A Case of Syntactical Learning and Judgment: How Conscious and How Abstract?

Don E. Dulany, Richard A. Carlson, and Gerald I. Dewey
University of Illinois at Urbana–Champaign

This study examined two possible bases for grammatical judgments following syntactical learning: unconscious representations of a formal grammar, as in Reber's (1976) hypothesis of implicit learning, and conscious rules within informal grammars. Experimental subjects inspected strings generated by a finite-state grammar, viewed either one at a time or all at a time, with implicit or explicit learning instructions. In a transfer test, experimental and control subjects judged the grammaticality of grammatical and nongrammatical strings, reporting on every trial the bases for their judgments. In replication of others' results, experimental subjects met the critical test for grammatical abstraction: significantly correct classification of novel strings. We found, however, that reported rules predicted those grammatical judgments without significant residual. Subjects evidently acquired correlated grammars, personal sets of conscious rules, each of limited scope and many of imperfect validity. Those rules themselves were shown to embody abstractions, consciously represented novelty that could account for abstraction embodied in judgments. The better explanation of these results, we argue, credits grammatical judgments to conscious rules within informal grammars rather than to unconscious representations of a formal grammar.

What is conscious and what unconscious is a central question for current cognitive psychology. Indeed, that question has provided some of our deeper historical issues (Hilgard, 1980; Klein, 1977) and more controversial experimental literatures: subception and perceptual defense (Dixon, 1971), learning without awareness or attention (Brewer, 1974; Dulany, 1968; Kellogg, 1980), and several varieties of preconscious processing (Broadbent, 1977; Dixon, 1981). But nowhere is the claim for unconscious processes stronger, or more significant if true, than when the hypothetical processes are among the most complex of which we are capable—processes such as abstraction, inference, decision, and judgment. This is the claim for a fully psychological unconscious.

In an important series of articles, Reber (1967, 1969, 1976) proposed just such a process, which he called implicit learning, and he and his associates reported a series of experiments designed to investigate its behavior (Allen & Reber, 1980; Reber & Allen, 1978; Reber, Kassin, Lewis, & Cantor, 1980; Reber & Lewis, 1977). From many passages we extract what we consider to be the essence of the implicit process: According to Reber (1976), (a) information may be passively encoded by a "nonconscious abstraction system" (p. 276). (b) What is learned is "tacit knowledge" (Reber & Lewis, 1977, p. 355), an unconscious and abstract representation of structure in the information given. (c) The judgment that new information does or does not satisfy that representation is "implicit in our sense that [subjects] are not consciously aware of the aspects of the stimuli which lead them to their decision" (Reber & Allen, 1978, p. 218). (d) The process is evoked, it is said, when "subjects are not actively trying to break the code," "the stimulus environment exhibits exceedingly complex structure"
For their series of studies, Reber and associates designed unusually interesting experimental tasks, with finite-state grammars, realizing the aim of complexity while permitting a precise description of the material to be learned. Displayed in Figure 1 is a representative grammar used in Reber and Allen (1978) and also in the present study. With each transition from State \(i\) to State \(j\) or recursion on State \(i\), the system writes a letter. In this way the grammar generates a letter string for each permissible sequence of transitions and recursions from the entrance state to the exit state (Chomsky, 1963; Miller & Chomsky, 1963). Using this and very similar experimental tasks, Reber and his associates replicated the central finding over a variety of learning conditions: Given grammatical strings and implicit learning instructions, subjects are significantly accurate when they have a later opportunity to judge the grammaticality of novel grammatical and non-grammatical strings.

In our view, however, the basic and frequently replicated finding does not in itself provide strong evidence for (a) unconscious abstraction, (b) abstract and unconscious representation, or (c) grammatical judgment based on criteria outside the subjects' awareness. That interpretation would require the assumption that (d) the complex task and instructions to learn passively did in fact evoke the full implicit learning and judgment process. It is at least as plausible that information embodying rulelike regularities would evoke a strategy of learning and using conscious rules despite instructions merely to learn passively. We need taps into the process.

In order to see whether an implicit process might elude introspection, Reber and Allen (1978) obtained open-ended verbal reports after learning, then left the tape recorder running throughout the test while “subjects were encouraged to keep up a running commentary . . . to provide reasons and justifications for their judgments wherever they could” (p. 198). “In summary,” they reported, “our subjects tell us that the observation procedure tends to produce knowledge which is abstract in nature but which feels intuitive” (p. 203). Moreover, “learning occurs in the absence of explicit code-breaking strategies; our subjects cannot tell us very much about what they know” (p. 204). Despite this summation, the authors also reported the following:

Specific aspects of the letter strings were often cited as important in decision-making . . . first and last letters, bi-grams, the occasional tri-grams and recursions were mentioned. . . . Introspections after [observation learning] abound with references that have abstract rule-like qualities. Subjects refer to what can (and what cannot) be, what feels right (or wrong) and what is coherent (or not), (p. 202)

Nevertheless, they concluded the following:
subjects emerged with a small but solid body of articulated knowledge which they used to make decisions about the well-formedness of novel letter strings, [but they] also have a solid but tacit apprehension of the grammatical structure which serves them on those occasions when they have no conscious criteria. (pp. 211–212)

Although these reports provide valuable exploratory data, they seem insufficient to show whether conscious rules could explain the grammatical judgments. Subjects reported justifications of their judgments on 821 of their aggregate 2,000 trials. Of these 821, 694 were “appropriate” in the sense that “the subjects’ reasons for a response were an accurate reflection of the constraints of the grammar” (Reber & Allen, 1978, p. 209). At a glance, the rules would seem to underpredict correct judgments. Appropriate justifications on 694 trials with 50% guessing accuracy on 2,000 − 821 = 1,179 trials would explain only 694 + 589.5 = 1,283.5 of the reported 1,620 correct judgments. The trouble is this: As Reber and Allen (1978, p. 210) recognized, an appropriate justification may yield an incorrect judgment if something else in the string is nongrammatical; and an “inappropriate” justification may yield a correct judgment if the justification itself is in error. A two-valued metric—consistent or inconsistent with the grammar—does not adequately capture what a subject may use to produce a particular level of correct judgments. The problem we see with Reber and Lewis’s (1977, p. 352) use of reports is virtually identical. If we are to determine how well reported rules could explain judgments, we must know the probability of a correct judgment given the use of any rule reported. We provide such a validity metric for reported rules.

The present study is motivated by two closely related questions:

1. Do subjects acquire consciously available rules that could explain their grammatical judgments? If so, this would challenge the postulation of an unconscious and abstract representation that guides their grammatical judgments. In fact, we ask whether the rules acquired might constitute something other than the formal grammar that generated the strings. It is a cornerstone of modern linguistic theory that any finite set of strings can be generated by or classified by an infinite number of grammars (Pinker, 1979), as Reber et al. (1980, p. 500) recognized. Far as the formal grammar in Figure 1 is from the conscious ken of imaginable subjects, the grammar each subject acquires may very well be a set of informal and consciously available rules. Although each of the rules alone may be limited in scope, and many imperfectly valid, they may in aggregate be sufficient to explain the significant but imperfect levels of performance observed. We examine the possibility that subjects acquire *correlated grammars*, an alternative conception of what might be abstracted from the structure of a domain.

2. Do conscious rules themselves embody abstraction in the sense of novelty? The key and conventional mark of grammatical abstraction is the successful classification (or generation) of novel strings, strings unencountered before but nevertheless classifiable by a grammar that generates all and only grammatical strings. If subjects’ judgments meet this test for abstraction and are also predicted by rules, we ask what it is about those rules that may permit that significant kind of abstraction. If conscious rules could be a locus of abstraction, this would enhance the theoretical utility of any formulation giving a prominent place to those rules.

In short, how conscious and abstract is syntactical learning and judgment, where *how* may be taken to mean both in *what sense* and in *what degree*?

**Method**

**Design**

During the acquisition phase, experimental subjects were arranged within a 2 × 2 design, with between-subjects conditions of implicit or explicit learning instructions and strings viewed either one at a time or all at a time. Control subjects did not participate in the acquisition phase. All subjects participated in a test phase in which they judged the grammaticality of strings and reported rules.

**Task**

We selected a finite-state grammar used by Reber and Allen (1978), the grammar they felt most likely “to keep these explicit processes at a minimum” (p. 194). It generates relatively short strings, and “the longer a grammatical string becomes, the more salient the internal structure is and the more likely it becomes that subjects will
try explicitly to ‘crack the code’” (Reber & Allen, 1978, p. 194). Shown in Table 1 are the 20 grammatical strings presented at acquisition and the 25 grammatical and 25 nongrammatical strings presented at the transfer test. With an arbitrary limit of 4 on the number of possible recursions at States 2 and 5, the grammar in Figure 1 generates exactly the set of 40 strings used, plus 1 other arbitrarily discarded by Reber and Allen. They constructed nongrammatical items by introducing the grammatically impermissible letter substitutions and reversals underlined in Table 1.

**Procedure**

**Acquisition.** All subjects participated in groups of 15 to 18, with a very general introductory instruction calling for their attention, care, and cooperation throughout. In all experimental groups, too, the task called for observation learning only, as in Reber and Allen (1978) and Reber et al. (1980). This is the most passive of the learning tasks used in this work.

Critical features of the explicit and implicit learning instructions were taken verbatim from Reber (1976), with minor rephrasing elsewhere to accommodate the instructions with the following addition:

The order of letters in each item of the set you are about to see is determined by a rather complex set of rules. The rules allow only certain letters to follow other letters. Since the task involves memorization of a large number of complex strings of letters, it will be to your advantage if you can figure out what the rules are, which letters may follow other letters, and which ones may not. Such knowledge will certainly help you in learning and remembering the items.

In the *sequential* conditions, modeled closely on Reber and Allen (1978), each grammatical string was projected onto a screen for 10 s, and the set of 20 items was presented 3 times in a different order each time. In the *all* conditions, modeled closely on Reber et al. (1980), the entire set of 20 grammatical items was presented on a single slide for 10 min, resulting in a total presentation time equal to that in the sequential conditions. In both conditions, the order of presentation of the strings was randomized, with randomization constrained to avoid introducing structure that would make the grammar especially salient. Before starting, subjects were told how long they would view each string or all strings.

**Test.** The test phase immediately followed the acquisition phase in experimental conditions and immediately followed the introductory instructions in the control condition. Given a paper-and-pencil test, subjects judged the grammaticality of 100 strings, reporting in each instance the rule by which they judged the string to be grammatical or nongrammatical. The 50 unique test strings, consisting of 25 grammatical and 25 nongrammatical items, were repeated once.

Control subjects were given the preliminary information that, “another group of subjects saw items made up of the letters M, R, T, V, and X. We can’t tell you what they were, but the items ran from three to six letters in length.” All subjects were then instructed as follows (with the variant for controls in parentheses):

The order of letters in each item of the set you (they) saw was determined by a rather complex set of rules. The rules allow only certain letters to follow other letters. Half of the strings of letters on your answer sheets are well-formed according to the rules that generated the letter strings you (they) studied, and half of the strings violate those rules. When you look at your answer sheets, for each item (1) Look at the item and decide whether it is well-formed or violates the rules. (2) If you think the item is well-formed, underline that part of the item that makes it right. (3) If you think the item violates the rules, cross out that part of the item you think violates the rules.

For emphasis and understanding, those three points were repeated twice, in sequence and in close paraphrase; subjects were reminded to “be as careful as possible in marking each item.”

### Table 1

<table>
<thead>
<tr>
<th>Strings Presented During Acquisition and Test Periods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acquisition</strong></td>
</tr>
<tr>
<td><strong>Grammatical</strong></td>
</tr>
<tr>
<td>MTTTV</td>
</tr>
<tr>
<td>MTTV</td>
</tr>
<tr>
<td>MTV</td>
</tr>
<tr>
<td>MTVRX</td>
</tr>
<tr>
<td>MTRVX</td>
</tr>
<tr>
<td>MVX</td>
</tr>
<tr>
<td>MVRX</td>
</tr>
<tr>
<td>MVRXX</td>
</tr>
<tr>
<td>MVRXX</td>
</tr>
<tr>
<td>MVRXXX</td>
</tr>
<tr>
<td>MVRXXV</td>
</tr>
<tr>
<td>VXX</td>
</tr>
<tr>
<td>VXRR</td>
</tr>
<tr>
<td>VXRR</td>
</tr>
<tr>
<td>VXRRR</td>
</tr>
<tr>
<td>VXTRV</td>
</tr>
<tr>
<td>VXTV</td>
</tr>
<tr>
<td>VXVT</td>
</tr>
<tr>
<td>VXV</td>
</tr>
<tr>
<td>VXVX</td>
</tr>
<tr>
<td>VXVXV</td>
</tr>
<tr>
<td>VXVXV</td>
</tr>
<tr>
<td>VXVXV</td>
</tr>
<tr>
<td>VXVXVM</td>
</tr>
<tr>
<td>VXX</td>
</tr>
<tr>
<td>VXX</td>
</tr>
<tr>
<td>VXX</td>
</tr>
</tbody>
</table>

* Items in both acquisition and test periods. * Strings are underlined at point of grammatical violation.
In response to this instruction, subjects reported both their judgments of grammaticality and rules implying grammaticality. For example, they might view VXXTV and underline VXT and or they might view VXXRT and cross out T. In doing so, they would report a string to be grammatical by underlining and nongrammatical by crossing out. In doing so, too, they would report a conscious rule, a rule in which the feature marked is asserted to imply the grammatical status of the string. For any feature, in the sense of a letter or sequence of letters, subjects reported a rule of the form, “Feature i in the string implies that the string is grammatical” or “Feature i in the string implies that the string is nongrammatical.” The rule is a state of propositional awareness, with feature as the subject and asserted grammaticality as the predicate. Because they could convey a judgment of grammaticality by the form of a mark and a feature in the rule by what was marked, this procedure minimizes intrusiveness of assessments and eliminates lag between assessments. In an effort to solicit care in responding, we also required subjects to mark confidence in each trial’s response on a 7-point scale ranging from complete uncertainty to complete certainty.

The experimenter paced the test phase by asking the subjects to move to the next item every 30 s for the first five items, then every 15 s thereafter. To prevent looking back at earlier responses, possibly increasing the consistency of answers, subjects were instructed to slide their answer sheets into folders as each item was completed.

Subjects
Subjects were 65 undergraduates, participating as a requirement in an introductory psychology course. The data from an additional 7 subjects were discarded because of their failure to follow instructions, either by marking all test items grammatical or all nongrammatical, or by marking single items both grammatical and nongrammatical. This resulted in two groups with 13 subjects, the explicit—all and explicit—sequential conditions; two groups with 12 subjects, the implicit—all and implicit—sequential conditions; and a control group with 15 subjects.

Results and Discussion
Judgments and Conditions
A preliminary check for string-type bias or response bias showed that proportions correct were closely equivalent when judging grammatical and nongrammatical strings (M = .635 and .630, N = 65) and also when judging strings to be grammatical and nongrammatical (M = .632 and .632, N = 65). Subjects judged the strings grammatical on .503 trials and nongrammatical on .497 trials. We therefore compared subjects with respect to proportion correct as the dependent variable.

Did subjects learn in the sense that their grammatical judgments benefited from inspecting grammatical strings? In all four experimental groups, the mean proportion correct exceeded the control subjects’ mean of .555; M = .644, t(26) = 2.74, p = .011, for the explicit—all group; M = .645, t(26) = 3.38, p = .002, for the explicit—sequential group; M = .695, t(25) = 5.95, p < .001, for the implicit—all group; and M = .630, t(25) = 2.85, p = .009, for the implicit—sequential group. Evidently, a substantial number of individual subjects learned; 46% of the experimental subjects exceeded all but 1 control subject. Although Reber and Allen (1978) reported .81 and .80 correct judgments using “ten specially selected advanced undergraduates and graduate students” (p. 194), our range of .63 to .70 mean correct judgments closely replicates the .62 and .66 correct in Reber et al. (1980, Experiment I, Random Display, and Experiment II, Implicit Instructions). This is the preceding work with the most comparable procedures and subjects: observation learning and students from an introductory psychology course.

It is important for a test of learning to compare experimental subjects with controls because the mean proportion correct for controls, .555, significantly exceeded the expected value of .5, t(14) = 4.03, p = .001. There was no evidence, however, that control subjects learned within the test series. The mean proportions correct were .561 at the first presentation of the strings and .549 at the second, t(14) = .55.

Implicitly instructed subjects did not learn more than did explicitly instructed subjects. The mean values for accuracy of judgment were .653 and .645, respectively, F(1, 46) = 0.818, MSe = 0.005. Reber (1976) found that implicit instructions were more effective than explicit instructions when subjects were tested after learning strings to a criterion. Our results, however, replicate those of Millward (1980) and Reber et al. (1980, Experiments I and II), where learning procedures were more comparable to ours. Neither type of presentation nor its interaction with instructions produced a significant effect (M = .668 and .638, for the all and sequential groups, respectively), F(1, 46) = 2.33 and 2.65, p = .13 and .11.

Was there an error consistency effect, such that only the explicitly instructed showed an inflated consistency of error? For the two
presentations of each string, the mean proportions of judgments that were correct-correct, error-error, error-correct, correct-error, and the average of the two mixed cases are displayed in Table 2. According to Reber et al. (1980):

If subjects' knowledge of the grammatical structure is veridical but incomplete, . . . these three values [EE, EC, and CE] should all be the same. On the other hand, if subjects are systematically using nonrepresentative rules to make their decisions, an inflated value of EE will emerge. (p. 496)

On their view, syntactic judgments are imperfectly accurate for either of two reasons: implicit learning has yielded a structure of rules, veridical but incomplete or explicit learning has yielded a set of rules, some representative but some nonrepresentative. Thus, only explicitly instructed subjects should show an inflated consistency of error. Taking proportions of designated error pairs as the dependent variable, an analysis of variance (ANOVA) examined the effects of type of instruction, conditions of presentation, and types of error pair (error-error vs. average of mixed cases). The mean proportion of error-error pairs was .20, reliably greater than the average proportion of the mixed cases at .15, \( F(1, 46) = 16.27, p < .001, MS_e = 0.004 \). The explicitly instructed groups were not alone, however, in their inflated consistency of error, as shown by the lack of a Type of Error Pair \( \times \) Type of Instruction interaction, \( F(1, 46) = 0.549 \). All other effects were also nonsignificant (ps > .140).

Reber (1976) did find an inflated error-error rate with explicit instructions but not with implicit instructions. The two parts of the effect appear to come and go with conditions, however. Explicitly instructed subjects have sometimes shown an inflated error-error rate (Reber et al., 1980, Experiment I, Random Display, and Experiment II), but sometimes they have not (Reber et al., 1980, Experiment I, Structured Display). Implicitly instructed subjects have sometimes shown no inflated consistency of error (Reber, 1967; Reber & Allen, 1978, Observation Learning), but sometimes they have (Allen & Reber, 1980, Paired-Associate Learning; Reber & Allen, 1978; Reber et al., 1980, Experiment II). Nevertheless, we closely replicate the findings in previous work with task and subjects most like ours:

### Table 2

<table>
<thead>
<tr>
<th>Judgment</th>
<th>Explicit A</th>
<th>Explicit S</th>
<th>Implicit A</th>
<th>Implicit S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct-correct</td>
<td>.512</td>
<td>.563</td>
<td>.467</td>
<td></td>
</tr>
<tr>
<td>Error-error</td>
<td>.205</td>
<td>.160</td>
<td>.212</td>
<td></td>
</tr>
<tr>
<td>Error-correct</td>
<td>.125</td>
<td>.133</td>
<td>.138</td>
<td></td>
</tr>
<tr>
<td>Correct-error</td>
<td>.146</td>
<td>.144</td>
<td>.183</td>
<td></td>
</tr>
<tr>
<td>Average mixed</td>
<td>.136</td>
<td>.139</td>
<td>.161</td>
<td></td>
</tr>
</tbody>
</table>

*Note. A = all, S = sequential; both refer to conditions of presentation of the strings.*

Reber et al. (1980, Experiment I, Random Display, and Experiment II, Implicit Instructions). With explicit instructions, their values for proportions of consistent and mixed errors were .25 and .14, and ours were .211 and .143; both cases were reliably different. With implicit instructions, their corresponding values were .19 and .15, not reliably different by a low-power \( \chi^2(1) \) but virtually identical to our reliably different values of .186 and .149.

Although the effect, when found, might suggest interesting patterns of rules related to instructions, we would emphasize that the presence or absence of the effect is nondiagnostic for the basic question we raise: Are the controlling rules—whatever error pattern they generate—conscious or unconscious?

### Rules and Judgments

Each of the subjects' reports is a rule in the sense that (a) it names a feature that could appear in a set of strings and (b) it predicates a grammatical status of the string containing that feature. In order to see how well conscious rules may explain these grammatical judgments, we need a metric for the variety of essentially qualitative rules that subjects report. A very general scale is provided by the \( P(\text{Event is in the correct category} | \text{Feature } i \text{ is in the event}) \). In the categorization literature, we know this scale more concretely as the validity of a feature: for example, the validity of wings or feathers for predicting membership in the category birds. It is also the scale on which subjects of propositional rules are
arrayed in the theory of propositional control described earlier by Dulany (1968).

In the present task, the correct category is (a) the set of grammatical strings when a rule asserts that a feature makes the string grammatical or (b) the set of nongrammatical strings when a rule asserts that a feature makes the string nongrammatical. In this task, too, a feature is a letter or sequence of letters that subjects underline or cross out in reporting a rule.

In this way, a validity metric for our rules is given by the following two expressions:

1. When the subject reports a rule asserting that “Feature \( i \) implies that String \( j \) is grammatical (G),”

   \[
   \text{Validity} = P(S_j \in G | F_i \in S_j)
   \]

2. When the subject reports a rule asserting that “Feature \( i \) implies that String \( j \) is nongrammatical (NG),”

   \[
   \text{Validity} = P(S_j \in NG | F_i \in S_j)
   \]

Simply put, the validity of any rule is the probability that it correctly categorizes a string, given the presence of the feature it represents. If a subject acted on and validly reported a conscious rule on all 100 trials, the proportion of correct judgments should closely approximate the mean validity of those 100 rules. This relation should be reduced only by a small degree of “sampling” error, the degree to which the trials on which features are reported are a biased selection of the trials on which they occur.

How strong was that relation? For subjects in the experimental groups, displayed in Figure 2 is the scatter plot of the proportion of correct judgments and mean validity of rules, showing \( r = .83 \), with a slope of .99 and an intercept of .01. Rules asserting grammaticality and rules asserting nongrammaticality were closely equivalent in their mean validities (.665 and .649, respectively). The value of \( r \) is .87 for all 65 experimental and control subjects taken together.

Do rules predict grammatical judgments without significant residual? For each subject we computed the signed difference between the mean validity of rules and the proportion of correct judgments. It is shown in Figure 3 that this signed error of prediction scatters around zero, with 49 of 50 subjects within the .05 sampling limits of a hypothetical zero by the binomial test. If grammatical judgments are controlled independently of those rules,
the proportion correct should significantly exceed the value that their mean validities predict. This was true of only 1 of 50 subjects. Because this is a 1-subject finding unreplicated in 49 opportunities, it is most plausibly explained by statistical or procedural deviation.

If control by conscious rules is a process with generality, we should also expect the error of prediction to be relatively stable over conditions. Within an ANOVA, absolute difference between the proportion of correct judgments and the mean rule validity did not vary with instructions, type of presentation, or their interaction, $F(1, 46) = 1.29, 0.40$, and $0.47; MS_e = 0.0008$. The overall grand mean for error of prediction was .029.

Rules and Conditions

Did subjects acquire correlated grammars? Consider the set of rules each subject reported to be a grammar in the general sense of a set of rules that will classify a set of strings into grammatical and nongrammatical subsets. The mean validity of those rules is simply the proportion of trials on which that grammar and the finite-state grammar make the same classification, both grammatical or both nongrammatical, which is to say the degree to which that set of rules is a correlated grammar. If those grammars reflected learning, their mean validities should be greater for experimental subjects than for control subjects. The mean rule validity, over subjects' means over trials, was .648 for experimental subjects, reliably greater than .564 for controls, $t(63) = 4.86, p < .001$. To a significant degree, subjects did acquire correlated grammars from inspecting grammatical strings. Moreover, with rule validities as with proportion correct, neither type of instruction, conditions of presentation, nor their interaction was a significant factor in that learning, $Fs(1, 46) = 0.620, 0.020$, and $2.65; all ps > .1, MS_e = 0.004$.

Not only were most rules imperfectly valid, but each was also limited in scope. Remember that the denominator of the validity index is the number of strings that contain the feature named in the rule. It is the number of strings for which that rule could be used. Over all experimental-group subjects, that mean scope of rules was 4.17 out of 100 strings, significantly below the mean scope of 4.7 for control subjects, $t(63) = 2.73, p = .008$. Through learning, subjects acquired a larger set of rules, each with more limited scope, than were available to control subjects by guessing.

As a further profile of these rules, the mean number of letters as a function of the number
of letters in the string is plotted in Figure 4. Although on a general conception of a rule, a subject may report all features, yielding a rule with scope of one (two with repetitions)—and on 20% of trials subjects did—it is clear that mean rule size was substantially less than string length (see Figure 4). It is also shown in Figure 4 that rule length did increase with string length, $F(3, 192) = 153.04$, $p < .001$, $MS_e = 0.166$, and was greater with grammatical than with nongrammatical judgments, $F(1, 64) = 173.20$, $p < .001$, $MS_e = 1.66$. On these dimensions, experimental and control subjects were indistinguishable, showing some commonality in the effects of guessing and making use of prior learning. Learning, as we have shown, was reflected in the key difference to be expected: greater validity of the rules that experimental subjects reported.

### Abstraction

Does the observed learning reveal itself in an ability to classify novel strings correctly? Or do subjects merely recognize old strings when encountered again at test, and sufficiently so to explain the overall effect? On the 5 old strings alone (10 with repetition), experimental subjects were substantially superior to control subjects, $M_{\text{correct}} = .73$ and .543, $t(63) = 2.91$, $p < .005$. Excluding these old strings, we reanalyzed over the remaining 40 novel grammatical and 50 novel nongrammatical strings. The mean proportions of correct judgments computed over novel strings were comparable to those computed over all strings. Values for the experimental groups differed reliably from the value of .566 for control subjects in three of four comparisons, $M = .634$, $t(26) = 2.97$, $p = .006$, for the explicit–all group; $M = .641$, $t(26) = 2.61$, $p = .015$, for the explicit–sequential group; $M = .697$, $t(25) = 5.75$, $p < .001$, for the implicit–all group; but $M = .612$, $t(25) = 1.64$, $p = .13$, for the implicit–sequential group. Abstraction was a little more common when strings had been presented all at a time rather than sequentially, $M = .666$ and .627, $F(1, 46) = 4.50$, $p = .039$, although neither type of instruction nor its interaction with condition of presentation was a significant factor ($M = .655$ and .638, for the implicit and explicit groups, respectively), $F(1, 46) = 0.645$ and 3.485, $MS_e = 0.005$. In this most conventional type of test, subjects’ judgments revealed a significant degree of syntactical abstraction.

We have already shown that rules could have a part in explaining that kind of abstraction. They predict correct judgments on novel as well as on old strings. But do those
rules themselves embody abstractions? What we have asked about the classification of novel strings can also be asked about the asserted classification by rules naming novel features. Does correctness benefit from inspecting grammatical strings at acquisition? The conventional test for abstraction at the level of strings is simply extended analogously to the level of rules.

Among the 20 novel grammatical strings, 13 contained novel features in the sense of a positionally indexed letter or sequence of letters that did not occur in strings at acquisition. Those 27 novel features, consisting of 13 whole strings and 14 parts of those strings, appear in Table 3. Within each of 25 nongrammatical strings, the appropriate novel feature is the grammatical violation introduced by Reber and Allen (1978), which as such cannot have appeared in the exclusively grammatical acquisition strings. Both sets are previously unencountered features that could be used to guide a correct grammatical judgment. We therefore examined \( P(\text{correct assertion} \mid \text{novel feature in rule}) \) over those 76 strings \( [(25 \text{ NG} + 13 \text{ G}) \times 2] \), where NG = nongrammatical and G = grammatical. We considered a novel feature to be in the rule if exactly that feature and nothing else was marked. In all experimental groups, the mean value of this conditional probability reliably exceeded the mean value of .501 for control subjects; \( M = .674, t(26) = 2.52, p = .008 \), for the explicit–all group; \( M = .692, t(26) = 3.26, p = .003 \), for the explicit–sequential group; \( M = .740, t(25) = 4.07, p < .001 \), for the implicit–all group; and \( M = .652, t(25) = 2.10, p = .046 \), for the implicit–sequential group. The last group is the same one that failed to show abstraction at the level of judgments on novel strings. Although this appears to be the most appropriate index of rule abstraction, note that all experimental-group subjects exceeded control subjects in \( P(\text{correct assertion} \mid \text{old feature in rule}) \), \( M = .289 \) and .185, \( t(63) = 3.08, p = .003 \). Furthermore, experimental and control subjects were essentially alike in both \( P(\text{correct assertion} \mid \text{old feature in rule}) \) and \( P(\text{correct assertion} \mid \text{old feature in rule}) \), \( M = .603 \) and .576, .350 and .368; \( t(63), ns. \) As a consequence of learning, rules could guide judgments more correctly just because they represent novel features and prescribe judgments more correctly. In this way, abstraction embodied in rules could provide a natural account of abstraction embodied in judgments.

**Features and Grammars**

As a check on our construal of the reported features, we ran a set of supplementary analyses. When subjects marked a feature, they marked a letter or a sequence of letters in an exact position, counted from the front and back of the string. Is this what they meant or did they mean something less restrictive? A letter or sequence counted only from the front? Or could they simply have meant that letter or sequence of letters in any position? Whether we can construe features as reported or whether we must assume something less restrictive would make for somewhat different validity computations, because the denominator of the validity fraction is somewhat different for each conception. Now each of these three conceptions of a feature entails a conception of grammar describing the set of each subject's reported rules. To check our construal of a feature, we examined the comparative fit of these grammars to the rules subjects reported. As shown in Table 4, the grammar designated G3, construing features as subjects reported them, consistently provided the best fit by two pairs

---

**Table 3**

<table>
<thead>
<tr>
<th>Whole strings</th>
<th>Parts of strings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MTTTVT</td>
<td>1. MTTTV_</td>
</tr>
<tr>
<td>2. MTTTV</td>
<td>—TTVT</td>
</tr>
<tr>
<td>3. MTTVRX</td>
<td>3. MTTVR_R</td>
</tr>
<tr>
<td>4. MTVRXX</td>
<td>—TTVRX</td>
</tr>
<tr>
<td>5. MTTV</td>
<td>4. TVRXR</td>
</tr>
<tr>
<td>6. MVT</td>
<td>8. VRXRM</td>
</tr>
<tr>
<td>7. MVRXM</td>
<td>9. XRM</td>
</tr>
<tr>
<td>8. MVRXRM</td>
<td>10. XRRRM</td>
</tr>
<tr>
<td>9. VXRM</td>
<td>—RRRM</td>
</tr>
<tr>
<td>10. VXRRRM</td>
<td>11. VXXTT</td>
</tr>
<tr>
<td>11. VXTTT</td>
<td>—XTTT</td>
</tr>
<tr>
<td>12. VXVRXR</td>
<td>12. XVXRX</td>
</tr>
<tr>
<td>13. MTVRXV</td>
<td>13. _TVRXV</td>
</tr>
</tbody>
</table>

**Note.** Among the 20 novel grammatical strings, 13 contained novel features in the sense of a positionally indexed letter or sequence of letters that did not occur in strings at acquisition.
Table 4
Indexes of Prediction and Acquisition for Three Grammars (G₁, G₂, G₃) Based on Three Construals of a Feature

<table>
<thead>
<tr>
<th>Index</th>
<th>G₃* &gt; G₂* &gt; G₁*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction of P correct from M rule validity Correlation</td>
<td>.866</td>
</tr>
<tr>
<td>1 -</td>
<td>Pred. - Obs.</td>
</tr>
<tr>
<td>Relation of rule validity to experimental group vs. control group M difference</td>
<td>.084</td>
</tr>
<tr>
<td>Student's t</td>
<td>4.83</td>
</tr>
</tbody>
</table>

Note. P correct = proportion correct. Pred. = predicted judgment. Obs. = observed judgment.
* A feature is a letter or sequence of letters (a) irrespective of position in string, on G₁; (b) in a given position counted from first letter, on G₂; or (c) in a given position counted from first and last letter, on G₃.

of indexes: (a) That grammar (G₃) is the best predictor of grammatical judgments. This is true for the correlation of the mean validity of the rules with the proportion of correct judgments and also for the mean of 1 — absolute error of prediction. (b) On that grammar, too, rules respond most strongly to acquisition. This holds for the mean difference between experimental and control groups in the validity of rules and for Student’s t of that difference (properly a measure of relation). Because G₃ was the grammar on which validities were computed for data in Figures 2 and 3, these findings support the basic analyses.

Confidence and Conditions

Although assessment of confidence was designed to enlist care in responding, its non-specific form allowed subjects to report confidence in judgments or in rules or in both, and they might have done so variably. We note incidentally, however, that experimental subjects were more confident of their responses than were control subjects, M = 6.69 and 5.38, t(63) = 2.87, p = .006. Furthermore, neither type of instruction, condition of presentation, nor their interaction affected confidence, Fs(1, 46) = 0.013, 1.187, and 0.069, respectively, MSₑ = 2.309. Evidently, learning was manifested in greater confidence in responding, and the lack of differences in confidence over groups is consistent with the lack of differences in learning over groups.

General Discussion

Judgments, Conscious Rules, and Grammars

On this evidence, it would seem that subjects’ conscious rules could indeed be central to a process explaining their grammatical judgments. Not only do the validities of reported rules strongly predict correct judgments, but they do so without significant residual. This is our most basic finding. It challenges the hypothesis of implicit learning and judgment, in particular Principles b and c, on which subjects are said to be unconscious of their grammatical representation and unconscious of aspects of strings guiding their judgment. (Refer to the beginning of this article for Principles a–d.) The finding seems strong. It is multiply replicated over individuals. We have provided the suggested conditions (the complex task and implicit instructions), which should, based on Principle d, evoke the implicit process. Furthermore, other aspects of these results closely replicate findings in the literature.

To the degree subjects learned, they could be said to have acquired correlated grammars. For each subject, the grammar can be simply characterized as the union of two sets of rules defined on features, strings, and grammatical classifications, (F₁ ∈ S₁) → (S₁ ∈ G) or (F₁ ∈ S₁) → (S₁ ∈ NG), where a feature, F₁, refers to a positionally indexed letter or sequence of letters. It is a personal grammar; every subject may have, and does have, a different set of rules. It is a correlated grammar to the degree that the subject’s rules and the finite-state grammar make the same grammatical classifications. Moreover, we know the grammars reflected learning, at least to a degree, because the validities of those rules benefited when subjects had inspected grammatical strings.

But could these sets of rules be more reasonably characterized as parts of the finite-state grammar rather than as separate and correlated grammars? We think not, at least...
not in the relevant sense of that part of a finite-state grammar that generates judgments. A finite-state grammar consists of a set of nonterminal symbols (the states in Figure 1), a set of terminal symbols (the letters labeling the arcs in that figure), and a set of productions for the letter-writing transitions among those states. With the grammar realized as a finite automaton, it classifies a string as grammatical if it can generate that string, and it classifies a string as nongrammatical if it cannot (Chomsky, 1963; Hunt, 1975).

When we construed features in rules as letters or as letter sequences, we described conscious states representing terminal symbols but not nonterminal symbols. Furthermore, rules declaring nongrammaticality as well as grammaticality are integral to the grammar; they are alike in form and in predictiveness. Having said as much, we acknowledge that we have brazenly characterized as "grammars" what might not seem to be proper grammars at all. To us, however, the more interesting point is this: Sets of rules that are intuitively but unconventionally grammars predicted without significant residual to the conventional index of grammatical abstraction—classification of novel strings. And those rules appeared in consciousness. They consciously represent novel arrangements of terminal symbols, thereby gaining functions analogous to the nonterminal symbols in conventional grammars: a key role in generating novel judgments. Simply put, some judgments may go beyond the information given because consciousness goes beyond the information given.

Abstraction

How abstract are the rules subjects report? In one sense, not very. On the average, each rule had dominion over only four strings, the number containing the feature on which that rule could prescribe a classification. With a minimal scope of one (two with repetitions), the rule completely described a single string and became equivalent to the very concrete exemplar representation in Brooks (1978) as a special case. Learning, in fact, moved rules toward greater concreteness in this sense; the rules that control subjects reported had greater scope than those of experimental subjects.

In another sense, however, these rules embody abstractions that could have a significant place in explaining the key and conventional index of grammatical abstraction, success in classifying novel strings. By an index analogous to the conventional index, subjects used novel features in rules to imply correct classification of novel strings—and significantly so in reflection of their experience with grammatical strings during acquisition. Within our data, this appears to happen in two ways, both evidently as abstractions from what is remembered and observed to what is reported. One way could be called combinatorial abstraction. When correctly classifying novel grammatical strings, the novel feature implying grammaticality is in fact a combination of old features already seen in two or more grammatical strings at acquisition. (This simply follows from our definition of a novel feature: If any letter or sequence of letters within a novel feature had not appeared during acquisition, our algorithm would have identified that letter or sequence as still another novel feature.) Following observations of features in separate grammatical strings, observations of their joint occurrence may suggest that the string is grammatical. Another way could be called substitutive abstraction. When correctly classifying novel nongrammatical strings, the novel features substituted for old features that would have left the string grammatical. Following observation of one feature in a grammatical string, observation of its replacement may suggest that the test string is nongrammatical. Of course, too, subjects exhibited selective abstraction—an older sense of abstraction (Lashley, 1929)—when they used old features to classify old strings as grammatical.

When abstraction yielded rules for novel judgments, the process evidently operated on information from both the earlier acquisition strings and the current test strings, a possibility that contrasts interestingly with the view that grammars must be completely formed by prior learning alone. In this study, however, we directly examined only what reflects abstraction and predicts judgments, not the earlier abstractive process itself. A fuller explanation would, of course, call for a detailed description of the processes that yield those conscious rules embodying abstraction.
**Conscious or Unconscious Control**

For this experiment, the one clear prediction of a theory of unconscious control—more correct judgments than predicted by conscious rules—was clearly disconfirmed. But are there suitable auxiliary assumptions that would enable a theory of unconscious control to predict these results? Subjects could entertain rules and act on them—or act and entertain rules, either independently of, or in justification of, that action. Our procedures are not so constraining that subjects could behave in only one way, and the data should in some degree reflect selectively on those possibilities.

We should first recognize that these data would not be explained by mechanisms of unconscious control of judgments and independent guessing of rules. Notice in Table 1 that every string of one grammatical class contains letters and sequences of letters appearing in strings of the other grammatical class. On any subset of trials, random guessing should select many features appearing in both grammatical classes. When grammatical judgments are incorrect, the rules would have a mean validity greater than zero and would overpredict correct judgments. When grammatical judgments are correct, the rules would have a mean validity less than one and would underpredict correct judgments. It follows, therefore, that as the proportion of correct judgments increases beyond .5, random guessing would produce rules with mean validities that progressively underpredict. We know from Figure 2, however, that the proportion correct was related to the mean validity of rules by essentially unit slope (.99).1

Is there some more systematic way that rules might come to track judgments without controlling them? In fact, two possibilities come fairly readily to mind. On both possibilities, subjects unconsciously abstracted varying portions of the finite-state grammar, or some other abstract but correlated grammar, and then carried over its unconscious representation and judged grammaticality while still unaware of anything in the strings guiding their judgments. On one possibility, the reported rules are learned in parallel with the unconscious grammar and then are recalled as cued by the string at hand. On the other possibility, each rule emerges only after its accompanying judgment, as a conscious reconstruction of some aspect of the unconscious grammar. In either case, the conscious rules are said to be merely noncontrolling justifications of an unconsciously controlled judgment.2 We see two kinds of problems, however, with this pair of accounts:

1. For neither account do we find linking assumptions, which is to say a description of a process that would strongly relate assumed amounts of unconscious grammatical learning to the observed mean validities of reports. Consider the first account: On what additional assumptions would subjects who learn just enough of the unconscious grammar to control P correct judgments also learn just the set of conscious rules whose validities predict P correct judgments without significant residual? The problem for the second account is analogous: On what additional assumptions would subjects who learn just enough of that grammar to control P correct judgments find themselves able to use that grammar—which is, after all, said to be unconscious—to reconstruct just the set of conscious rules with validities that predict P correct judgments without significant residual?

2. At this stage of inquiry, too, the hypotheses of conscious and unconscious control

---

1 On A. S. Reber's (personal communication, February 25, 1984) convincing us that we should further examine the question, we simulated rule guessing on the computer with a program that randomly selected a featural rule for each judgment of each subject, simulating our 50 experimental subjects 100 times. The mean rule validity over 100 simulations did progressively underpredict as the proportion correct increased (slope = 1.65, intercept = −.38). Furthermore, all 100 simulations of the 50 subjects showed both progressive and overall underprediction of correct judgments from the mean validity of guessed rules. Our experimental subjects had a best-fitting function and an essentially zero (.004) mean error of prediction that stood clearly apart from the generated sampling spaces for guessing. A full description of the simulations will be submitted for publication.

2 In interesting ways this question parallels an earlier question about learning without awareness (Dulany, 1968; Brewer, 1974). These are the same "parallelist" and "emergentist" assumptions that behaviorists used to defend the theory of unconscious reinforcement from experimental challenge. Furthermore, forming correlated grammars parallels forming correlated hypotheses, an alternative to the process of learning by the automatic action of reinforcement.
have different statuses on an important consideration: the identification of postulated processes. For one kind of explanation, the controlling states are conscious and assessed. For the other, one postulates an unconscious and unassessed process that conveniently does all the work. The two problems are obviously connected: Without some constraining identification of the processes, the unconscious grammar can be asserted to take whatever values are needed, namely, to be present in each subject in just that degree coordinate with the correctness of judgments and the validities of rules. Every parameter is freely adjustable after the fact.

The difficulty for the augmented formulation is that it does not in fact predict these results—and could not, we think, in the absence of independent indexes of an unconscious grammar and still further process assumptions that would link that grammar to the validities of reported rules. It is also a challenge, of course, that may eventually be met in further work, perhaps aided by formulating the problem in this way. In view of this difficulty, however, we think the more tenable account of these data would simply say that reported rules predicted grammatical judgments because subjects were conscious of rules controlling their judgments. Furthermore, the explanation gains utility in covering rules that may generate novelty by representing novelty in consciousness.

Possible Extensions

We confronted the question of implicit learning where it was raised, focusing on a finite-state grammar that lacked a semantic interpretation. Generalization to natural language is obviously limited because natural language embodies a semantic component and a higher order grammar with hierarchical structures. The results may not, however, be irrelevant to learning and judgment with natural language material. Although semantic interpretation may aid the learning of an artificial grammar (Moeser & Bregman, 1972; Morgan & Newport, 1981), Morgan and Newport (1981) found that it was not essential to learning and that semantic representation of dependencies was of no additional benefit. Furthermore, English submits to a finite-state description within hierarchically organized constituents. Our results at least raise an interesting and researchable question: What, if any, may be the role of correlated grammars, with conscious rules of limited scope, when learning a semantically interpreted and hierarchically organized artificial language—and by further extension, a natural language? Much the same could be said and asked of structured domains of play and social ritual.

The generality of these results also seems to be challenged by very strong intuitions from within structured domains of the real world. We often intuitively judge the grammaticality of a sentence or the legality of a move or the propriety of an act without conscious access to the formal syntax of the domain. But let us turn the tables somewhat. It is an interesting possibility that each of those intuitions is one of a set of informal rules of limited scope and perhaps imperfect validity. The intuitions seem quite conscious. We know something that seems right or wrong, even when we don’t think of or know the proper rule from a formal system. With intuition reclaimed for consciousness, we would not disagree with Allen and Reber (1980, p. 178) that “decisions about the well-formedness of test strings are made largely on an intuitive basis.”

Because the average amounts of learning in this study and in earlier studies were relatively small, we should also ask how well this paradigm models syntactical learning and judgment at large. Even as individuals achieved up to 83% correct judgments, however, our results were the same: no significant error of prediction from conscious rules. Furthermore, the syntaxes of games, rituals, and second languages are often imperfectly mastered, and all late learning is preceded by early learning, with consequences interesting in their own right. What happens with automatization in very advanced stages of learning does indeed raise new questions, but they are questions with significant ties to the questions examined here. On one very common view, rules that were once conscious may continue to control but at an unconscious level. Consequently, if entirely automatic judgments do occur, we can ask whether they express the internalized syntax formalists have in mind or the automatized residue of informal rules the learner once had very consciously in mind. If the common
view is correct, these findings would suggest that unconsciously controlling rules may be more informal than has commonly been thought. But is this what really happens or is control passed to still other informal and conscious rules, perhaps at some higher level of syntactic or conceptual organization? A number of interesting questions are only suggested but not answered by the present experiment.

References

Received October 18, 1983
Revision received May 21, 1984