Journal of Memory and Language **41,** 223–252 (1999)
Article ID jmla.1999.2647, available online at http://www.idealibrary.com on IDEAL®

The Strategic Basis of Performance in Binary Classification Tasks: Strategy Choices and Strategy Transitions

Lyle E. Bourne, Jr., Alice F. Healy, and James T. Parker

University of Colorado

and

Timothy C. Rickard

University of California, San Diego

Three experiments were conducted using a probe procedure in binary classification tasks to track strategies and strategy transitions over blocks of trials and to determine the effects of task difficulty (easy or hard classification rule) and task rule familiarity (artificial rule or natural language rule) on classification performance and strategy use. Two primary response strategies were identified, one based on the use of a classification rule and one based on the retrieval of instances from memory. There was clear evidence of rule use during the early trials, but this rule strategy eventually was replaced by the instance strategy except when the easy natural language rule was in effect. Response time speed-up followed a power function that was specific to the strategy in use.

Researchers have come to accept the idea that performance in any learning or memory task might be governed at different times by different underlying mechanisms. Siegler and Jenkins (1989), for example, examined the performance of young children as they learned basic single-digit addition, observing that performance changes that occur with practice reflect a transition from the use of a counting algorithm (or rule) for computing the answer of a given addition problem to the retrieval of answers directly from memory. Similarly, Logan (1988) and Rickard (1997), in studies of pseudoarithmetic performance with adult participants, agreed that an algorithmic (rule-based) approach to answer production is required and used on initial problems but is replaced by direct answer retrieval as participants become more skilled at the task (see also Palmeri, 1997). These theorists differ, however, as to whether the two bases of performance, rule-based responding or instance memory, are attempted in parallel (Logan, 1988), with the faster procedure coming to dominate performance, or are selected and executed on a one-at-a-time basis (Rickard, 1997). Both Logan (1988) and Rickard (1997) have given detailed quantitative accounts of a transition from algorithmic (or rule-based) performance to instance-based performance, which apparently always occurs in these tasks if enough practice and experience is given with the instances or problems to be learned.

Strategy transitions have been reported occasionally in other research areas. For example, in studies of category learning, Homa, Sterling, and Trepel (1981) argued that participants adopt an instance memory strategy (essentially rote

This research was supported in part by Army Research Institute Contracts MDA903-90-K-006, MDA903-93-K-0010, and DASW01-96-K-0010 to the University of Colorado. The authors thank Bill Oliver for help with modeling the response time data, Nancy Sherrod for help with the preparation of experimental material and the testing of participants in Experiment 2, and Grant Sinclair for testing participants and analyzing data in Experiments 1 and 2. We are also grateful to John Anderson, Thomas Palmeri, Gordon Logan, Robert Nosofsky, Richard Carlson, and Barbara Malt for helpful comments on earlier versions of this manuscript.

Address correspondence and reprint requests to L. E. Bourne, Jr., Department of Psychology, University of Colorado, Boulder, CO 80304-0345. Fax: (303) 492-8895. E-mail: lbourne@clipr.colorado.edu.

0749-596X/99 $30.00

learning) when the number of instances per category is small (less than five) but rely on the formation of a prototype (in this case essentially a rule) to assign instances to that category when the number is large. Further, in the Homa et al. study, participants appear to begin the learning task with an instance approach, making a transition to a prototype-based approach when and if the number of instances becomes large (see Smith, Murray, & Minda, 1997, for more recent evidence on the transition to prototype-based responding). (Note, however, that Busemeyer, Dewey, & Medin, 1984, and Shin & Nosofsky, 1992, demonstrated that prototype abstraction mechanisms may not be needed to explain these results.) This putative transition from instance-based to rule-based performance also resembles the transition from specific bigram and trigram knowledge to grammaticality as a basis for categorizing instances of an artificial grammar (e.g., Meulemans & Van der Linden, 1997). In still other domains, Harvey and Anderson (1996; see also Lundy, Wenger, Schmidt, & Carlson, 1994; Singley & Anderson, 1989) have argued for the presence of a composition process, which produces a transition between declarative and procedural bases of response, paralleling the transition from instance memory to rule use.

In verbal memory tasks, Tulving (1985) and others (e.g., Gardiner, Gawlik, & Richardson-Klavehn, 1994) used an introspective reporting technique to show that correct recognition of previously-seen words sometimes depends on explicit remembering of those words but at other times is attributed to a mere feeling of knowing. Tulving (1985) argued that remembering and knowing are functionally independent routes to the correct response in a recall task, the former being more akin to instance (episodic) knowledge and the latter to (semantic) knowledge of rules or concepts. These studies not only identified different bases of response in the same task, but also implied the possibility of transitions over time or over practice from one basis to another. Indeed, recently, Conway, Gardiner, Perfect, Anderson, and Cohen (1997) demonstrated a transition in the reported basis of course content knowledge over a retention interval, from remembering at short intervals to knowing at longer intervals, using an adaptation of the Tulving remember/know introspective technique. Note that in all of these cases, the transitions observed, instance to rule, are essentially the reverse of the one reported by Logan (1988) and Rickard (1997), which involves a change from rule-based to instance-based responding.

Thus, apparently learners can and do change the basis of performance in a learning or memory task as experience with or practice on the task proceeds. The research just reviewed documented unidirectional transitions either from rule to instance or instance to rule. It is not inconceivable (see Anderson, Fincham, & Douglass, 1997) that multiple transitions might take place over learning trials or over retention intervals in some cases. That is, as in the task used by Homa et al. (1981), participants might begin with a rote learning strategy that gives way in time to a rule covering the correct response for all instances. But later still, as the participant becomes thoroughly familiar with the complete and delimited stimulus domain, rule-based performance might give way to direct instance retrieval, as in the tasks used by Logan (1988) and Rickard (1997). To date, there is no direct evidence for multiple, bidirectional strategy transitions within a single task. Indirect evidence of such multiple strategy transitions can be found in a recent study by Anderson et al. (1997), who proposed a multistage model of cognitive skill acquisition which provides a theoretical rationale for expecting these transitions based on the relative speed and accuracy with which the different strategies can be employed in a given task. The models of Logan (1988) and Rickard (1997) might also be extended to accommodate multiple strategy transitions in circumstances in which the learner, rather than being given the rule (algorithm) for responding, is required first to discover it. Neither Logan nor Rickard has as yet reported such an extension, however. In any event, this issue is worth pursuing, because a demonstration of multiple transitions is the kind of tour de force that makes it clear that learning is not simply a matter of just one fundamental kind of underlying process, as suggested, for example, by Logan and Stadler (1991) and by Kramer, Strayer, and Buckley (1990).

Strategy shifting raises another important theoretical issue. Newell and Rosenbloom (1981) proposed a Power Law of Practice to account for the process of acquiring skill in any task. The Power Law formalizes the relationship between trials of practice and time to make a correct response as a power function whose general form is:

$$T = AN^{-\beta}, \qquad [1]$$

where T is response time on any trial N, A is response time on Trial 1, and β is a rate of change parameter. Thus, the relationship between response time and trial number is linear in log-log coordinates,

$$\log T = \log A - \beta \log N, \qquad [2]$$

which provides a convenient though informal test by inspection of the theoretical fit. Power Law theory implies a general, systematic process underlying improvement in performance. Strategy changes, if they exist, would challenge that assumption and might produce significant variations in performance from a single monolithic power function. Indeed, Rickard's (1997) evidence for an algorithm to instance retrieval strategy transition in pseudoarithmetic is in part dependent on significant deviations in performance from an overall power function fit to his data, coupled with close power function fits to the data from each strategy separately. A similar finding has been reported more recently by Delaney, Reder, Staszewski, and Ritter (1998) in double-digit multiplication. It is noteworthy that in both of these cases, the power function fits for the two separate strategies differed only in intercept (i.e., in mean response time), not in slope, and the observed transitions were from the slower to the faster strategy. The overall mean response time differences between strategies in these experiments were large, which should be expected as performers transition from answer calculation for a problem like 23 × 17 to the direct retrieval of its product, 391, after many closely spaced repeated presentations. These findings by Rickard and by Delaney et al. not only challenge the applicability of an overall Power Law to skill acquisition data, but also serve to validate the existence of separable strategies and the use of strategy probing as a means of identifying them. They leave open the possibility, however, that, under some circumstances, power functions for distinguishable strategies might converge, as the learner develops disproportionately greater skill with one of the strategies. The results of the present experiments speak to this possibility.

Our initial goal in this research was to create a classification learning task in which strategy transitions are feasible and to develop a method for tracking them. In this study, we consider the use of a rule, of prior instances, or of pure guessing to be different strategies for such a task, adopting essentially the same logic as Tulving (1985) and Conway et al. (1997) for defining recall performance strategies or Delaney et al. (1998), Rickard (1997), and Rehder and Pennington (in press) for defining algorithmic and instance-based strategies in calculational tasks. By "strategy" we mean whatever procedure a participant identifies, retrospectively, as the source of a classification response, again following the logic of Tulving (1985), Rickard (1997), Delaney et al. (1998), and others. "Strategy" might not be quite the right term because it implies deliberate planful behavior over a series of trials or problem-solving attempts, whereas we measure performance on a trial-by-trial basis. But is it not inconsistent with the way the term has been used recently by others (e.g., Rickard, 1997; Delaney et al., 1998). In addition to these introspective reports, we use the separability of strategy specific power functions for response times as additional evidence for the selection and use by participants of different bases of response on different trials.

We needed a small and well-defined population of stimuli and a set of associated classification responses which might either be memorized or be organized by a rule. We chose to use initially 12 strings of three letters each as the stimuli, 6 of which were successive letters of the alphabet when put in proper order (e.g., HGF) and 6 of which were not successive (e.g., FGJ). These stimuli were presented one at a time, in random order, and the participant, not knowing the rule at the outset, was asked to give a “classification” response, *valid* (for alphabetic stimuli) or *invalid* (for nonalphabetic stimuli), to each. Feedback was provided after each stimulus and response. The participant was told that there were several ways of responding correctly, viz., by guessing, by remembering the classification of a particular stimulus instance from a previous trial, by knowing the classification by means of a rule, or by some other means. After each stimulus, response, and feedback combination, participants were asked to indicate the strategy used for or the basis of their response—guess, instance, rule, or other. We expected performance, measured by proportion of correct responses over blocks of trials and by response time, to improve monotonically. The idea was to relate any systematic characteristics of strategy reports to these measures of performance.

The initial experiment was intended to provide evidence on the reliability and validity of the strategy report procedure. As a check on reliability, we wanted to determine whether all or most participants produced the same pattern of overall performance and strategy use over an extended series of trials. As a check on validity, we examined the relationship between strategy use and overall performance to assess certain expectations about strategy transitions. In particular, we expected that participants might guess at the outset, but that guessing should vanish over early blocks of trials as participants memorized certain stimulus–response combinations or found an organizing rule. Note that this task lends itself to guessing, to remembering instances, or to generating answers by rule, but to little else. Therefore, we expected the “other” strategy option to be chosen minimally throughout practice. On the basis of prior work, one might expect participants to adopt a rote learning approach initially, but to make a transition to rule use, if and when a rule was identified. But such a transition might never occur for some participants. Indeed, the operable rule might never be identified by some participants. Or, if it is identified and used, participants might revert to instance use as the limited number of instances are repeated many times over a long series of trials. We expect that the strategy participants eventually adopt should be the one that allows for most rapid correct responding. Experiment 1 was designed in part simply to determine what actually happens in this set of circumstances.

In fact, the results of Experiment 1 demonstrated the adequacy of this procedure to produce different response strategies, to allow for strategy transitions, and to reveal interesting relationships between strategy usage and speed and accuracy of classification responses. The outcome of Experiment 1 raised further questions. First, many but not all participants discovered and used the alphabetic rule (i.e., valid strings consisted of successive letters of the alphabet when put in proper order). But, although it was simple on the surface, the rule either eluded some participants or was more difficult to use than instance memory. Overall response times were faster on instance-based trials as contrasted with rule-based trials. If we create a task based on an even simpler rule, we might enhance rule usage. The examination of a simpler rule is especially interesting in light of an argument by Anderson et al. (1997) that instance retrieval “is even more direct and faster than production rule use” (p. 945). To the contrary, rule use might be more direct and faster than instance retrieval in the case of some very simple rules.

Second, the task of Experiment 1 used a relatively artificial rule and stimuli, as was the case in the study by Anderson et al. (1997). Rule usage might be more common in familiar or naturalistic tasks. It seemed appropriate therefore to compare the letter string task to one in which a commonly used linguistic rule provided the basis of performance. For Experiment 2, we chose the normative rule for pronouncing the definite article *the,* either as “thuh” or “thee.” This rule, according to virtually all English dictionaries and grammar books, is: Use the pronunciation “thuh” with the schwa phoneme /ə/ before words (adjectives or nouns) that begin with a consonant sound, and "thee" with the phoneme /i/ before words that begin with a vowel sound. This rule is especially interesting because, although used consistently by most native speakers of English, it appears never to have been explicitly taught or learned by most individuals, who prior to any explicit training or feedback treat it more in terms of instances (or instance fragments) than as a single, abstract rule (Healy & Sherrod, 1994). The outcome of Experiment 2 suggested the further examination of a somewhat more difficult linguistic rule, the rule for distinguishing count from mass nouns in the English language, which we used in Experiment 3.

Overall, the goal of these experiments was to study the relationship between strategy usage and measures of performance in both simple and difficult and artificial and naturalistic (i.e., unfamiliar and familiar) classification learning tasks, to identify some of the empirical characteristics of strategies and strategy transitions, to explore the generality of these empirical findings, and to examine the results in the light of contemporary theoretical models for strategy choice and strategy–response relationships.

EXPERIMENT 1

The aim of Experiment 1 was to check on the feasibility of a straightforward, although somewhat artificial classification learning task as a method to demonstrate strategy-based performance, strategy transitions, and strategy-performance relationships. As noted earlier, this task allows for a variety of strategy possibilities, which include guessing, trying to remember particular instances, and formulating a concept or rule for knowing the answer. Do participants adopt a single strategy, instance, or rule throughout practice? Do they start with one strategy, but later shift to another? Do they show multiple transitions between strategies? With practice, we expected most if not all participants to gravitate toward either rule usage or instance memory, although, because the two strategies do not seem to differ in difficulty as widely as they do in mental multiplication (Delaney et al., 1998), which would dominate was unclear. On the basis of prior work, one might expect participants to adopt a rote learning or instance memory approach initially (Homa et al., 1981), make a transition to rule use, if and when a rule was identified, and then possibly return to instance memory with extended practice because the number of instances to remember is small (Logan, 1988; Rickard, 1997). Experiment 1 was designed to determine what participants actually learn and do in these circumstances.

We did not ask the participants to verbalize any strategy they used, rule-based or otherwise, but rather required them to select one of four strategy options after each classification response, which is similar to the memory-probing procedure of Conway et al. (1997) and the strategy-probing procedure of Delaney et al. (1998). In addition, on occasion, we required participants to respond to novel stimuli inserted into the trial sequence, both as a check on the strategy-probing procedure and as a measure of the generalizability of any rule that the participant might have acquired. Note that instance-based generalization was possible but was minimized by using novel stimuli that were highly dissimilar to the training stimuli.

Method

Overview and design. Each participant was trained on a single set of 12 strings of three letters (e.g., HGF) presented one at a time repeatedly over 30 blocks of practice. A block is defined as one presentation of each of the 12 stimuli in a random order. Participants were asked to indicate whether each letter-string is "valid" or "invalid" and each response was followed immediately by (correct/incorrect) feedback. Validity was determined by a rule unknown to the participant at the onset of the experiment. By the rule, a letter-string is valid only if it can be rearranged to correspond to a sequence of adjacent letters in the alphabet (e.g., HGF is valid because it can be rearranged as FGH). By this definition, six of the letter-strings were valid and six of the letter-strings were invalid. Also, every third block of practice (i.e., every 36 trials) was followed by a novel string of three letters. For the 10 novel stimuli that were used, five were valid and five were invalid. Responses to these novel stimuli were also followed immediately by feedback. Participant strategies were probed after the feedback was presented on every trial to determine whether they (a) guessed, (b) used a rule, (c) remembered the answer to the particular instance from a previous trial, or (d) did something else (other, unspecified). The dependent measures were the response times, the proportion of correct responses, and the strategies reported for training and novel stimuli.

Participants. Twenty volunteer undergraduate college students participated for credit in an introductory psychology course.

Materials and apparatus. Four Zenith personal computers and four Zenith monitors were used in conjunction with the Micro Experimental Laboratory (MEL) program v1.00 (Schneider, 1988) for presenting stimuli and tabulating data. Stimulus items were chosen pseudorandomly with the following constraints. All of the letters used were alphabetically greater than E and less than X. Each letter was used once across the set of valid items and once across the set of invalid items (so that participants would not be able to use the presence of individual letters as the basis of string classification). For the invalid items, there were no more than two missing letters between each pair of letters in the string when the three letters are arranged in order. Participants used specially marked keys to indicate whether a stimulus was valid or invalid and to select one of the four strategy options on each trial.

Procedure. Participants were tested individually or in pairs (with each participant in a pair using a different computer) in quiet rooms, free from outside distractions. The instructions, which outlined the procedure in detail, were read aloud to the participants:

> Today's experiment explores basic learning processes. You will be presented with strings of three letters presented one at a time on a computer. During the experiment, you will see each of twelve letter strings several times. You can think of these letter strings as examples of a secret spy code. Half of the strings are meaningful examples of the secret code (that is, they convey some information such as "all is well"). Half of the strings are nonmeaningful examples (that is, they are not part of the secret code and they convey no information). It is your job to learn which letter strings are meaningful (or *valid*) and which strings are not meaningful (or *invalid*). Note that your job is not to try and guess what each string means. Rather, you are simply to learn which strings are valid and which are invalid. There may be a relatively simple rule which will discriminate between valid and invalid strings and which will help you in this task. Or, you may be able to learn the correct response by simply memorizing which strings are valid and which are invalid. Whether you discover a rule or not, the overall goal is to perform as quickly and accurately as possible on each string. After making your response to each letter string, you will be asked to indicate which of the following strategies you used to respond. Your strategy options will be "Guess," "Rule," "Remember," or "Other." You should select the "Guess" option if you made your response having no idea whether the string was valid or not. Select the "Rule" option if there was some aspect of the letter string, taken as a whole, which led you to believe that the string was or was not valid. You should select "Remember" if you simply remembered that the string was or was not valid based on an earlier experience with it. Select "Other" if none of the first three strategies applies. Note that whether you get the answer right or wrong should not influence the strategy that you choose for that trial. Here we are interested in the strategy that you used, regardless of whether it produced the correct answer. Most of the letter strings will appear in every block of the experiment. Occasionally, though, a new string may appear that you haven't seen before.

Results and Discussion

The proportion of correct responses per block of training trials increased significantly, $F(29,551) = 17.12$, $MS_e = .15$, $p < .01$, from about chance in Block 1 to around 95% by the end of practice, as shown in Fig. 1. Changes in strategy use underlying this improvement in accuracy, in contrast, were quite complex. An analysis of strategy selection that includes all four strategy options creates a potential problem of nonindependence because participants must select one and only one strategy on each trial. Thus, increases in the use of one strategy necessarily force a reduction in the use of one or more of the other strategies. To mitigate this problem, our analysis of variance includes only the two most interesting strategies, rule and instance. This analysis reveals a significant overall difference in proportion of use between

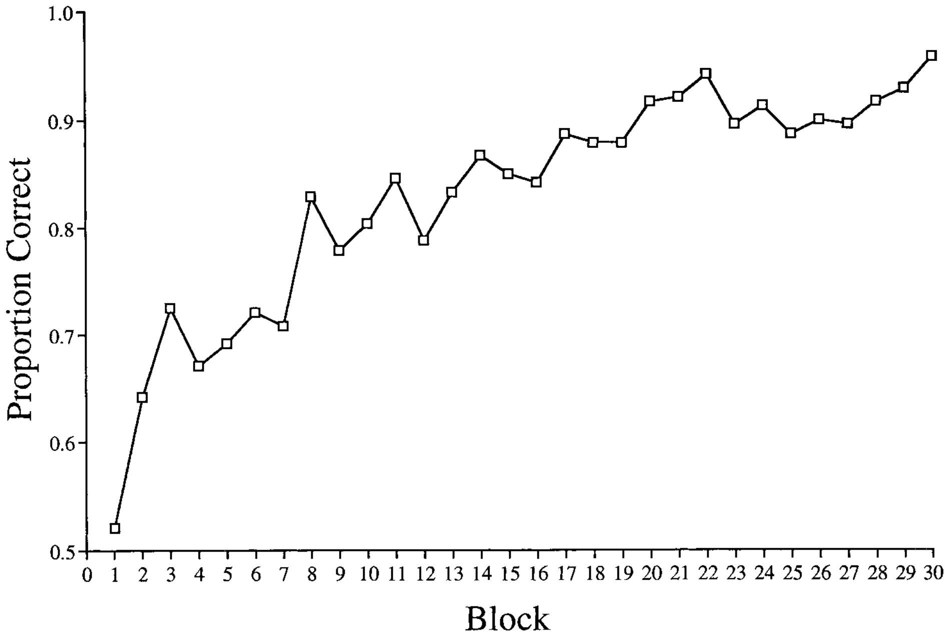

FIG. 1. Proportion of correct responses per block of training in Experiment 1.

instance (.67) and rule (.18), $F(1,19) = 39.64$, $MS_e = 1.87$, $p < .01$. It is worth noting that the proportion of use was lower for the guessing (.14) and other (.02) strategies. Figure 2 shows these proportions, collapsed across all participants and training problems and plotted as a function of block of practice. Although all participants guessed initially, many participants

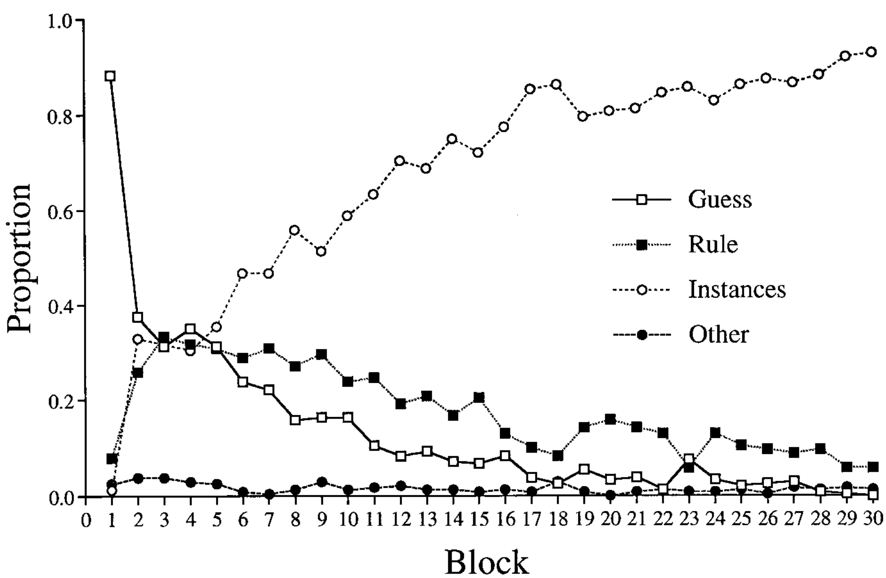

FIG. 2. Proportion of trials within each block of training on which participants reported each strategy in Experiment 1.

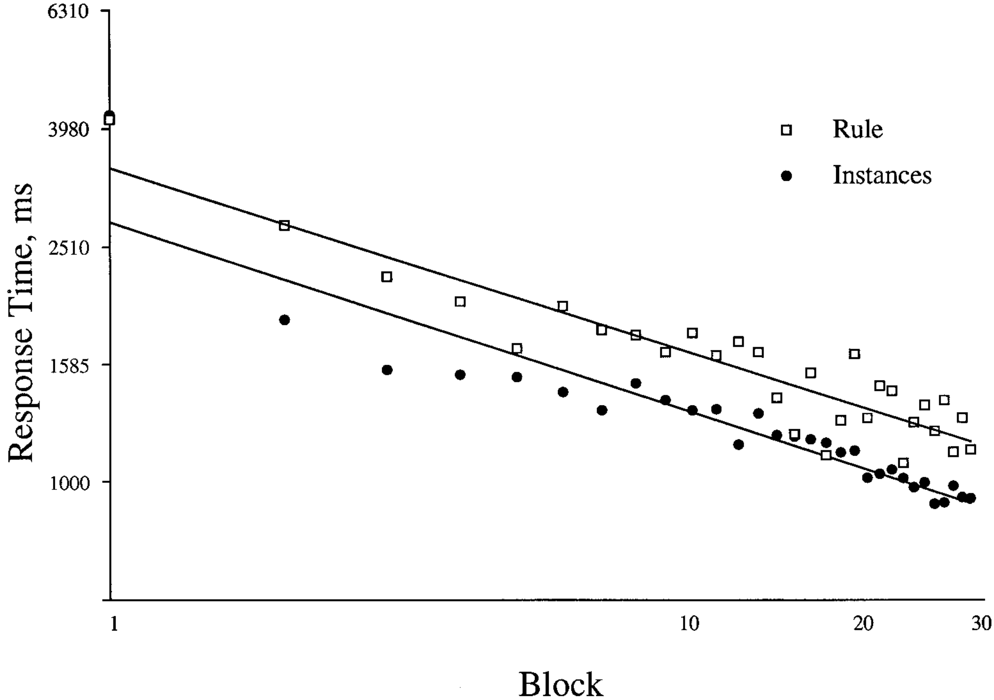

FIG. 3. Response times for correct responses trials during training as a function of block and the strategy reported, rule, or instance in Experiment 1.

discovered and used the alphabetical rule within the first three blocks (36 trials). However, by about Block 6, rule use began to give way to an instance strategy so that by the end of 30 blocks of practice, participants relied on their memory for instances almost exclusively. The Strategy × Blocks interaction was significant, $F(29,551) = 13.22$, $MS_e = .07$, $p < .01$. Clearly, the overall accuracy curve fails to reveal all the important changes in behavior that take place with practice.

The proportion of correct responses for the rule strategy (.971) was slightly higher than for the instance strategy (.958); an analysis of variance averaged across participants with blocks as the random effect showed this difference not to be significant. Although the difference was unstable, it was largest in the early blocks after the participants had begun to report rule use with some frequency (Blocks 7–14).

Consistent with accuracy changes across blocks, response times decreased with blocks of practice. But this overall representation of the data obscures another strong relationship between performance and strategy use. The data for each participant were divided into those trials on which the rule was reported and those on which instance memory was reported, using correct responses only. There is a speed-up process over both types of trials, which is consistent with a strategy-specific power law account of skill acquisition (e.g., Rickard, 1997). Averaged over all participants contributing to a given block, speed-up is approximately linear for each strategy on a log-log plot, as shown in Fig. 3. (Note that in these plots, the log values on each axis have been transformed back to normal scale. Also note that not all participants contributed to all points because not all participants used both strategies on all blocks.) Best fitting linear equations are

$$\log T = 3.443 - .325 \log N, \quad r^2 = .870 \text{ (for instance-based performance)} \qquad [3]$$

TABLE 1

Proportion of Strategy Reports as a Function of Stimulus Type (Experiment 1)

Stimulus type	Strategy: Guess	Rule	Instance
Training	.14	.18	.67
Novel	.45	.37	.18

$$\log T = 3.534 - .315 \log N,$$

$$r^2 = .856 \text{ (for rule-based performance)}, \quad [4]$$

where T is average response time for a block of trials and N is the trial block number. The slopes of these functions are nearly the same, differing in the second decimal place. But, there is a substantial difference in intercepts, reflecting the fact that over all trials, responses were slower when the rule-based as opposed to the instance-based strategy was used, $F(1,29) = 106.30$, $MS_e = .002$, $p < .01$, as would be expected if rule use is a less direct cognitive route to the correct answer than retrieval of the instance representation (Anderson et al., 1997). (This F statistic used blocks, not subjects, as the random effect because, as mentioned earlier, not all participants contributed to the response time functions for each block. Hence, to determine whether there was a difference between the two strategies, we used the Block X Strategy interaction as the error term.) The difference in mean response time between rule and instance strategy trials is, however, not nearly as large as the differences reported by other investigators (Delaney et al., 1998; Rickard, 1997) using quantitative calculational tasks.

Novel stimuli provide a check on the validity of strategy probing and an opportunity to examine the generalizability of the participants' acquired knowledge at any point in the training sequence. Table 1 shows the proportion of trials on which participants reported a guess, a rule-based response, or an instance-based response on the training stimuli and on the novel stimuli. (Other strategies were selected on less than 2% of all trials.) In an analysis restricted to the guess, rule, and instance strategies, the interaction of strategy by stimulus type (novel and training) is significant, $F(2,38) = 28.71$, $MS_e = .07$, $p < .01$. During training, guesses were relatively infrequent except during the first block or two, but guesses were reported more often than instance or rule strategies on novel stimuli. Likewise, rule use was higher and instance use was lower on novel stimuli relative to training stimuli, all of which would be expected if, in fact, participants generally recognized that novel stimuli had not been seen previously. The overall proportion of correct responses on the novel stimuli, all of which were unique and seen only once, was .64; this proportion did not change systematically or significantly over trials, $F(9,171) < 1$. The total number of choices of each strategy on novel stimuli, the proportion of correct responses, and the mean response time on correct responses for each strategy are shown in Table 2. In an analysis using novel instances (i.e., items) as the random effect, we found that rule-based performance produced significantly more correct responses than did instance-based performance and guesses, $F(2,18) = 11.06$, $MS_e = .04$, $p < .01$, which is consistent with the conclusion that many participants acquired the correct rule. In a similar analysis, there was no significant difference among the three most frequently used strategies in response times on correct responses, $F(2,18) = 2.02$, $MS_e = .01$, although, consistent with the training trials, the

TABLE 2

Responses to Novel Stimuli (Experiment 1)

	Strategy: Guess	Rule	Instance	Other
Total cases	89	73	36	2
Proportion correct	.46	.88	.61	0
Correct response time (in ms)	1919	2360	1879	na

trend is for rule-based responses to be slower than instance-based responses.

EXPERIMENT 2

Experiment 1 demonstrated that, in an artificial classification learning task where correct responses can be based on a simple deterministic rule that is not initially known by or described for the learner, overall performance improves while participants simultaneously report systematic changes in the basis of their responses. Are there analogous strategy transitions in a more naturalistic task where the rule is simple and familiar and in some sense known, implicitly or explicitly, to the learner at the outset? One purpose of Experiment 2 is to address this question. There is another, more important reason to look for strategy transitions in a task different from the one used in Experiment 1. We found in Experiment 1, as had Rickard (1997) and Delaney et al. (1998), that response times were significantly faster for the instance strategy than for the rule strategy. It is possible in this and in the experiments of Rickard (1997), Delaney et al. (1998), and others that participants simply reported "instance" whenever their response times were more rapid. If so, the strategy data would not reflect different cognitive processes, or bases of response, but instead they would only reflect perceived response times. If we use a task in which the rule generates a more rapid response than instance retrieval, then the foregoing possibility could be eliminated. This is another reason why we chose a simple, familiar naturalistic rule as the task in Experiment 2.

We adapted the task used in Experiment 1 to study changes by participants over trials in the pronunciation of the definite article *the* used in conjunction with a small subset of English words. As noted earlier, the rule is to use the pronunciation "thuh" before words (adjectives or nouns) that begin with a consonant sound and "thee" before words that begin with a vowel sound. This rule was not specifically mentioned in instructions to the participant nor is it commonly taught in school except in courses on English as a second language (although the analogous distinction for the indefinite articles, "a" and "an" is taught as a rule in school and is used by adults with very few errors). According to Keating, Byrd, Flemming, and Todaka (1994), however, individuals who speak a standard American English dialect generally, although not perfectly, conform to the *thee/thuh* pronunciation rule. Healy and Sherrod (1994) found that, when asked directly to choose between the two alternative pronunciations, all college students tested were individually somewhat inconsistent in their selections and conformed to the rule on average only about 75% of the time. This finding implies that this natural rule, though simple, is not known explicitly by most speakers and that individuals may even be resistant to learning and using it under ordinary conditions. Healy and Sherrod found, however, that rule usage can be increased to nearly 100% by response feedback over training trials. We wondered whether rule usage, encouraged by feedback, might eventually give way, as in Experiment 1, to instance-based performance when the number of instances involved in training (practice) is small.

Although the formal rule for the *thee/thuh* pronunciation distinction is based on the classification of the initial sound of the word as a vowel or a consonant, Healy and Sherrod (1994) have provided evidence that the rule actually used by college students may be less abstract and more variable. For example, there was evidence in the Healy and Sherrod data that individuals based their pronunciation on the specific initial letters or syllables (e.g., individuals might say "thee" before the letter *i* or the syllable *in*-) rather than the abstract categories of vowel and consonant sounds. Participants in the present experiment might also use rules other than the formal classification-based rule. Although we did not specify to participants the level of rule that might be used, to distinguish rule (knowing the answer) use from instance (remembering the answer) use, we gave explicit instructions to respond with the instance strategy only when they remembered a previous occurrence of the same word in the trial sequence.

Method

Overview and design. Each participant was trained on a single set of 12 stimuli consisting of high-frequency words (nouns and adjectives) one at a time repeatedly over 30 blocks of practice with feedback on each trial. A block is defined as one presentation of the 12 stimuli in a random order. Participants were asked to indicate whether "thuh" or "thee" is the correct pronunciation for *the* if it were to precede the stimulus word presented on the screen. Six of the stimulus words began with a consonant, taking the pronunciation "thuh," and six of the stimulus words began with a vowel, taking the pronunciation "thee." Also, every third block of practice (i.e., every 36 trials) was followed by a novel word. For the 10 novel stimuli that were used, five of the words began with a consonant and five began with a vowel. Responses to these stimuli were also accompanied by correct/incorrect feedback. As in Experiment 1, participant strategies were probed after every trial. The dependent measures were the same as in Experiment 1.

Participants. Twenty undergraduate college students participated for credit in an introductory psychology course. All of the participants tested were native English speakers.

Materials and apparatus. The apparatus used in Experiment 2 was the same as in Experiment 1. The stimulus list of words used was taken from Experiment 2 of the study by Healy and Sherrod (1994). All of the words used were high-frequency words (i.e., they had a frequency equal to or greater than 145 per approximately 1 million words of text; Kučera & Francis, 1967).

Procedure. Participants were tested individually or in pairs (with each participant in a pair using a different computer in a separate quiet room). The instructions, which were parallel but not identical to those used in Experiment 1, were read aloud to the participants. Participants were told that the word *the* is sometimes pronounced "thee" and sometimes "thuh." They were told that they would be presented with 12 different words one at a time on a computer screen. The same 12 words would be presented repeatedly over 30 blocks of practice, where each block is one exposure to each of the 12 words. Using the computer keyboard, they were told to select either the key specially marked "thuh" or the key specially marked "thee" depending on which of these pronunciations they would use preceding the word presented on the screen. They were told that a grammatical rule determines the appropriate pronunciation of *the* for each word and use of this rule might prove helpful. However, the main goal was to learn to respond quickly and accurately to each word. After each response, the participant was told to expect two messages on the screen, the first informing them whether their choice of pronunciation for the preceding word was correct or incorrect and the second asking them to indicate with a key press which of four strategies they used to make their choice: "guess" if they simply guessed the correct pronunciation, "rule" if they used a grammatical rule to make the choice, "memory" if they knew the answer to that specific item based on memory of its previous occurrence, and "other" if they felt that their strategy for the preceding word did not match any of the other three categories. They were told that, occasionally, a novel item would appear and that they were to respond to it in whichever way seemed correct. Novel trials were arranged in the same way as training trials. After a strategy had been chosen on each trial, participants were told that the computer would display a new word and the entire cycle would be repeated.

Results and Discussion

The proportion of correct responses was above chance at the outset of training and continued to increase systematically across blocks, $F(29,551) = 3.48$, $MS_e = .08$, $p < .01$, as shown in Fig. 4. Participants were clearly and understandably more familiar with the basis for correct responding at the beginning of Experiment 2 than they were in Experiment 1. In a cross-experiment comparison, participants were overall more accurate on the pronunciation task (Experiment 2) than on the letter string task (Experiment 1), $F(1,38) = 5.93$, $MS_e = 4.21$, $p < .02$. There was, in addition, an interaction

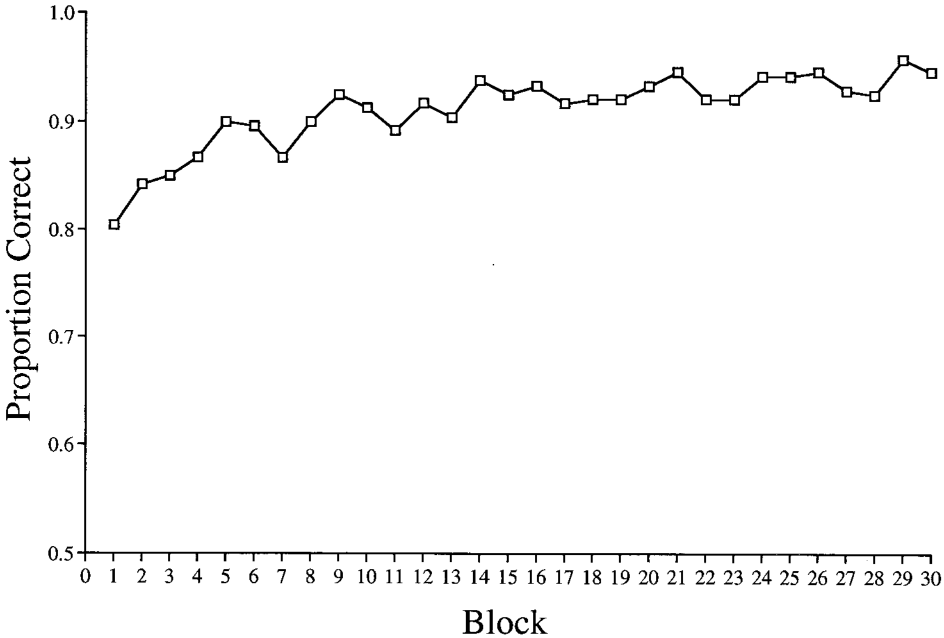

FIG. 4. Proportion of correct responses per block of training in Experiment 2.

of task (string versus pronunciation) by blocks of trials, $F(11,418) = 2.19$, $MS_e = .10$, $p < .02$, showing that the change in accuracy across training depends significantly on the demands of the two different tasks. But, there are ceiling effects that also could produce this interaction given that performance approaches the same limit in both tasks as training progresses.

Figure 5 depicts the proportion of trials on which each strategy was reported, collapsed across all participants and training items, and plotted as a function of trial block in Experiment 2. Both rule-based and instance-based responses were reported approximately 40% of the time initially; guessing and other strategies in this case were minimal. The high initial percentage of rule reports once again indicates that the pronunciation rule was known, in some sense, to some if not all participants at the outset. The high initial percentage of instance reports may reflect at least in part the possibility that participants, contrary to instructions, based their strategy selections on extraexperimental familiarity with the words or training instances. Over blocks of practice, the rule strategy began to dominate, although instance-based responding continued to be reported frequently throughout practice. By the end of practice, participants reported rule use about 65% of the time and instance use about 33% of the time. The difference in proportion of use between the rule (.58) and instance (.37) strategies over all trials was nonsignificant, $F(1,19) = 1.52$, $MS_e = 9.10$. The proportion of trials on which guessing (.03) and other (.01) strategies were reported was much lower than the two primary strategies. The interaction of strategy by blocks of trials in the proportion of use of rule and instance strategies was reliable, $F(29,551) = 1.52$, $MS_e = .06$, $p < .05$. When the rule and instance strategy choice data of Experiments 1 and 2 are combined, a strong three-way interaction appears between task, strategy, and blocks of trials, $F(29,1102) = 11.40$, $MS_e = .06$, $p < .01$, showing that the pattern of changes in strategy choice over trials is dependent on task requirements. The major difference in outcome between the two experiments is that, whereas the rule emerges over trials as the dominant basis of responding in the pronunciation task, instance-based responding is clearly the dominant strategy in the letter string task.

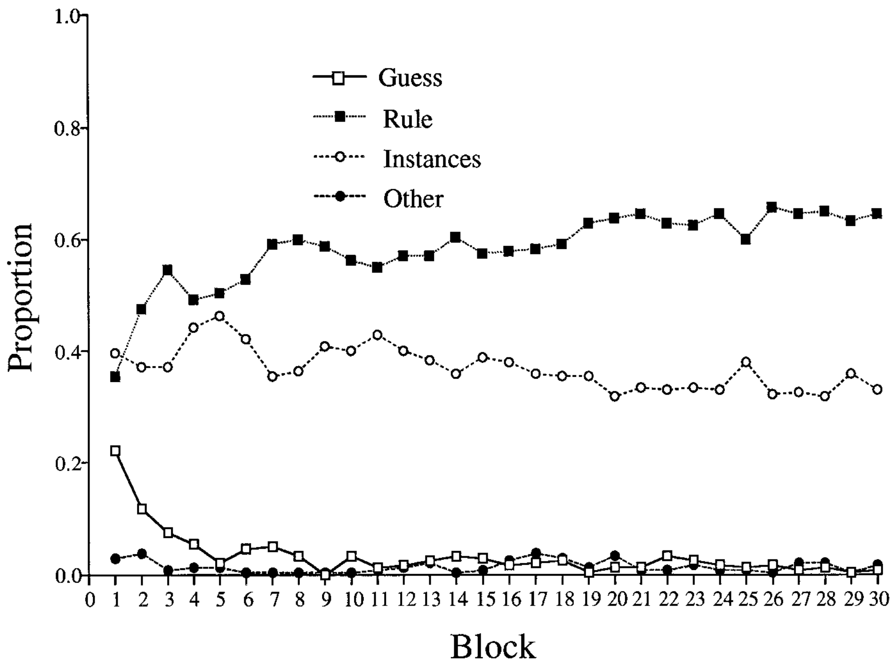

FIG. 5. Proportion of trials within each block of training on which participants reported each strategy in Experiment 2.

The proportion of correct responses for the rule strategy (.97) was higher than for the instance strategy (.89). An analysis of variance averaged across participants with blocks as the random effect showed this difference, which was especially marked between Blocks 5 and 28, to be statistically significant, $F(1,29) = 140.28$, $MS_e = .001$, $p < .001$.

Response time patterns provide evidence consistent with the strategy probing data, showing that, in the pronunciation task, performance is reliably slower when instances were reported than when the rule was reported, $F(1,29) = 201.61$, $p < .01$. As presented in Fig. 6, response time speed-up was relatively well fit by two power functions (although there appear to be some deviations which are difficult to explain on a simple strategy-specific argument):

$$\log T = 2.989 - .178 \log N, \quad r^2 = .77 \text{ (for rule-based performance)} \qquad [5]$$

$$\log T = 3.167 - .237 \log N, \quad r^2 = .81 \text{ (for instance-based performance).} \qquad [6]$$

The overall difference in response times between rule and instance strategy trials reverses that of Experiment 1, but supports our predictions based on the simplicity of the required pronunciation rule (i.e., the difference between vowel and consonant sounds). In this experiment, instance retrieval is likely to be the less direct, more explicit cognitive route to the correct answer than rule use (cf. Anderson et al., 1997). There is an interesting and potentially important discrepancy between the outcome of this analysis and the similar analyses of Experiment 1 and by Rickard (1997) and Delaney et al. (1998). In these earlier cases, the strategy-specific power functions were always essentially parallel, and the major difference between strategies was a difference in overall response time. In contrast, the strategy-specific functions

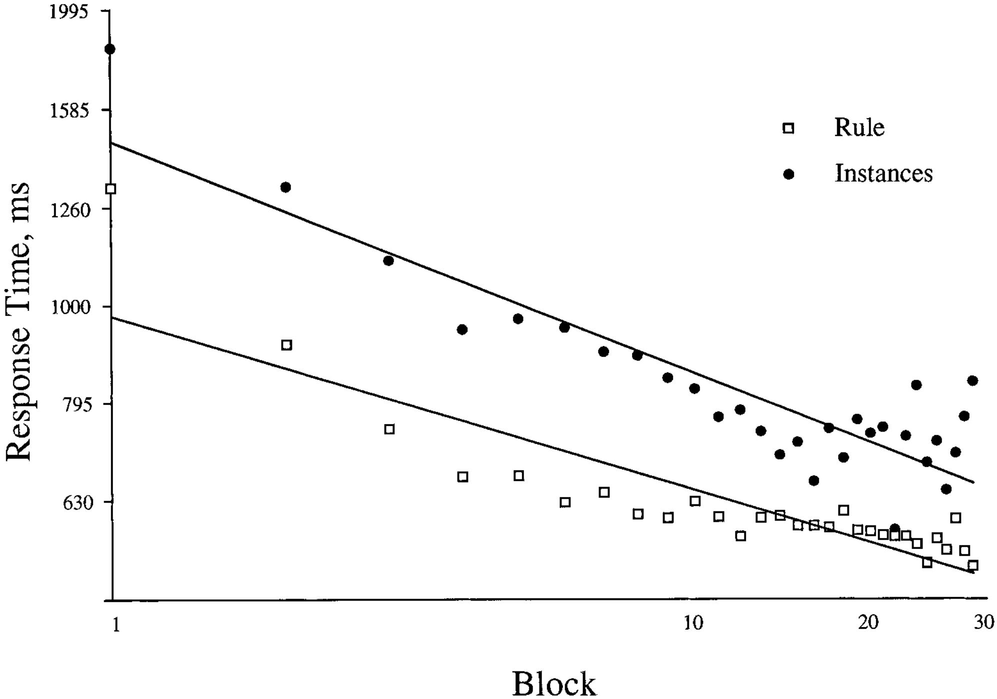

FIG. 6. Response time for correct responses trials during training as a function of block and the strategy reported, rule, or instance in Experiment 2.

in the present experiment differ primarily in slope, with a steeper slope for the instance function than for the rule function. The steeper slope of the instance function implies that strategy-specific response times are converging over trials, although the decreasing frequency of the instance-based strategy suggests that response times might never reach parity. The steeper slope of the power function for the instance strategy may occur because the instances in this task are common words, which are readily learned and retrieved from memory, unlike the letter-strings used in the task of Experiment 1. In any case, in both experiments, the strategy that emerges as the dominant basis of response over trials is the strategy that generates the more rapid responses.

Table 3 shows the proportion of trials on which participants reported guess, rule-based, and instance-based strategies during training and on the novel instances. The difference in frequency of the rule, instance, and guessing strategies was significant for the novel stimuli, $F(2,38) = 8.57$, $MS_e = .18$, $p < .01$. An analysis that included training stimuli as well as novel stimuli yielded an interaction of strategy by stimulus type that was only marginally significant, $F(2,38) = 2.62$, $MS_e = .03$, $p < .10$. In this experiment, guesses were uncommon and rule usage was the most frequently reported strategy on novel instances just as it was on training instances (not the case in Experiment 1). Instance-based responding was reported on .27 of the trials involving novel stimuli, which is less often than for training stimuli

TABLE 3

Proportion of Strategy Reports as a Function of Stimulus Type (Experiment 2)

	Strategy		
Stimulus type	Guess	Rule	Instance
Training	.03	.58	.37
Novel	.08	.62	.27

TABLE 4

Responses to Novel Stimuli (Experiment 2)

	Strategy			
	Guess	Rule	Instances	Other
Total cases	16	124	54	6
Proportion correct	.50	.96	.88	.83
Correct response time (in ms)	1633	766	733	1040

(.37). But the novel instance frequency is high enough to reflect the possible involvement either of extraexperimental memories or of the occurrence of stimulus generalization (despite the lack of similarity between training and novel instances), as might be expected on the basis of the exemplar-based random walk model of Palmeri (1997). The relatively large proportion of rule reports for the novel stimuli supports the claim that the rule, used or discovered by most participants during practice, generalized to new items. The overall proportion of correct responses on the novel stimuli was .88 and did not change significantly over occurrences, $F(9,171) = 1.24$, $MS_e = .09$. When the data of Experiments 1 and 2 are combined, analysis of variance shows that participants were significantly more accurate on novel stimuli in the pronunciation task than in the string task, $F(1,38) = 13.16$, $MS_e = .44$, $p < .01$. The total number of reports of each strategy on novel stimuli, the proportion of correct responses, and the mean response times on correct responses are shown in Table 4 for the pronunciation task alone. An analysis of variance on the novel stimulus data using items as the random effect shows that the proportion of correct responses depends on strategy, $F(2, 18) = 8.26$, $MS_e = .06$, $p < .01$, with the rule-based strategy producing a slightly higher proportion of correct responses than the instance-based strategy, both of which were superior to guessing. These results suggest that participants acquired the correct rule and that there was both rule-based and instance-based generalization to novel stimuli. More correct rule-based responses than instance-based responses to novel stimuli is consistent with Palmeri's (1997) model, according to which the instance-based responses reflect retrieval of similar, but possibly incorrect training instances. Mean response times in (anti-log) milliseconds on rule- (766) and instance-based (733) trials were nearly identical for the novel stimuli in Experiment 2 and responses based on both strategies were considerably faster than responses on guessing trials (1633). There was an overall significant difference among the three most frequently reported strategies on correct response times, $F(2,18) = 13.06$, $MS_e = .03$, $p < .01$. When the two experiments are compared, analysis of variance shows that participants were faster overall on novel stimuli in the pronunciation than in the string task, $F(1,18) = 13.93$, $MS_e = .45$, $p < .01$, suggesting that pronunciation is an easier or simpler task than the alphabetic string task.

EXPERIMENT 3

The results of the first two experiments differed in a number of ways, perhaps most importantly in terms of which strategy became dominant with practice. In Experiment 1, involving artificial strings of letters, an instance-based strategy dominated at the end of practice. In Experiment 2, using real words and a familiar linguistic pronunciation rule, a rule-based strategy dominated. In both cases, the trend in strategy use over trials was quite regular, but it is not entirely clear why different dominant strategies should emerge. The tasks used in these experiments differed in at least three potentially important ways that could have contributed to the observed trends. First, although cross-experiment comparisons are always dubious, it appears from overall response times that the rule for the string task was less obvious and more difficult to use than the rule for the pronunciation task. Difficulty of rule is a likely candidate to affect strategy usage, and, tentatively, we assume that whichever strategy is easier to apply in a given task will become dominant with practice. But, second, although both tasks involved, in a sense, language materials and language-related rules, they differed in how natural or familiar the materials and rules were to participants at the outset. The rule for the letter-strings creates a more artificial and unfamiliar classification learning task than the rule for the pronunciation task. Initial familiarity with the task or with the stimuli comprising the task could also plausibly influence strategy use. Third, the instructions to the participants were not completely parallel in Experiments 1 and 2, and the subtle differences in the descriptions of the strategies might have affected participants' strategy reports.

Rule difficulty and task familiarity were manipulated systematically in Experiment 3 to determine whether either or both of these variables influences the strategy reported by participants. Specifically, we compared two letter-string rules, the one used in Experiment 1 and a simpler one in which the alphabetic sequence is more obvious. Further, we compared two linguistic rules, the relatively simple pronunciation rule used in Experiment 2 and a more difficult linguistic rule based on the count/mass noun distinction that entails a choice between "many" and "much" as a quantifier (Serwatka & Healy, 1998). Rules differed in difficulty within the letter-string and the linguistic tasks, but obviously there is no difficulty metric that allows us to compare across tasks. Lack of a common metric creates some potential ambiguity for statistical analysis, which we address in the Method section. In all four conditions, participants were required to learn the proper response to a small subset of stimuli, either three-letter strings or familiar words. The instructions were written to be parallel in the four conditions, so that it would be unlikely that any differences in performance among them could be attributed to differences in the way that the strategies were described. As previously, we expected to see evidence of both instance- and rule-based responding early in practice. However, in all tasks, a rule-based strategy should eventually dominate when the rule is relatively simple, whereas an instance-based strategy should dominate when the rule is complex, assuming that the strategy that eventually dominates is the one that can be executed most rapidly. Further, because both natural tasks involve familiar words, whereas the artificial task involves arbitrary letter strings, learning the instances should be easier in the natural tasks than in the artificial task. Thus, parallel strategy-specific power functions are expected for the artificial task, as found previously in Experiment 1 and by Logan (1988) and Rickard (1997), but converging power functions are expected instead for both natural tasks, as found in Experiment 2.

The count/mass distinction provides for another potentially important window on strategy use. Serwatka and Healy (1998) have discussed the linguistic complexity of this distinction. Whether "much" or "many" is appropriate as the quantifier depends largely on a countability criterion: If a noun refers to something that can be counted (e.g., "carrot"), the proper quantifier is "many," otherwise (e.g., "spinach") "much" is appropriate. Countability is fairly obvious when the items are concrete, as in the foregoing examples. However, when the reference items are abstract, such as "idea" ("many") or "advice" ("much"), the countability criterion is hard, if not impossible, to apply (e.g., ideas should be no easier to count than advice). Rule use is likely to be more difficult in the abstract than in the concrete case. For this reason, both concrete and abstract singular nouns are included in the *many/much* lists of Experiment 3 to determine whether this variable has any impact on the use of the rule versus the instance strategy.

Method

Overview and design. There were four experimental conditions, generated from the 2 × 2 factorial combination of task difficulty (easy/hard) and task familiarity (artificial/natural). These four conditions are labeled (1) string-easy, (2) string-hard, (3) *thee/thuh* (easy), and (4) *many/much* (hard). The string-easy and string-hard conditions were designed to replicate and extend the findings of Experiment 1, whereas the *thee/thuh* and *many/much* conditions were designed to replicate and follow-up the findings of Experiment 2. A conservative conception of the design of this experiment would describe it as having a single factor of task with four levels. To emphasize the two distinct variables manipulated in this study, we have chosen to analyze the data instead using a 2×2 factorial format, despite the fact that the manipulation of difficulty is admittedly not equivalent for the artificial and the natural tasks. Because the analysis we have chosen provides a finer breakdown of variance than other possibilities, readers preferring a more conservative analysis will have sufficient information provided in the Results section to derive their own conclusions about differences among the four conditions.

In the string-easy and string-hard conditions, each participant was trained on a single set of 12 strings of three letters (e.g., HGF) one at a time repeatedly over 30 blocks of practice. Participants were asked to indicate whether each letter-string is "code" or "noncode." The "code" classification was determined by the same rule used in Experiment 1. Six of the letter-strings were classified as code and six of the letter-strings were classified as noncode. Also, every second and third block of practice was followed by a novel string of three letters. For the novel stimuli that were used, 10 were code strings and 10 were noncode strings. The difference between the string conditions was that all of the strings were in alphabetical order for the string-easy condition but none were in alphabetical order for the string-hard condition.

In the *thee/thuh* condition, each participant was trained on a single set of 12 stimuli consisting of high-frequency words (nouns and adjectives) presented one at a time repeatedly over 30 blocks of practice. Participants were asked to indicate whether "thee" or "thuh" is the correct pronunciation for *the* if it were to precede the stimulus word presented on the screen. Pronunciation was determined by the same linguistic rule used in Experiment 2. Six of the stimulus words began with a consonant and six began with a vowel. Also, every second and third block of practice was followed by a novel word. For the novel stimuli that were used, 10 began with a consonant and 10 began with a vowel.

In the *many/much* condition, each participant was trained on a single set of 12 singular nouns presented one at a time repeatedly over 30 blocks of practice. Participants were asked to indicate whether "many" or "much" is the correct modifier if it were to precede the stimulus word presented on the screen. The correct modifier was determined by a grammatical rule. By this grammatical rule, a noun is preceded by the "many" modifier only if it is a count noun (e.g., brain, rock). If the noun is a mass noun (e.g., dust, furniture), then the preceding modifier is "much." The nouns were all presented in singular form because the count/mass distinction also determines whether a noun can be pluralized. Six of the stimulus words were count nouns and six were mass nouns. Three stimuli of each type were abstract and three were concrete. Also, every second and third block of practice was followed by a novel word. For the novel stimuli, 10 were count nouns and 10 were mass nouns, with 5 of each type abstract and 5 concrete.

Participants received feedback on every trial in each condition, even on trials involving novel stimuli. Also as in Experiments 1 and 2, participant strategies were probed after every trial in each condition. The dependent measures were the same as in Experiments 1 and 2.

Participants. Forty-eight undergraduate college students in an introductory psychology course volunteered to participate for course credit. All of these participants were native English speakers. Participants were assigned by a fixed rotation to the four conditions with 12 in each condition.

Materials and apparatus. The materials and apparatus used in Experiment 3 were the same as in Experiments 1 and 2 with the following exceptions. A new set of letter strings was created for the string-easy condition by rearranging the same letter strings used in Experiment 1 into alphabetical order. The stimuli used in the string-hard condition were taken directly from Experiment 1. Also, 10 more letter strings were added to the novel set of the string-easy and string-hard conditions. The stimuli used in the *thee/thuh* condition were the same as in Experiment 2 except that 4 of the words were changed. Also, 10 more words were included in the novel set. These new words had the same constraints as in Experiment 2. The stimuli used were taken from Experiment 3 of the study by Healy and Sherrod (1994). All of the words used in the *thee/thuh* condition were high frequency words (i.e., they had a frequency equal to or greater than 84 per approximately 1 million words of text; Kučera & Francis, 1967). Lastly, a new set of words was included for the *many/much* condition. These words were taken from the study reported by Serwatka and Healy (1998). All of the words used in the *many/much* condition were singular nouns divided into four categories of count concrete, count abstract, mass concrete, and mass abstract. For the training stimuli, there were three count concrete, three count abstract, three mass concrete, and three mass abstract words. For the novel stimuli, there were five count concrete, five count abstract, five mass concrete, and five mass abstract words.

Procedure. Participants were tested individually or in pairs (with each participant in a pair using a different computer) in a quiet room, free from outside distractions. The instructions for the string conditions and the *thee/thuh* condition, similar to those in Experiments 1 and 2, respectively, but modified as necessary to improve their coherence and to make the two sets of instructions as similar as possible, were read aloud to the participants. The instructions for the *many/much* condition, prepared to be as similar as possible to those in the string and *thee/thuh* conditions, were as follows.

> Today's experiment explores basic learning processes. You will be presented with words one at a time on a computer screen. During the experiment, you will see each of 12 different words several times. These same 12 words will be presented repeatedly over 30 blocks, where each block is one exposure to each of the 12 words. Some words use the modifier "many," and some words use the modifier "much." In this experiment, half of the words are examples of the use of "many," and half of the words are examples of the use of "much." It is your job to select either the key marked "many" or the key marked "much" depending on which of these modifiers you think is correct preceding the word presented on the screen. You will use specially marked keys on the computer keyboard to record your answer. Note that your job is not to guess at each modifier. Rather, you are to learn which words use the modifier "many" and which words use the modifier "much." There is a relatively simple rule which will discriminate between the use of "many" and the use of "much" and which, if you discover it, can help you in this task. Or, you may be able to learn the correct response by memorizing which words follow the modifier "many" and which words follow the modifier "much." Whether you discover a rule or not, the overall goal is to perform as quickly and accurately as possible on each trial. After making your response to each word (that is, whether it was preceded by "many" or "much"), the displayed word will disappear from the screen and the computer will display two new messages on the screen, one after the other. The first message will inform you whether your choice of modifier for the preceding word was correct or incorrect. The second message will ask you to indicate which of four strategies you used to make your choice. Your strategy options will be "Guess," "Rule," "Memory," or "Other." You should select the "Guess" option if you made your response having no idea whether the preceding modifier was "many" or "much." Select the "Rule" option if you used the rule to make your choice. You should select "Memory" if you remembered that the preceding word used a modifier of "many" or "much" based on an earlier experience with it. Select "Other" if none of the first three strategies applies. Note that whether you get an answer right or wrong should not influence the strategy that you choose for that trial. You can be right or wrong for a variety of reasons. Here we are interested in the strategy that you used, regardless of whether it produced the correct answer. Most of the words will appear in every block of the experiment. Occasionally, though, a new word may appear that you haven't seen before. Don't forget that on each trial, it is important that you respond "many" or "much" as quickly and accurately as you can. You can take as much time as you need, however, to perform the strategy selection part of the task.

Results and Discussion

The proportion of correct responses increased systematically over blocks of trials in all conditions of the experiment, $F(29,1276) = 24.58$, $MS_e = .09$, $p < .01$, as shown in Fig. 7. There was an overall difference in accuracy between the artificial letter-string tasks and the more natural linguistic tasks, $F(1,44) = 5.71$, $MS_e = 2.72$, $p < .02$, with performance in the natural tasks starting (in Block 1) and remaining more accurate than in the artificial task throughout practice. Easy tasks were performed more accurately than hard tasks, but the difference was not statistically reliable, $F(1,44) = 3.04$, $MS_e = 2.72$, $p < .09$. The interaction of type of task by difficulty level was nonsignificant, as were all interactions with blocks of trials. In a separate analysis of proportion of correct responses over blocks in the *many/much* (hard

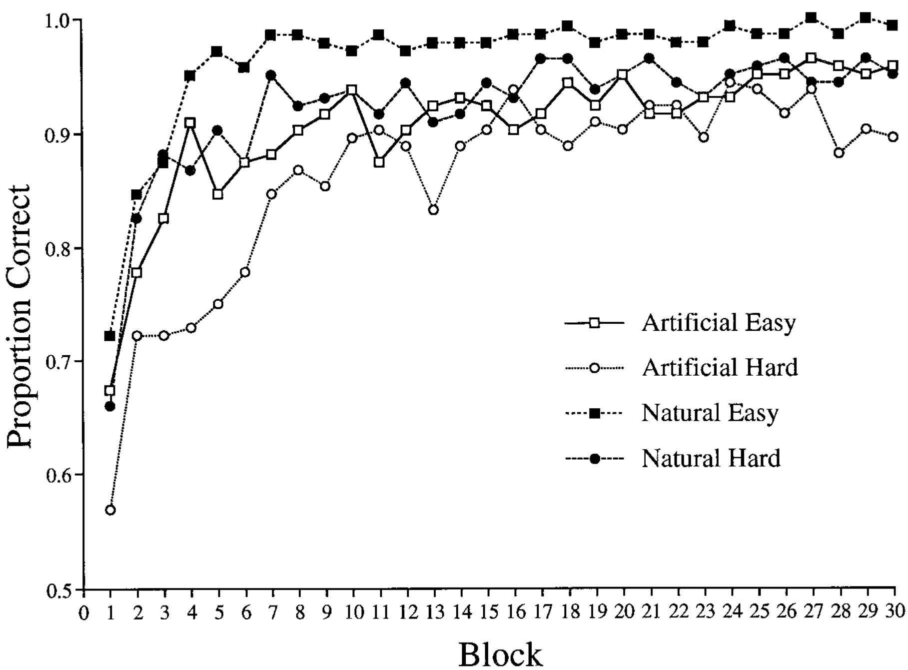

FIG. 7. Proportion of correct responses per block of training as a function of type of task in Experiment 3.

natural) task alone, participants were more accurate on concrete (.946) than on abstract (.898) words, $F(1,11) = 6.24$, $MS_e = .21$, $p < .03$, as had been found in previous work (Serwatka & Healy, 1998), suggesting that, whatever the basis for using a quantifier, it is easier to apply when the target is concrete and presumably easier to count.

Over all trials, strategies differed in proportion of use, with the instance strategy more frequent (.60) than the rule strategy (.30), $F(1,44) = 4.96$, $MS_e = 6.07$, $p < .05$. The proportion of use was even lower for guessing (.08) and other strategies (.02). Instance and rule strategy use interacted with blocks of trials, $F(29,1276) = 11.77$, $MS_e = .08$, $p < .01$, and with blocks and task familiarity (artificial/natural), $F(29,1276) = 5.21$, $MS_e = .08$, $p < .01$. There was a significant four-way interaction among rule and instance strategy, blocks, task familiarity, and task difficulty, $F(29,1276) = 2.86$, $MS_e = .08$, $p < .01$, shown (for all four strategies) in Fig. 8. The pattern of strategy use over blocks was much the same for three of the four task conditions: artificial-easy, artificial-hard, and natural-hard. In each of these conditions, use of the instance strategy increased and was clearly the dominant strategy by the end of practice. Rule use first increased and then decreased, as instances became dominant. Guesses were frequent at the outset of practice, especially in the artificial tasks, but quickly dropped to insignificance. Other strategies were rarely reported. The picture was qualitatively and quantitatively different in the natural easy task, which was based on the *thee/thuh* distinction. As in Experiment 2, rule use increased from the beginning to the end of practice and was clearly the dominant strategy after about 15 blocks of trials. Instance use increased and then decreased in frequency as the rule became dominant. Guesses and other strategies followed essentially the same pattern over trials as they did in the other three conditions.

In a separate analysis of rule and instance strategy use in the natural hard condition, significant effects attributable to the difference between abstract and concrete nouns were observed. Specifically, the main effect of concrete/

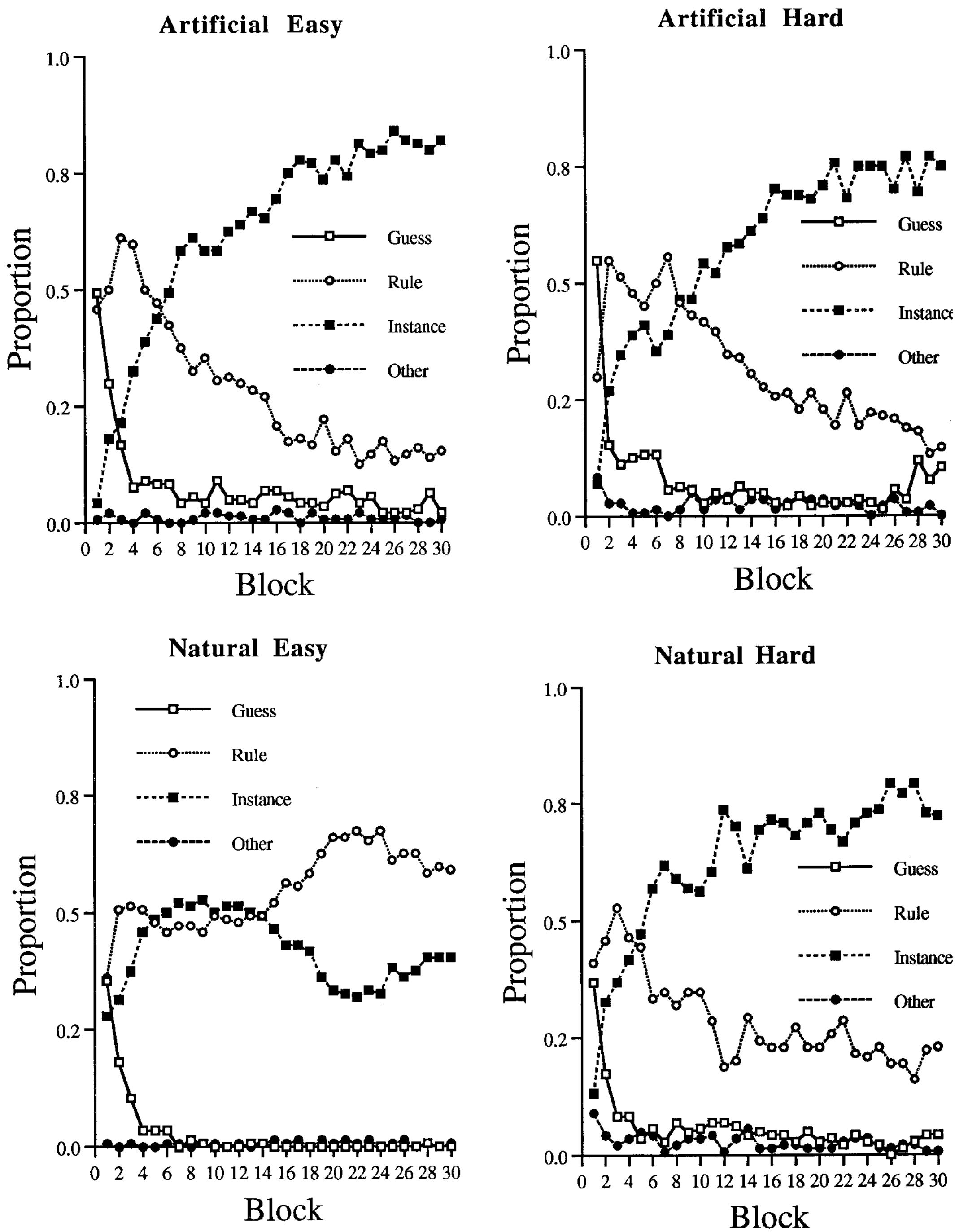

FIG. 8. Proportion of trials within each block of training on which participants reported each strategy as a function of type of task in Experiment 3.

abstract was reliable, $F(1,11) = 7.47$, $MS_e = .01$, $p < .02$, and this variable interacted with strategy, $F(1,11) = 5.01$, $MS_e = .09$, $p < .05$. As shown in Fig. 9, the instance strategy, which was overall the most frequently reported strategy in this task, was especially common on

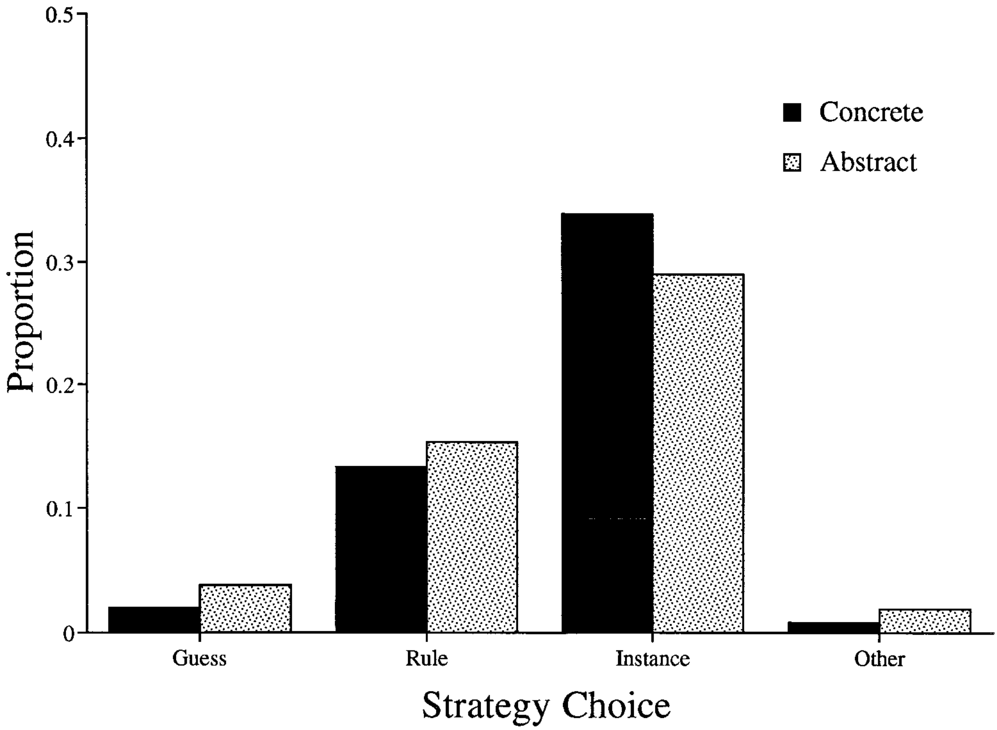

FIG. 9. Proportion of training trials containing abstract and concrete words on which participants reported each strategy in the natural-hard task in Experiment 3.

concrete nouns, whereas abstract nouns elicited the rule (and other strategies and guesses) more often than did concrete nouns. This outcome might result from the fact that concrete nouns are easier to remember than abstract nouns (Paivio, 1986) and thus are more readily available to support instance-based responding in the present task. At present, we have no memory data to provide independent evidence for this hypothesis.

For all four conditions, the proportion of correct response was higher with the rule strategy (artificial-easy: .971; artificial-hard: .957; natural-easy: .994; natural-hard: .987) than with the instance strategy (artificial-easy: .958; artificial-hard: .885; natural-easy: .969; natural-hard: .969); this difference between the rule and instance strategies was significant for the artificial-hard condition, $F(1,29) = 15.95$, $MS_e = .005$, $p < .001$, for the natural-easy condition, $F(1,29) = 764$, $MS_e = .001$, $p < .01$, and for the natural-hard condition, $F(1,29) = 13.50$, $MS_e = .003$, $p < .001$, in analyses with blocks as the random effect. In all four conditions, the advantage for the rule strategy was greatest in the early blocks of trials. Thus, in all four conditions, the participants seem to have induced at the outset of practice the appropriate rule for the task.

The response-time analysis shows a clear and significant difference between the instance and the rule strategies in the artificial tasks, with performance being reliably slower when the rule strategy was used. Over blocks of training instances, mean antilog response times in milliseconds were 1313 for rule in the easy version, 1091 for instance in the easy version, 1794 for rule in the hard version, and 1315 for instance in the hard version. This outcome is consistent with Experiment 1 in which only the harder artificial string task was used. The mean difference between strategies was larger in the harder than in the easier artificial task, but is still considerably smaller than the comparable dif-

TABLE 5

Power Functions and Variance Accounted for in Experiment 3 as a Function of Task Familiarity, Task Difficulty, and Strategy

	Strategy	
	Rule	Instance
Artificial task		
Easy	$\log T = 3.412 - .271 \log N$, $r^2 = .62$	$\log T = 3.314 - .255 \log N$, $r^2 = .94$
Hard	$\log T = 3.526 - .252 \log N$, $r^2 = .89$	$\log T = 3.400 - .261 \log N$, $r^2 = .93$
Natural task		
Easy	$\log T = 3.110 - .239 \log N$, $r^2 = .94$	$\log T = 3.197 - .306 \log N$, $r^2 = .84$
Hard	$\log T = 3.305 - .225 \log N$, $r^2 = .85$	$\log T = 3.423 - .353 \log N$, $r^2 = .96$

ference found by Rickard (1997) and Delaney et al. (1998) in tasks that involve calculation. Differences in response times between strategies were more complex in the natural tasks. Mean antilog response times in milliseconds averaged over blocks were 711 for rule in the easy version (*thee/thuh*), 735 for instance in the easy version, 1151 for rule in the hard version (*many/much*), and 1102 for instance in the hard version. Response-time speed-up over blocks of trials was fit by power functions, presented in Table 5 and Figure 10. The power function fits were good ($r^2 \geq .84$) in all cases except for the rule strategy in the artificial-easy task ($r^2 = .62$). The relatively poor fit in this condition may be due in part to the fact that there are not many reports of rule use late in the training trial sequence. It should be noted that the power functions reported here were based on group averages and that participants did not contribute equally to these averages because not every participant chose both a rule-based and an instance-based strategy on every block of trials. Nevertheless, when we removed data points from the power function fits if 5 or fewer of the 12 participants in a condition contributed to the points, the fits were essentially unchanged.

In the artificial tasks, response-time speed-up functions were nearly parallel and responses made under the instance strategy were faster both at the beginning and at the end of practice. Functions in the natural tasks converged when the discrimination was based on *thee/thuh*, just as they seemed to do in Experiment 2, with the instance strategy producing slower responses at the outset (and overall). In the natural-hard task, instance-based responses were slower at the outset, but rule and instance speed-up functions intersected and rule based responses were slower by the end of practice (and overall). The intersecting functions for both natural tasks are due to the fact that the response time speed-up functions have a particularly steep slope for the instance strategy in those cases. This steep slope may occur because the instances in both natural tasks are common words, which are readily learned and retrieved from memory, unlike the letter strings used in the artificial tasks.

Response times were related to the dominant strategy at the end of practice. In artificial tasks, the dominant strategy was instance based in both the easy and the hard versions. In the natural linguistic tasks, the rule was dominant by the end of practice in the easy version, and rule-based responses were faster overall (although the functions converge). Although instance-based responding was slower at the outset of practice in the hard version, it was faster at the end when the instance strategy was dominant. In all tasks then, the dominant strategy is associated with the basis of responding that is faster. The analysis of variance on log response times for correct trials with blocks as the random effect reveals that, over all conditions, rule-based responding was slower than instance-based responding, $F(1,29) = 59.12$, $MS_e = .003$, and that both task familiarity (artificial/natural), $F(1,29) = 96.22$, $MS_e = .002$, and task difficulty, $F(1,29) = 12.66$, $MS_e = .002$, interacted with strategy, $ps < .01$.

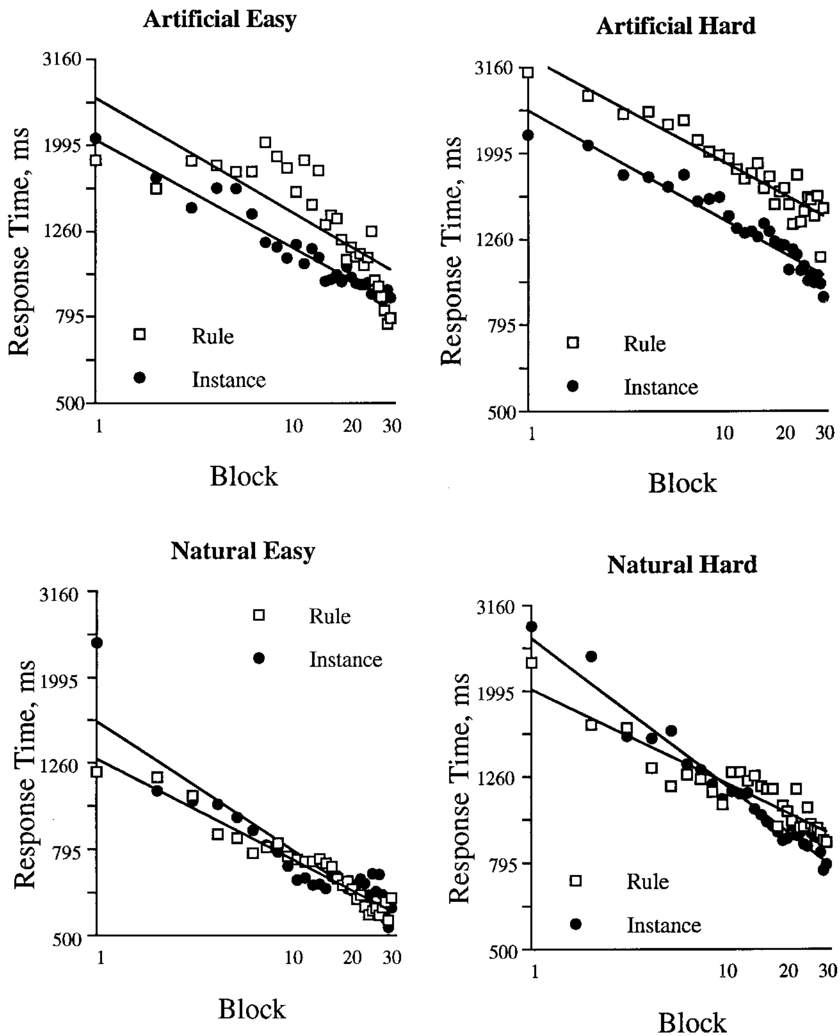

FIG. 10. Response time for correct responses trials during training as a function of block, the strategy reported, rule, or instance and type of task in Experiment 3.

Even though there are differences in response time as a function of strategy at the end of training, both rule-based and instance-based strategies yield accuracy close to the ceiling over the last 10 blocks of training trials in each of the four conditions (artificial easy, instance based $M = .997$, rule based $M = .994$; artificial-hard, instance based $M = .993$, rule based $M = .983$; natural-easy, instance based $M = .966$, rule based $M = .961$; natural-hard, instance based $M = .964$, rule based $M = .989$).

Table 6 shows the proportion of trials on which participants reported guessing, rule-based, and instance-based strategies during training and on novel instances. The pattern of results was nearly the same for both artificial-easy and -hard tasks. During training, guesses were relatively infrequent (except during the first block or two). As found in Experiment 1, however, guesses were reported frequently and more often than instance strategies on novel stimuli. Instance-based responses were most common on the training stimuli, but rule-based responses were most common on the novel stimuli, as would be expected if participants

TABLE 6

Proportion of Strategy Reports as a Function of Condition and Stimulus Type (Experiment 3)

	Strategy		
	Guess	Rule	Instances
Artificial-Easy			
Training	.08	.29	.63
Novel	.23	.58	.15
Artificial-Hard			
Training	.08	.32	.58
Novel	.23	.57	.17
Natural-Easy			
Training	.03	.55	.42
Novel	.18	.63	.19
Natural-Hard			
Training	.06	.29	.63
Novel	.13	.60	.16

recognized that novel stimuli had not been seen previously. Strategy choices differed between the natural-easy and -hard tasks. For the easier, *thee/thuh* task, guesses were uncommon and rule usage was the most frequently reported strategy on novel as well as on training instances. There was a strong mixture of both rule-based and instance-based strategies for training items throughout practice, but rule use predominated on novel stimuli. These results support the claim that the rule was used or discovered by most participants during practice and that use of this rule generalized readily to new items. In the natural-hard task, the pattern of strategy reports on novel and training instances was similar to that on the artificial tasks. Instance-based performance was typical of training, whereas rule-based performance dominated on novel stimuli. These effects are supported by an analysis of variance on strategy use (i.e., the proportion of trials on which participants selected a given strategy). In this analysis, the main effect of strategy (excluding the "other" category), $F(2,88) = 15.40$, $MS_e = .20$, $p < .01$, and the interaction of stimulus type (novel/training) and strategy, $F(2,88) = 55.25$, $MS_e = .05$, $p < .01$, were significant, and the interaction of stimulus type, strategy, task familiarity (artificial/natural), and task difficulty, $F(2,88) = 2.53$, $MS_e = .05$, $p < .09$, was close to statistical reliability.

The strategy report data might reflect either a mixture of strategies within participants or a mixture of participants each adopting a single unique strategy at a particular practice level. We examined individual participant strategy selections over all training trials in the four conditions of Experiment 3. We classified each block of trials as rule based or instance based only if the participant used that strategy on more than half of the trials in that block. Then, we examined the sequence of selections over blocks of trials for transitions in strategy choices, from either rule to instance or instance to rule. We found that 9 or 10 participants of 12 in each condition showed at least one clear transition, from rule based to instance based, for the two artificial tasks and for the natural-hard task. In the natural-easy task (the task involving the *thee/thuh* pronunciation distinction), only 5 of the 12 participants showed a clear transition, in this case from instance to rule, with the remaining participants reporting rule-based responses from the outset or after a series of guesses. The implication is that strategy transitions do indeed occur on a per participant basis and do not

TABLE 7

Responses to Novel Stimuli (Experiment 3)

	Strategy			
	Guess	Rule	Instances	Other
Artificial-Easy				
Total cases	56	140	36	8
Proportion correct	.412	.964	.861	0
Correct RT (in ms)	1153	1542	1377	na
Artificial-Hard				
Total cases	54	137	41	8
Proportion correct	.389	.912	.659	0
Correct RT (in ms)	1052	2897	1268	na
Natural-Easy				
Total cases	42	152	46	0
Proportion correct	.619	.993	.783	0
Correct RT (in ms)	1807	780	1007	na
Natural-Easy				
Total cases	32	143	38	27
Proportion correct	.656	.972	1.000	.926
Correct RT (in ms)	798	1690	1374	782

reflect merely an unsystematic mixture of participants at different practice levels.

Table 7 shows the total number of reports of each strategy on novel stimuli, the proportion of correct responses, and the mean log response times on correct responses in the four conditions of this experiment. The overall proportions of correct responses on the novel stimuli were .788, .721, .887, and .929 for artificial-easy and -hard and natural-easy and -hard, respectively. The task difference (artificial/natural) was reliable, $F(1,44) = 10.11$, $MS_e = .56$, $p < .01$, with performance being more accurate on the natural tasks, but there was no effect of difficulty or an interaction of task by difficulty. There was a small but significant effect attributable to successive novel instances, $F(19,836) = 1.86$, $MS_e = .11$, $p < .02$, and to the interaction of task, difficulty, and successive instances, $F(19,836) = 1.77$, $MS_e = .11$, $p < .03$, but there were no clear trends in these cases.

In an analysis using only those novel trials on which guess, rule, or instance was the reported strategy and using items as the random effect, mean proportion of correct responses were significantly higher for the natural tasks than for the artificial tasks, $F(1,19) = 31.29$, $MS_e = .02$, $p < .001$. In addition, rule-based responses were more accurate overall on novel stimuli than were instance-based responses or the responses based on guessing, $F(2,38) = 74.10$, $MS_e = .02$, $p < .001$ Within conditions, only the natural hard task showed a reversal of this pattern, presumably because of a ceiling effect in this case; the interaction of strategy and task familiarity was significant, $F(2,38) = 7.44$, $MS_e = .02$, $p < .01$, as was the three-way interaction of strategy, task difficulty, and task familiarity, $F(2,38) = 5.17$, $MS_e = .03$, $p < .02$.

In a parallel analysis using only those novel trials on which guess, rule, or instance was the reported strategy and using items as the random effect, mean correct response times were significantly faster for the natural tasks than for the artificial tasks, $F(1,19) = 12.86$, $MS_e = .04$, $p < .01$, and participants were faster for stimuli on which the instance strategy was used than on those for which the rule strategy was used, $F(2,38) = 5.70$, $MS_e = .07$, $p < .01$. Only in the natural easy condition was the rule strategy faster than the instance strategy; the interaction of strategy and task familiarity was significant, $F(2,38) = 10.71$, $MS_e = .05$, $p < .001$, as was the interaction of strategy and task difficulty, $F(2,38) = 23.07$, $MS_e = .05$, $p < .001$, and the three-way interaction of strategy, task difficulty, and task familiarity, $F(2,38) = 4.95$, $MS_e = .07$, $p < .02$.

GENERAL DISCUSSION

Changes in overall measures of performance, such as accuracy and speed of response, during practice in binary classification tasks might mask and certainly do not always reveal the complexity of changes in the underlying basis of trial-by-trial responses. The strategies learners developed in these experiments did not typically move in parallel during training with overall performance measures. In the experiments reported here, overall measures changed systematically toward an asymptote, while strategy measures showed increases and decreases in the frequency of various strategies and transitions from one strategy to another at certain points in practice.

Strategy transitions were not the same for all binary classification tasks. In tasks with artificial stimuli and an initially unknown classification rule, participants often began by guessing the classification response for a given stimulus. As the rule was acquired or identified over successive instances, rule-based responses became more common and, in some cases, the rule became the primary basis of response. Because there were relatively few unique training stimuli to be categorized, because each was repeated a large number of times, and because the required rule needed to be executed explicitly, a transition in response basis eventually occurred from rule to instance memory, and learners increasingly reported using recall of previous instances as a basis of classification response.

Although we did not systematically vary the number of instances per classification, the transition from rule- to instance-based responding would almost certainly be sensitive to the number of unique training stimuli to be categorized. With similarity controlled, the more stimuli to be remembered, the more demanding instance memory would be relative to rule use, implying that rule-based responding might become dominant in a large stimulus domain (but see, e.g., McKinley & Nosofsky, 1995). Even in the present case where stimuli are few, participants tended to revert to rule-based responding when novel stimuli were presented. When training trials on which participants reported rule-based responding were separated from those on which they reported instance-based responding, a clear difference in mean response time emerged, with instance-based responding being significantly faster than rule-based responding, especially for the harder task. Similar effects have been reported previously by Rickard (1997) and Delaney et al. (1998) in calculational tasks. Because the alphabetic rule, when used by participants, had to be consciously or explicitly applied to each stimulus, instance memory should be a less cognitively demanding or time-consuming strategy than rule-based responding, leading to faster response times.

The outcome was somewhat different in natural language tasks where participants had some familiarity at the outset of training with the stimuli and with the classification rule required. Accuracy changes over trials in the *many/much* quantifier task appeared essentially to be the same as accuracy changes in the two artificial tasks. Moreover, strategy patterns were similar, with learners conforming to the quantifier rule and reporting rule use as a basis of response in the early blocks of trials. But the rule eventually gave way to instance-based performance, which became dominant in the later blocks of trials. Changes in response times are consistent with this trend; instance-based responses were slower than rule-based responses in the early blocks of trials but became faster when instance-based responding dominated. The trend in response times was different, however, in the quantifier task than it was in the artificial tasks, in which instance-based responding was faster than rule-based responding over all trials. We think that this difference can be attributed, in part, to the difference in familiarity and memorability of instances in the two types of tasks. Well-known words already in semantic memory were used as instances in the quantifier task, whereas meaningless letter strings were used as instances in the artificial task. Most likely, instances were more readily stored and retrieved from memory in the natural task than in the artificial tasks. Moreover, participants might claim rule use early in practice based on a general feeling of knowing or familiarity with the quantifier task and the noun stimuli it employed. Later, as the word stimuli were repeated and the limited number became clear to participants, remembering that specific instances had occurred earlier in the trial sequence provided an easier access to correct responses than rule application. It has been well documented that memory is better for concrete than for abstract nouns (Paivio, 1986). In the present study, instance-based performance was more common with concrete than with abstract nouns, which is consistent with the observation that responding in later trials was dependent on memory for the specific stimuli used in training.

The data obtained in the *thee/thuh* pronunciation task were essentially the same across Experiments 2 and 3 but, in important respects, different from data in both the artificial alphabetic tasks and the *many/much* natural task. Accuracy increased systematically, and there were reliable strategy transitions across trials, as in the other conditions. But, in the *thee/thuh* case, classification responses were ultimately controlled by the rule rather than memory for instances. Contrary to the suggestion of Anderson et al. (1997), rule-based responses dominated through most of practice and were generally faster than instance-based responses. The dominance of the pronunciation rule over instance-based responding can probably be traced to the fact that this rule, although not necessarily verbalizable by all participants, is simpler, more straightforward, and easier to apply than the *many/much* quantifier rule and faster to invoke than instance memory. (In fact, the rule for the *thuh/thee* task in our study required that participants consider only the first letter of each stimulus. The lack of instance-based responding may reflect the attentional demands of this task because participants need not attend to the entire word to respond correctly. Therefore, the entire word may not be processed and stored in memory; see, e.g., Logan & Etherton, 1994.) The simple pronunciation rule is more likely to become proceduralized (Anderson et al., 1997) with practice than either the alphabetic rule of the artificial tasks or the quantifier rule of the *many/much* task. Even so, the difference between frequency of rule-based and instance-based responding toward the end of practice was smaller in the *thee/thuh* case than in any of the other conditions of Experiment 3. There is some suggestion in the data, in fact, that instance-based responding might eventually overtake rule-based responding in frequency, if practice had continued further.

Accuracy improved as participants made the transition from the guessing strategy to the rule-based or instance-based strategies. In the late stages of practice, there were no differences between rule and instance strategies in accuracy. Nonetheless, at all stages of practice, there were important differences between strategies in the time required to make a correct response. Participants generally came to use whichever highly accurate strategy yielded the faster response time. Thus, a careful and complete analysis of trial-by-trial strategy usage typically will have much to offer toward a full understanding of the role of practice and of the retention interval in classification and in skilled performance. Relevant to this conclusion are the results of Rickard (1994; see also Rickard & Bourne, 1995), who observed that, when participants were retested on previously practiced psuedoarithmetic problems after a delay interval, the accuracy and speed of their responses depended on whether they retained the specific instances that they had been trained on. All subjects gravitated from a rule-based (a calculational algorithm) to a faster instance-based strategy during learning. When participants remembered an instance at retention, response time on that problem was no different from that observed at the end of practice. When remembering failed, however, participants used the rule-based strategy that returned response time levels to where they were at the outset of practice. These results not only are consistent with the outcome of the present experiments, but also imply that retention, as well as acquisition, can be strategy specific (see also Conway et al., 1997).

Rule-based responding was dominant at the end of practice on the *thee/thuh* task. But, as noted above, the instance-based strategy showed an upward trend later in practice in the *thee/thuh* task of Experiment 3. There is some possibility that, with further practice, instance-based responding might have become dominant even in the *thee/thuh* task, as would be expected by Anderson et al. (1997). The fact that instance-based response times converged on those of rule-based trials in the log-log plots is consistent with this possibility. Such an outcome would not, however, invalidate our argument that the less cognitively demanding (i.e., faster) strategy eventually dominates. It would merely mean that the cognitive demands of strategies change with experience and use, and there is plenty of evidence for that conclusion in the other data of these experiments. We cannot be sure of this outcome, of course, and at least one additional experiment employing the *thee/thuh* task with extended practice will be necessary to verify or refute this possibility.

Do the trial-by-trial reports of participants validly reflect their real strategies in these experiments? Although the trial-by-trial reports (like any responses that might be required in an experiment) are undoubtedly subject to some response biases, there are several kinds of evidence that they do reflect real strategies. For one thing, reports taken in these experiments are essentially the same as those effectively employed by Tulving (1985) and Conway et al. (1997), in pure memory tasks, and by Rickard (1997) and Delaney et al. (1998), in calculational tasks, to identify the participants' basis of responding. Further, in all conditions and tasks of our experiments, mean response times differed reliably between strategies and were predictable from reported strategies during the training trials. Finally, strategy reports were different for novel and training trials. Novel instances had never been seen by participants on earlier trials. They might elicit a guess initially, but on later trials rule-based responses should be reported, not instance-based responses. In fact, the frequency of instance-based reports on novel trials was low relative to reports on training trials, even when instance-based responding had become dominant. This contrast between responses on training and novel stimuli clearly supports the argument that participants acquired or discovered the operative rule and used it when necessary (on novel stimuli) even though they might employ a less demanding instance strategy when it was applicable. Still, instance reports were not uncommon for novel stimuli. We suspect that nearly all of these reports can be accounted for by one of two factors. First, there might have been some stimulus generalization (or confusion) between training and novel stimuli leading a participant to believe that a novel instance had indeed occurred earlier on a training trial. Second, despite instructions to the contrary, some participants might have been uncertain about whether extraexperimental experience with stimuli (words) could serve as the basis of an instance-based reports.

The validity of strategy reports might be challenged in another way. Suppose strategy choice depends on response time, so that participants report rule use only when they have responded slowly. This possibility is unlikely in the present experiments for three reasons in our view. First, although significant for the artificial task, the overall difference in response times between rule and instance strategies was much smaller in our experiments than in the earlier experiments by Rickard (1997) and by Delaney et al. (1998), which involved rules that were considerably more complex than the ones we used. Hence, response time would have been a much weaker cue to strategy selection in our study than in those earlier studies. Second, participants in the *thee/thuh* pronunciation task showed the opposite pattern of results from those in the artificial tasks (and in experiments by others), responding more *quickly* when they reported using the rule than when they reported instance-based responding. Third, unlike for the artificial task, the strategy-specific power functions were not parallel for our natural tasks involving common words, but rather converged. Thus, there was not only a difference in intercept between the rule and instance functions but also a difference in slope, reflecting greater speed-up on the instance strategy.

Another interpretation of our findings might be that participants rely exclusively on an instance strategy but report rule use whenever the instance is retrieved implicitly so that the participant is unaware of the instance. There is no conclusive evidence in these experiments to refute this possibility, and it is not clear what would constitute sufficient evidence to the contrary. Thus, this is a possibility that will challenge any rule-based interpretation of performance including the prior work of Rickard (1997) and of Delaney et al. (1998) as well as the present study. Alternatively, if all responses are determined by a race between instance retrieval and rule use (e.g., Logan, 1988; Palmeri, 1997), a response could be generated by memory retrieval even though the rule use process was nearly completed. Suppose that any time participants have sufficient evidence of having engaged in rule-based processing, they report "rule" usage. If so, it is conceivable that participants might report "rule" even though their responses were actually generated by instance retrieval. Again, this possibility would challenge our conclusions as well as those of other investigations demonstrating strategy-specific power laws (e.g., Delaney et al., 1998; Rickard, 1997).

Much but not all of the data of these experiments is consistent with extant models. For example, although the models of Logan (1988) and Rickard (1997) were not explicitly designed to account for multiple strategy shifts, our results do not present them with a major challenge. Both of these models assume that instance strength accrues on every exposure to an item. The greater the strength, the more likely instance memory retrieval will prevail over other strategies. In most conditions of our experiments, participants discovered a rule and used it temporarily en route to a final transition to instance-based responding. Within an instance theory framework, the rule-based strategy might be conceptualized as an intermediate state on the way to answer retrieval after sufficient memory strength for instances has accrued. However, neither of these models is explicit about such an intermediate state at this time. Indeed, only the recent multistage model of Anderson et al. (1997) specifically predicts the rise and fall of rule strategy usage over training trials. But even Anderson et al.'s theory offers no account of the variations we found in the use of rule and instance strategies as a function of rule difficulty and task familiarity.

All of these models predict that instance memory retrieval should eventually dominate for any task. Our results for the *thee/thuh* task are interesting as a possible exception to that expectation. Although more evidence is needed, it may be that rule application will persist indefinitely provided the rule can be applied more quickly than instance retrieval in the later stages of practice. Perhaps, if the rule is easy enough, instance strengthening is somehow disengaged and does not continue. Such an instance disengagement mechanism would be at odds with all models as they now stand, and in particular with Logan's (1988) instance theory, which takes as a fundamental assumption that new instances are encoded into and retrieved from memory as an obligatory and unavoidable consequence of attending. Follow-up research addressing this issue is warranted.

REFERENCES

Anderson, J. R.. Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **23,** 932–945.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **10,** 638–648.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General,* **126,** 393–413.

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science,* **9,** 1–7.

Gardiner, J. M., Gawlik, B., & Richardson-Klavehn, A. (1994). Maintenance rehearsal affects knowing, not remembering: Elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review,* **1,** 107–110.

Harvey, L., & Anderson, J. (1996). Transfer of declarative knowledge in complex information-processing domains. *Human-Computer Interaction,* **11,** 69–96.

Healy, A. F., & Sherrod, N. B. (1994, November). *The/Thee pronunciation distinction: A local model of linguistic categories.* Paper presented at the 35th Annual Meeting of the Psychonomic Society. St. Louis, MO.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory,* **7,** 418–439.

Keating, P. A., Byrd, D., Flemming, E., & Todaka, Y. (1994). Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication,* **14,** 131–142.

Kramer, A. F., Strayer, D. L., & Buckley, J. (1990). Development and transfer of automatic processing. *Journal of Experimental Psychology: Human Perception and Performance,* **16,** 505–522.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown Univ. Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review,* **95,** 492–527.

Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 1022–1050.

Logan, G. D., & Stadler, M. A. (1991). Mechanisms of performance improvement in consistent mapping memory search: Automaticity or strategy shift? *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **17,** 478–496.

Lundy, D. H., Wenger, J. L., Schmidt, R. J., & Carlson, R. A. (1994). Serial step learning of cognitive sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 1183–1195.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance,* **21,** 128–148.

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

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the power law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Paivio, A. (1986). *Mental representation: A dual coding approach.* New York: Oxford Univ. Press.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psy-*

chology: Learning, Memory, and Cognition, **23,** 324–354.

Rehder, B., & Pennington, N. (in press). Procedural interference. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Rickard, T. (1994). *Bending the power law: The transition from algorithm-based to memory-based performance.* Unpublished doctoral dissertation. University of Colorado, Boulder.

Rickard, T. (1997). Bending the power law: A CMPL model of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General,* **126,** 288–311.

Rickard, T. C., & Bourne, L. E., Jr. (1995) An identical elements model of basic arithmetic skills. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Learning and memory of knowledge and skills.* Thousand Oaks, CA: Sage.

Schneider, W. (1988). Micro experimental laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, & Computers,* **20,** 206–217.

Serwatka, M., & Healy, A. F. (1998). On the status of the count-mass distinction in a mental grammar. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 113–140). Mahwah, NJ: Erlbaum.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General,* **121,** 278–304.

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies.* Hillsdale, NJ: Erlbaum.

Singley, K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard Press.

Smith, J. D., Murray, M. M., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **23,** 659–680.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist,* **26,** 1–12.

(Received October 12, 1998)
(Revision received February 3, 1999)