. The subjects were 36 Rice university undergraduates who participated as part of a course requirement or in return for extra credit in a psychology course.

Results and Discussion

The average probability of a subject choosing the correct answer in each of the four conditions is presented in the bottom panel of Table 8. The overall level of learning was not reliably affected by whether ungrammatical symbols were from the same or from a different vocabulary, $F(1, 35) = 2.18$, $MSe = .011$, $p = .15$. Informing subjects of the location of the ungrammatical symbols did improve learning, $F(1, 35) = 16.22$, $MSe = .015$, $p < .001$. The informed advantage was true of the same-vocabulary conditions, $t(35) = 3.28$, $p = .001$, and of the different-vocabulary conditions, $t(35) = 2.19$, $p = .02$. Thus, subjects showed some degree of willful control in the both same and different condition. Was the extent of willful control increased by presenting the ungrammatical symbols in a different vocabulary? The data do no support that hypothesis, $F(1, 35) = 0.64$, $MSe = .017$. 

Using the confidence ratings, the subjects’ responses were converted to a 6-point confidence score. The average confidence scores were computed for each condition and are presented in parentheses in the bottom panel of Table 8. There was no main effect of vocabulary, $F(1, 35) = 2.80, \text{MSe} = .075, p = .10$, but there was a main effect of informing subjects, $F(1, 35) = 16.51, \text{MSe} = .165, p < .001$. There was an informed advantage when the ungrammatical symbols were from the same vocabulary as the rest of the grammar, $t(35) = 3.72, p = < .001$, and when the ungrammatical symbols were from a different vocabulary, $t(35) = 1.86, p = .04$. The advantage of informing subjects did not differ significantly as a function of whether the ungrammatical symbols were drawn from the same vocabulary as the grammar or from a different vocabulary, $F(1, 35) = 1.04, \text{MSe} = .202, p = .31$. Hence, there was evidence of willful control for both the same and different conditions, but the extent of control did not vary reliably between these two conditions.

Finally, the confidence ratings were used to determine different criteria for a correct answer. For each criterion, the mean probability correct at each criterion was computed for each of the four conditions. These means are shown cumulatively in Figure 8. These data were also statistically analyzed, and the results of these analyses are presented in Table 10. For the same-vocabulary conditions, there was a reliable advantage of the informed condition over the uninformed condition for the three strictest criteria, but not for the more lenient criteria. For the different-vocabulary conditions, however, the advantage of the informed condition was only marginally reliable for the two strictest conditions, not at all reliable for the “$\geq 3$” criterion, and not even in the predicted direct for the most lenient criterion. The middle criterion, “$\geq 4$”--which corresponds to the probability of the subjects choosing the correct answer, regardless of confidence rating--was reliable. Although analyses of the same-vocabulary data tended to show evidence of willful control and the analyses of the different-vocabulary data did not show much evidence of willful control, there was no
Figure 8. Probability of recall as a function of study condition for five response criteria that range from most stringent, "= 6," to least stringent, "≥ 2" for Experiment 10.
Table 10

Statistical Analyses for Each of Five Response Criteria in Experiment 10

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comparison</th>
<th>Statistics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$t(35)$</td>
<td>$r$</td>
<td>$p$</td>
</tr>
<tr>
<td>=6</td>
<td>same: informed &gt; uninformed</td>
<td>2.43</td>
<td>.38</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>different: informed &gt; uninformed</td>
<td>1.45</td>
<td>.24</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.87</td>
<td>.15</td>
<td>.39</td>
</tr>
<tr>
<td>≥5</td>
<td>same: informed &gt; uninformed</td>
<td>3.73</td>
<td>.53</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>different: informed &gt; uninformed</td>
<td>1.58</td>
<td>.26</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>1.08</td>
<td>.18</td>
<td>.29</td>
</tr>
<tr>
<td>≥4</td>
<td>same: informed &gt; uninformed</td>
<td>3.28</td>
<td>.48</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>different: informed &gt; uninformed</td>
<td>2.19</td>
<td>.35</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.80</td>
<td>.13</td>
<td>.42</td>
</tr>
<tr>
<td>≥3</td>
<td>same: informed &gt; uninformed</td>
<td>1.59</td>
<td>.26</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>different: informed &gt; uninformed</td>
<td>1.00</td>
<td>.17</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.31</td>
<td>.05</td>
<td>.76</td>
</tr>
<tr>
<td>≥2</td>
<td>same: informed &gt; uninformed</td>
<td>0.46</td>
<td>.08</td>
<td>.32</td>
</tr>
<tr>
<td></td>
<td>different: informed &gt; uninformed</td>
<td>-0.15</td>
<td>-.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>interaction</td>
<td>0.42</td>
<td>.07</td>
<td>.68</td>
</tr>
</tbody>
</table>

Note: $p$-values are rounded to two decimal places.
evidence of an interaction between vocabulary and information, irrespective of response criterion.

The lack of any effect of changing the vocabulary might be artifactual. Because the subjects participated in all four conditions, it is possible that those subjects who saw the informed-different condition followed by the uninformed-different condition realized that the ungrammatical symbols were different from the grammatical symbols. Consequently, they may have been able to identify the ungrammatical items in the uninformed condition, at least after seeing several strings. Any enhancement in willful control derived from choosing the ungrammatical symbols from a different vocabulary may have been masked. Such an enhancement should emerge if the subjects' first block of trials are analyzed separately. In other words, if the experiment is treated as though it were a between-subjects design, ignoring responses collected after the subjects saw the first grammar, the data could not be biased by knowledge that the subjects carried over from one condition to another. Even with this potential artifact eliminated, however, the pattern of results was essentially the same as with the total data. The advantage of informed over uninformed presentation was somewhat greater in the same-vocabulary condition (.65 vs. .57) than in the different-vocabulary condition (.66 vs. .62), although the interaction was not reliable $F(1,32) = 0.14$, $MSe = .024$, $p = .54$. Obviously, using a different vocabulary did not enhance willful control; the effect is in the opposite direction. Furthermore, the higher level of performance shown in the uninformed different-vocabulary condition relative to the uninformed same-vocabulary condition in the total data is also present in these between-subjects data; hence, this small difference should not be attributed to the effects of practice.

In summary, there was evidence that subjects were able to willfully ignore ungrammatical symbols, regardless of whether these symbols were drawn from the
same vocabulary as the grammar or from a different vocabulary. The hypothesis that
drawing the ungrammatical letters from a different vocabulary would enhance willful
control was not supported. Finally, this pattern of results cannot be attributed to
practice effects because they are found even in the first grammar that subjects learned.

SUMMARY AND CONCLUSIONS

The research reported here was concerned with the extent to which people can
willfully control their learning of complex systems. Some theories predict that subjects
should have no such control, whereas other theories predict that they should have
considerable or even complete control. Experiments 1-4 measured the extent of and
limitations on willful control using the nominal irrelevance technique. In two
conditions, irrelevant information in the form of two ungrammatical letters, was added
to each string. In one of these conditions the subjects were informed of the locations of
the ungrammatical letters during study; in the other, they were not informed. The
advantage of the informed over the uninformed condition served as a measure of willful
control. In these first four experiments, a third, “baseline” condition was added in
which no ungrammatical letters were added to the strings. Any disadvantage of the
informed condition relative to the baseline condition was evidence of a limitation on
willful control. Experiment 1 showed that although there was some degree of willful
control, it was less than complete. Experiment 2 replicated this result using a slightly
different test procedure. Experiment 3 replicated it using a recall procedure at study.
Experiment 4 replicated this pattern of results with an incidental learning procedure.
These four experiments showed evidence for a limited degree of willful control,
contradicting the position that the learning of complex systems is beyond willful
control.

The remaining experiments were an attempt to identify the conditions governing
willful control. Experiments 5-8 explored the effect of varying the number of
ungrammatical letters added to each string. Intuitively, it seemed likely that including
more ungrammatical letters would weaken willful control, but in none of these experiments was the expected interaction statistically reliable. In Experiment 5, four ungrammatical letters were added to each string, sometimes distributed throughout a string, sometimes clustered as a set. For both distributed and clustered conditions, the subjects were either informed or uninformed of which letters were ungrammatical. It was expected that if the subjects treated the clustered letters as a perceptual group, there would be a greater degree of willful control in the clustered condition. The difference between the informed and uninformed conditions was essentially the same in the distributed and clustered conditions. There was either a failure of perceptual grouping or reducing the amount of irrelevant information was not important. The next three experiments directly tested the effect of varying the number of ungrammatical letters. Experiment 6 compared one letter with two letters; Experiment 7 compared two letters with four letters. Varying the number of letters had little impact on willful control. Experiment 8 again compared two letters with four letters, this time preventing subjects from articulating the letters subvocally so as to preclude a strategy of rehearsing only the grammatical letters. Again, varying the number of letters had no effect on the degree of willful control.

Experiments 9 and 10 were attempts to find variables that would increase willful control. Experiment 9 tested the hypothesis that initial exposure to strings without ungrammatical letters would enhance subjects' ability to ignore ungrammatical letters when they were presented later. Experiment 10 tested the hypothesis that willful control would be increased if the ungrammatical letters came from a different vocabulary from the grammar. There was no evidence for either of these hypotheses.

A number of general conclusions can be drawn from these results. First, the data are fairly clear in indicating that subjects have a degree of willful control over their learning of artificial grammars. In all 10 experiments, there was evidence of an
advantage of the informed condition over the uninformed condition. In particular, Experiment 4 showed that even when the rules of the grammar were learned incidentally, irrelevant information could be ignored to some degree. If subjects had no willful control over the learning process, they should not have benefitted from having the irrelevant information identified.

Similarly, if learning were not under willful control, it should occur whenever the subject is exposed to complex stimuli. Accordingly, subjects in this study should have shown evidence for learning during the test phase of each experiment. The test procedure involved the presentation of grammatically correct strings with only a single letter missing. Processing this information should have improved subjects’ knowledge of the grammar throughout the course of the test. To check this possibility, in each condition of each experiment, the test items were divided into five equal sections, the mean score for each consecutive section was computed, and these means were submitted to a linear trend analysis. This test would reveal evidence of scores increasing (or decreasing) from the beginning to the end of the test. Only in Experiment 4 was there any evidence of a linear increase in performance over the course of the test. For the other nine experiments, the effect size3, r, of the linear trend was computed; the average effect size across all experiments was −.07, a slight decrease, on average.

In contrast, Experiment 4 showed a reliable positive trend in the baseline, t(14) = 2.57, p = .02, r = .57, and in the uninformed conditions, t(14) = 2.24, p = .04, r = .51. There was also a positive, but non-reliable, trend in the informed condition, t(14) = 1.54, p = .15, r = .38. These results suggest that in Experiment 4, the subjects acquired additional information about the grammar during the test period. Because they

3Specifically, r is the point-biserial correlation between scores and condition membership. For within-subjects design, the inter-correlation between subjects' scores in the two conditions are partialed out (Rosenthal & Rosnow, 1984, p. 313). Note that r can be directed computed from t and degrees of freedom (df) according to the formula: r = (t^2/(t^2+df))^{1/2}.
were initially unaware of the rule-based nature of the strings, it is understandable that these subjects would have tried to learn from the test items. Likewise, it is understandable why the subjects in the other experiments would not have tried: They probably felt they already understood the grammar. Assuming that subjects could decide whether or not to extract information from the test items offers a simple explanation for these data. It is not at all clear how these findings could be accounted for in terms of nonwillful or automatic processing.

A second conclusion to be drawn from the results of this study is that the extent of willful control is rather limited. Experiments 1-4 show that performance in the informed condition falls far short of that in the baseline condition. Complete willful control would be indicated by a lack of a difference between these two conditions. The subjects are unable to completely ignore the irrelevant information, even when it is clearly identified. Those theories of complex system learning that adopt a willful control position (e.g., Dulany, Carlson, & Dewey, 1984, 1985; Brooks & Vokey, 1991) do not offer an explanation for why there were any limitations on willful control. Nevertheless, in fairness to these theories, they simply assume that whatever information about the strings is encoded will contribute to performance, and information that is not encoded will not contribute. These theories could, therefore, explain the present results by modifying their encoding assumptions.

One implication of the fact that learning is not completely under willful control is that more attention should be paid to the characteristics of the stimuli. Some recent research has begun to address just this problem. Stadler (1992) analyzed the effect of structure of the stimuli in a serial reaction time task. In this task, subjects react to an on-screen stimulus by striking the corresponding key as quickly as possible. If the stimuli consist of a repeating sequence, subjects' reaction times decrease, even when they unaware of the repetitions (Cohen, Ivry, & Keele, 1990; Nissen & Bullemer,
1987; Stadler, 1989). Stadler (1992) showed that learning in this task was directly dependent on the extent to which a particular stimulus was statistically predictable from the stimulus that preceded it (cf. Shiffrin & Schneider, 1977). Stadler's (1992) study illustrates that characteristics of the stimuli can govern learning. The importance of stimulus characteristics in learning complex systems should not be overlooked. If, as the present study shows, complex system learning is not completely controlled by the efforts of the subjects, then a full understanding of this kind of learning requires that the nature of the system be taken into account as well.

A third conclusion to be drawn from these experiments is that the knowledge subjects acquire about artificial grammars is partially based on memory. The similarity in performance between Experiment 1 and Experiment 4 implies that subjects intention to learn rules is unimportant to grammar learning. Subjects treated Experiment 4 as though it were a simple memory experiment with the added requirement, in the informed condition, that they were to ignore certain stimuli. The fact that the subjects in Experiment 1 knew the strings were governed by rules did not seem to enhance their performance relative to that of the subjects in Experiment 4. It is possible that the similarity in results is coincidental: The processes and strategies used in the two experiments could have been very different. Only further research can address that question. Currently, the most parsimonious explanation of these results is that all of the subjects relied on memory for the studied strings. These data therefore tend to support the memory-based theories of Brooks (1978; Vokey & Brooks, 1992), Dulany, Carlson, and Dewey (1984, 1985), and Mathews (1991; Mathews et al. 1989).

A fourth conclusion to be drawn from this study is that the amount of irrelevant information has little or no effect on the degree of willful control. Of the four experiments to address this question (Experiments 5-8), not one showed evidence that willful control was influenced by varying the number of ungrammatical letters; that is,
in no experiment was there a statistically significant interaction between informing subjects and varying letters. Nevertheless, one might argue that a small effect existed and could have been found with a more powerful experiment. If such an effect existed, it might be detected by combining the results of these four experiments in a meta-analysis (Rosenthal & Rosnow, 1984). Rather than emphasizing significance tests, a meta-analysis focuses on effect sizes; hence, four non-significant results could still contain a trend for effect sizes to change as a function of the number of ungrammatical letters.

The steps of this meta-analysis were as follows. First, the difference between the informed and uninformed conditions was computed for each condition for each subjects. Second, each difference score for each subject in each experiment was assigned a 1 or a 0, depending on its condition membership (e.g., two ungrammatical letters or four ungrammatical letters). Next, the point-biserial correlation between the difference score and coded condition membership was computed. Note that because these were within-subjects design, the inter-correlation between subjects' scores in the two conditions (i.e., the variance due to subjects) had to be partialed out (Rosenthal & Rosnow, 1984, p. 313). The resulting Pearson's product-moment correlation coefficient, or \( r \), for each experiment represented the effect size for that interaction. For Experiments 5, 6, 7, and 8, \( r = .10, -.14, -.15, \) and \( .00 \), respectively, where a negative sign represents the fact that adding ungrammatical letters actually improved willful control--i.e., it increased the difference between the informed and uninformed conditions. An average effect size was computed simply by summing the rs and dividing by the number of studies (Rosenthal & Rosnow, 1984, p. 505). If Experiment 5 (in which there were always four ungrammatical letters, either distributed or clustered) is included, \( r = -.05 \); if Experiment 5 is excluded, \( r = -.07 \). In either case, the effect is exceedingly small, and its direction indicates that adding irrelevant information increases willful control very slightly. In summary, this analysis implies
that varying the number of ungrammatical letters has, at most, a minor effect on willful control.

The results of this study indicate that the subjects were able to ignore ungrammatical letters to some extent, but far from completely. Unfortunately, the present data say little about subjects’ partial success in learning the grammatical letters while ignoring the ungrammatical letters. Experiment 8 did test the hypothesis that subjects were repeating the grammatical letters to themselves and omitting the ungrammatical letters, but this conjecture was not supported by the evidence. Common sense might suggest that the subjects simply avoided looking at the ungrammatical letters, but the arrangement of the letters on the screen would have made selective attention of this sort very difficult. Research by C. W. Eriksen and his colleagues (Colgate, Hoffman, & C. W. Eriksen, 1973; B. A. Eriksen & C. W. Eriksen, 1974; C. W. Eriksen & Hoffman, 1973) has shown that people have little ability to ignore irrelevant visual information when it is in close proximity to relevant information in a visual search task. Thus, in order to see all of the grammatical letters, subjects could not avoid encoding the ungrammatical letters. To further explore the role of visual selective attention, the physical proximity of the letters within strings could be manipulated, or the letters could be presented sequentially. In any case, the data reported here do not shed much light on the question of precisely how the grammatical letters were learned while the ungrammatical letters were ignored.

Another interesting issue raised by these experiments concerns the apparent immutability of willful control. The degree of control that subjects managed to bring to bear was roughly the same over a variety of experimental manipulations. Although it is not possible to fully explain this constancy, relating it to a similar case may aid understanding. Another phenomenon which is hard to increase or decrease experimentally is the suffix effect, mentioned earlier in the context of the nominal
irrelevance technique. Normally, the suffix effect shows little evidence of willful control. It can be modified slightly by changing the physical characteristics (e.g., voice of presentation, location of origin) of the suffix item relative to the rest of the stimulus items; however, the extent of this modification is modest. The suffix effect is usually unaffected by an increase in the number of irrelevant items presented to the subjects. For instance, adding three suffix items to the end of a list does not increase the magnitude of the suffix effect (Watkins & Sechler, 1989; Watkins & Watkins, 1982). Some experiments have even shown a very small decrease resulting from additional suffix items (Crowder 1978; Morton, 1976). The present failure to find any effect of increasing irrelevant information is, therefore, not unique to the nominal irrelevance technique.

The present series of experiments did not directly address the role of consciousness in learning complex systems, perhaps the major focus of research on this topic; nevertheless, these experiment do speak to that issue, if only tangentially. If learning were unconscious, it would necessarily be nonwillful because willful control is, by definition, conscious. The data reported here show that learning can be controlled willfully, even the subjects are unaware of the presence of rules, as in Experiment 4. Therefore, the strong position that the learning of complex systems is completely unconscious is no longer tenable. Still, the possibility remains that some aspects of learning are accomplished unconsciously, especially in light of the fact that the present data show willful control to be limited. In general, however, the case for unconscious learning is weakened by these data.

All of these conclusions are based on experiments using a single experimental method and should, therefore, be generalized with considerable caution. The learning of complex systems has been explored using a number of experimental methods, including the dynamic control task and the reaction time task described in the introduction. Using these other methods to study willful control might yield different
results, which would be valuable because these tasks are often assumed to depend on the same underlying processes (e.g., Berry & Broadbent, 1988). If the extent of willful control differed substantially among these tasks, this assumption would have to be rejected. Still, as the review of the literature showed earlier, many findings do generalize across methods (e.g. Mathews et al., 1989 and Berry & Broadbent, 1988). Thus, it is not unreasonable to expect that the present findings regarding willful control will prove at least somewhat general.

What are the implications of these findings, assuming that they can be generalized? Perhaps the most important implication is that people have some control over their learning of complex systems. Specifically, people should be able to partially ignore “noise” in a system’s output, at least when the noise is readily distinguishable from the “signal.” For instance, when people learn to interact with a computer system, they are often exposed to aspects of the system’s “behavior” that are irrelevant to them. The present results suggest that if they were able to discriminate, to some extent, between the relevant and the irrelevant information, learning would be facilitated. Note that if this kind of learning were accomplished by an automatic process, identifying the irrelevant information would be useless.

In summary, the present series of experiments show that there is a limited ability to ignore irrelevant information when learning artificial grammars. This ability can be found even when the subjects are unaware that there are rules governing the strings they are learning. Furthermore, this ability to ignore irrelevant information seems to be relatively unaffected by the amount of irrelevant information present. It is also apparently unaffected by limiting initial exposure to error-free examples or by drawing the relevant and irrelevant information from different stimulus sets. In other words, the extent of willful control does not appear to be easily modified. These experiments support theories of complex system learning that claim knowledge is at least partly conscious and memory-based. Finally, these experiments illustrate the
utility of asking questions about the locus of control with respect to particular phenomena. This approach promises to reveal the importance of considering both willful and nonwillful control in a variety of experimental and real-life tasks. Ultimately, investigating the contributions of both willful and nonwillful control offers the opportunity for a more complete understanding of psychological processes.
References


