Journal of Experimental Psychology:
Learning, Memory, and Cognition
1991, Vol. 17, No. 3, 478–496

Copyright 1991 by the American Psychological Association, Inc.
0278-7393/91/$3.00

Mechanisms of Performance Improvement in Consistent Mapping Memory Search: Automaticity or Strategy Shift?

Gordon D. Logan
University of Illinois

Michael A. Stadler
Louisiana State University

Five experiments compared four possible mechanisms underlying changes in performance with practice in consistent-mapping memory search. Subjects were presented with a catch trial in which the probe was a member of the 'superset' from which memory sets were sampled but was not a member of the current memory set. Process-improvement mechanisms predict fast correct rejections of these probes; process-switching mechanisms predict false alarms or slow correct rejections. Overall, half of the subjects false alarmed on catch trials and the remainder rejected the probe slowly, which supports process-switching mechanisms over process-improvement mechanisms. Experiments 3–5 attempted to distinguish among three process-switching mechanisms and found evidence more consistent with a strategy shift than with item-based learning (automatization).

Practice has dramatic effects on performance in consistent task environments. What was once difficult becomes easy; what was slow and effortful becomes fast and effortless; what was deliberate becomes automatic (for reviews, see Logan, 1985b; Schneider, Dumais, & Shiffrin, 1984). Several mechanisms may produce these dramatic changes. The process itself may improve, or it may be replaced by another, more efficient process. The replacement may be deliberate or strategic, or it may occur inevitably as the new process strengthens. It is important theoretically to be able to tell which mechanism or set of mechanisms operates in a given case because all theories cannot account for all mechanisms. The appropriateness of a theoretical approach to a given task hangs in the balance.

The purpose of this article is to examine four mechanisms that might produce performance improvements in consistent mapping memory-search tasks, one process improvement mechanism, and three process-switching mechanisms. Memory-search tasks have been the paradigmatic focus of Schneider and Shiffrin's (1977; Shiffrin & Schneider, 1977) theory of automaticity, and debates about their theory have focused on mechanisms of performance improvement (e.g., Cheng, 1985; Ryan, 1983). The present data bear on the validity of the memory-search task as a paradigm for investigating their theory of automaticity.

Memory Search

In a memory-search task, subjects are presented first with a set of items to memorize and then with a probe item. Their task is to indicate as quickly as possible whether the probe is a member of the memory set. Experimenters vary the number of items in the memory set and find that reaction time increases linearly with set size, up to sets of 5–7 items, with a slope of 20–50 ms per item (for a review, see Sternberg, 1975).

In consistent mapping (CM) memory search, the population of items is divided into two sets—positive and negative—and memory sets are sampled only from the superset of positive items. The mapping of stimuli onto responses is consistent in that the same items require the same responses throughout practice; targets are never distractors, and distractors are never targets. If subjects practice extensively with consistent mapping, the slope of the set-size function decreases from the initial 20–40 ms per item to nearly zero (e.g., Logan, 1978; Strayer & Kramer, 1990).

Consistent mapping is often contrasted with varied mapping (VM). In VM memory search, memory sets are sampled from the entire population of items, so that items that once were distractors can become targets, and targets can become distractors; the mapping of items onto responses varies. Extensive practice with varied mapping has little effect on the slope of the set size function. It remains stable at 20–40 ms per item over many sessions (e.g., Burrows & Murdock, 1969; Logan, 1978). Theories of performance change must explain why consistent mapping benefits from practice but varied mapping does not.

Schneider and Shiffrin on Automaticity

Schneider and Shiffrin (1977; Shiffrin & Schneider, 1977) distinguished between two qualitatively different modes of processing, automatic and controlled processing. Automatic processing was fast, parallel, and not limited by the capacity of short-term memory, whereas controlled processing was slow, serial, and limited by the capacity of short-term memory. In broad conception, their distinction was similar to then-

---

This research was supported by National Science Foundation Grant BNS 8510365 and BNS 8811026 to Gordon D. Logan. Experiments 1–3 were presented at the annual meeting of the Psychonomic Society in 1986. We would like to thank Liz Olzsyck and Tom Rodenbeck for running subjects in Experiments 1 and 2 and Julie Delheimer for running subjects in Experiments 4 and 5. This research began when we were both at Purdue University and continued when we moved to our present institutions. We are grateful to our colleagues at Purdue for supporting us while we were there and to Alexander Graham Bell and the inventor of electronic mail for sustaining us since we left.

Correspondence concerning this article should be addressed to Gordon D. Logan, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820. Electronic mail may be sent to glogan@s.psych.uiuc.edu.

contemporary distinctions by LaBerge and Samuels (1974) and Posner and Snyder (1975). The details were different, perhaps because Schneider and Shiffrin applied their theory primarily to memory-search and visual-search tasks.

In memory search, controlled processing involved a serial, self-terminating comparison of the probe with a representation of the memory set in short-term memory, whereas automatic processing involved an immediate, parallel detection of a match between the probe and the memory set. The serial nature of controlled processing predicts a substantial slope in the set-size function: The parallel nature of automatic processing predicts a slope of zero. All memory-search tasks were assumed to require controlled processing initially. Automatic processing could develop with CM but not with VM. Thus with CM set-size-function slopes should decrease with practice and asymptote near zero, as automatic processing takes over from controlled processing. With VM, however, slopes should remain relatively constant because controlled processing is always required.

The theory contained in these assumptions accounted for existing data very well. It stimulated the growth of a large literature on practice effects in memory-search tasks, which is mostly consistent with the theory (for a review, see Schneider et al., 1984). In much of this literature, the logic is inverted: Zero slopes are interpreted as evidence of automatic processing, and slopes substantially above zero are interpreted as evidence of controlled processing. The reduction in slope with practice with consistent mapping is interpreted as a transition from controlled to automatic processing.

Critics of the theory took issue with the inversion of the logic (e.g., Cheng, 1985; Ryan, 1983). They proposed other mechanisms to account for the performance changes and argued that zero slopes and so on could not distinguish their proposals from those of Schneider and Shiffrin (1977, but see Schneider & Shiffrin, 1985; Shiffrin & Schneider, 1984). When many theories predict the same observation, one cannot reason backwards from observation to a single theory.

Critics also took issue with Schneider and Shiffrin's (1977) failure to specify the details of the learning mechanism (Cheng, 1985; Ryan, 1983). Shiffrin and Schneider (1977) argued for association learning (establishing connections between nodes in long-term memory) and the development of an automatic attention response (whereby targets practiced under consistent mapping conditions appear to pull attention), but they said nothing about the processes that underlay them and nothing about the quantitative details of their functioning. More recently, Schneider (1985a; Schneider & Detweiler, 1987) and Shiffrin (Shiffrin & Czerwinski, 1988) have worked on these issues, but their earlier papers said little about them. Moreover, what little they said can be interpreted many ways; at least two of four performance improvement mechanisms discussed next can be construed as consistent with what Shiffrin and Schneider originally wrote.

Mechanisms Underlying Performance Change

There are at least four distinct mechanisms that could produce the performance improvements seen in CM memory-search tasks. All four predict that slopes will flatten with practice and asymptote at zero.

Process-Based Learning

Process-based learning is a process improvement mechanism in which the task is performed in essentially the same way throughout practice, only more efficiently. In memory search, the probe is initially compared with an explicit representation of the memory set in short-term memory. The comparison may be done in series, as Shiffrin and Schneider (1977) claimed, or in parallel in a capacity-limited fashion so that the set-size slope is greater than zero (see Townsend & Ashby, 1983). Practice improves the efficiency of the comparison process somehow (e.g., by reducing resource requirements) so that the probe can be compared with the memory set in a parallel unlimited-capacity fashion, producing a slope of zero. An important characteristic of this view is that at all stages of practice, performance depends on comparing the probe with an explicit representation of the memory set in short-term memory. Learning changes the characteristics of the comparison process from serial to parallel or from limited to unlimited capacity, but the process remains the same. Probes are compared with an explicit representation of the memory set.

Many readers may attribute this view to Shiffrin and Schneider (1977). It is certainly consistent with their claims about the serial nature of controlled processing, the parallel nature of automatic processing, and the idea that one develops from the other. However, it is not a necessary implication of their writings, so its success or failure should not bear directly on the validity of their theory.

This view is consistent with the zeitgeist of the 1970s in which Shiffrin and Schneider's (1977) theory was published. Researchers had been modeling search tasks for a decade, and the serial or parallel nature of the search task and the nature of its capacity limitations were subject to vigorous debate (for a review, see Townsend & Ashby, 1983). Researchers also knew that some search tasks produced results suggestive of serial limited-capacity processing and that others produced results suggestive of parallel unlimited-capacity processing (e.g., Corballis, 1975; Sternberg, 1975). The idea that practice could transform one kind of search into another was easy to accommodate with this zeitgeist. Researchers thought that capacity limitations forced processing to be serial (e.g., Shiffrin & Gardner, 1972), and many believed that capacity demands could diminish with practice (e.g., LaBerge & Samuels, 1974; Logan, 1978), which would allow processing to become parallel. Hardly anyone was interested in learning, so the question of how such changes were possible was not a high priority.

Item-Based Learning

Item-based learning is a process-switching mechanism in which subjects associate particular items with particular responses or response categories and base subsequent performance on retrieved associations (Logan, 1988; Schneider, 1985a). In memory search, probes are initially compared with explicit representations of the memory sets in short-term memory, but with consistent mapping practice, probes retrieve responses or response categories associated with them directly from long-term memory. They are not compared with an explicit representation of the memory set in short-term memory (Schneider, 1985a; Strayer & Kramer, 1990). The slope of the set-size function asymptotes at zero because retrieval from long-term memory is independent of set size. Subjects do not compare probes with memory sets and may not even bother to encode the current set in short-term memory.

Many readers may attribute this view to Shiffrin and Schneider (1977) as well. It is consistent with their claim that automatic and controlled processing are qualitatively different, though the idea that automatic processing is parallel needs some reinterpretation: Subjects are not switching from a serial to a parallel version of the same process; they are switching from one process (short-term memory search) to another (long-term memory search) that happens to be parallel. Again, it is not clear that this view is a necessary implication of Shiffrin and Schneider's theory, so its success or failure need not bear too strongly on the validity of their theory.

Item-based learning is not unique to Shiffrin and Schneider's (1977) theory. Indeed, it was the alternative that Ryan (1983) offered in his criticism of their theory. Item-based learning is also predicted by several modern approaches to learning, including Schneider's (1985a) connectionist model, Logan's (1988) instance theory, and Shiffrin and Czerwinski's (1988) strength model. These modern approaches specify the learning mechanism in sufficient detail that quantitative aspects of performance change can be predicted quite precisely. These theories stand or fall on evidence of item-based learning.

Category Comparison Strategy

The category comparison strategy is a process-switching mechanism. It involves learning a general category that distinguishes positive from negative items and then responding to probes by assessing their category membership. Category membership can be decided on the basis of information that the probe retrieves from long-term memory; it does not require the probe to be compared with the memory set in short-term memory.

The category comparison strategy predicts a zero slope because category membership of the probe is independent of the number of items in the memory set. Subjects pay no attention to the memory set, so naturally their performance should be independent of its size.

The category comparison strategy can be viewed from different perspectives, some more consistent and some less consistent with Shiffrin and Schneider's (1977) theory of automaticity. It depends mainly on how category comparison is done. If category comparison is done by comparing probes with previous instances or exemplars in long-term memory (e.g., Hintzman, 1986; Medin & Schaffer, 1978), it is not much different from item-based learning (e.g., Logan, 1988) and, thus, is consistent with Shiffrin and Schneider's theory. But if category comparison involves assessing some superficial property like color (e.g., Ellis & Chase, 1971) or something readily accessible like membership in a taxonomic category (e.g., Jones & Anderson, 1982), it seems more strategic, a parlor trick but not a skill, and therefore unlike Shiffrin and Schneider's conception of automaticity. This was essentially Cheng's (1985) view of category comparison when she offered it as a theoretical alternative to automaticity. If clearly strategic category comparison is responsible for the performance improvements seen in memory search, then Shiffrin and Schneider's theory of automaticity cannot apply to memory search.[1]

---

[1] Shiffrin and Schneider (1977; Schneider & Shiffrin, 1985) took great pains to distinguish automaticity from categorization. Their argument in 1977 and 1985 rested primarily on Experiment 3 from their 1977 article. That experiment used a multiple-frame search task in which memory-set size was two or four. For the first 24 training sessions, the items (eight letters) were divided into two sets of four (i.e., two categories). The procedure involved varied mapping, in that each category served as target as often as it served as distractor, but the categorical distinction was preserved, in that targets and distractors were never drawn from the same category on the same trial. When targets were drawn from the first category, all of the distractors were drawn from the second; when targets were drawn from the second category, all distractors were drawn from the first. Shiffrin and Schneider thought that this procedure would allow subjects to develop categories because the two sets of items were kept distinct, but it would not allow them to develop automaticity because mapping of stimuli onto responses varied over trials.

Over the 24 training sessions, performance improved markedly, relative to traditional VM controls, and in the last 4 sessions there was no difference between Set Size 2 and Set Size 4. Both of these effects would be expected if automaticity had developed. However, Shiffrin and Schneider (1977) did not interpret these effects as evidence of automaticity. They ran an additional 10 sessions in which the category group and the VM control group were transferred to traditional CM; in the category group, targets were drawn consistently from one set of items and distractors were drawn consistently from the other. Performance improved appreciably in these pure CM sessions. Shiffrin and Schneider interpreted the improved performance as evidence that automaticity developed after the switch to pure CM and as evidence that performance improvements during the initial 24 sessions were due to categorization rather than automaticity.

One can question Shiffrin and Schneider's (1977) interpretation on several grounds. First, the elimination of the set-size effect in the first 24 sessions is suggestive of automaticity. Second, the improvement relative to VM controls in the first 24 sessions is suggestive of automaticity. Third, Shiffrin and Schneider attempted to provide converging evidence by presenting two targets on some trials and examining the effects of lag between the two presentations. In the categorical condition, the pattern was more characteristic CM performance than VM performance. Moreover, when subjects were transferred to the pure CM condition, the lag effects for repeated stimuli were the same as those observed in the categorical condition. If that pattern is diagnostic of automaticity after the switch to pure CM, it should also be diagnostic of automaticity before the switch. Fourth, the improvement in performance following the switch to pure CM can be interpreted as a "letting go" phenomenon that Schneider refers to in later publications (e.g., Schneider, 1985b): He reported that subjects are sometimes reluctant to abandon controlled processing and rely on automatic processing. He suggested that if subjects are encouraged to let go, they will switch from controlled to automatic processing, and their performance will improve suddenly and dramatically. It is entirely possible that this is what was observed after the shift to pure CM in Shiffrin and Schneider's Experiment 3. Consequently, automatic processing and categorization may not be as distinct as Shiffrin and Schneider (1977; Schneider & Shiffrin, 1985) claimed they were.

Superset Strategy

The superset strategy is a process-switching mechanism that involves learning the superset of positive items from which all memory sets are sampled and then comparing the probe against the superset rather than the current memory set. The superset, not the current memory set, is represented explicitly in short-term memory. This strategy would lead to an apparent performance improvement because performance would no longer depend on the size of the current memory set. The slope would be zero, not because comparisons were done in parallel, but rather because the probe would be compared with the same superset regardless of the size of the memory set. Performance would be independent of set size because subjects ignore the memory set.

As it stands, this view predicts that superset-based responding must be at least as slow as responding to the largest memory set because the superset must be at least as large as the largest memory set. Reaction time should increase as the slope decreases. This is never found (see, e.g., Logan, 1978; Strayer & Kramer, 1990). However, the view need not make this prediction if other reasonable assumptions are made: Preparing a new memory set for comparison with a probe takes time and effort (see, e.g., Logan, 1978); preparing the same superset on each trial should be easier and faster. It should benefit from memory for past preparations, and it need not be time-locked to the experimental events, such as the presentation of the memory set. Also, it may be easier to learn associations between probes and one superset than between probes and several different memory sets, so subjects who use the strategy may be further down the learning curve than subjects who do not. Either or both of these factors could speed up nonsearch processes and offset increases in search time caused by the superset strategy. The point is not to argue for any given degree of compensation, but rather to argue that the possibility of compensation weakens the prediction that superset-based responses must be at least as slow as responses to the largest memory set.

No reader would attribute the superset strategy to Shiffrin and Schneider (1977). If the superset strategy is responsible for the performance improvements in memory search, then memory search cannot reveal much about their conception of automaticity.

Catch-Trial Method

The different mechanisms can be distinguished from each other on catch trials in which subjects are presented with a probe that is a member of the positive superset but not a member of the current memory set (cf. Gleitman & Jonides, 1976). Performance on catch trials allows process improvement to be distinguished from process switching in a straightforward manner: Process improvement predicts that subjects will respond *no* on catch trials, correctly rejecting the probe, whereas process switching predicts that they will respond *yes*, false alarming to the probe. The different kinds of process switching are harder to distinguish because they all predict false alarms, but they differ somewhat in the conditions under which false alarms will be emitted.

Process-based learning. Process-based learning predicts that subjects will respond *no* on the catch trial because performance is based on a comparison of the probe with an explicit representation of the memory set in short-term memory. Practice makes comparison more efficient, but probes are still compared with memory sets. The catch-trial probe is not in the memory set, so it will be rejected correctly.[2]

Item-based learning. Item-based learning predicts that subjects will respond *yes* on the catch trial because subjects respond in accord with associations that they retrieve from long-term memory. Catch-trial probes have been consistently associated with *yes* responses, so subjects will false alarm to them. The probe is not compared with the memory set in short-term memory, so there is no tendency to say *no*.

Category comparison strategy. In the category comparison strategy, subjects evaluate the category membership of the probe and ignore the current memory set. Because the probe is a member of the same category as all memory sets, subjects will respond *yes*.

Superset strategy. In the superset strategy, subjects compare the probe with the positive superset, not with the current memory set. Because the probe is a member of the positive superset, subjects will respond *yes*.

False Alarms and Slow Correct Rejections

The tendency to respond *yes* on catch trials may appear as a slow correct rejection instead of a fast false alarm (Hock, Rosenthal, & Stenquist, 1985). The probe may evoke a tendency to respond *yes* that competes with the correct *no* response and slows its execution, or subjects may deliberately initiate a *yes* response, decide that it is a false alarm, inhibit it, and then execute a *no* response. Error detection, inhibition, and reprogramming all take time, so catch-trial responses would be slower than controls, but no less accurate.

Interpretation is clearest when subjects overtly respond *yes* on the catch trial. Then, there is little doubt that one of the process-switching mechanisms was in operation and subjects did not compare the probe with the memory set. But when subjects respond *no* slowly, they must have compared the probe with the memory set. The memory-search process could be the winner of the response competition or the basis of correcting a deliberate *yes* response, but either way it seems logically necessary. In order for process-switching mechanisms to influence the correct rejection, they would have to precede memory search or go on in parallel with it. The long-term memory retrieval involved in item-based learning might go on in parallel with short-term memory search. Deliberate strategies such as category comparison and superset search would have to precede memory search (and fail) in order to influence performance.

---

[2] Subjects may produce false alarms if they misperceive or misremember the memory set and happen to substitute the catch-trial item for the misperceived or forgotten memory-set item but that is unlikely given the generally high accuracy in memory-search tasks. However, process-based learning predicts fast correct rejections and no false alarms. We consider the possibility of misperception and misremembering in detail below.

Regardless of the detailed interpretation, slow correct rejections suggest an underlying tendency to respond *yes*, and that is important. It means that the false-alarm rate may underestimate the proportion of trials on which subjects used a process-switching mechanism. Underestimation may be significant early in practice during the transition from memory search to the new process, when subjects do not rely exclusively on one process or the other.

The Experiments

We conducted 5 experiments. The first 2 asked whether subjects would respond on catch trials, and the last 3 asked why. The first 2 discriminated process improvement from process switching, and the last 3 attempted to discriminate among process-switching mechanisms. All of the experiments used a single session, which is a low level of practice for many studies that compare CM and VM (e.g., Shiffrin & Schneider, 1977). We should not expect asymptotic differences between consistent and varied mapping in a single session, but the differences should begin to emerge. In fact, several studies have shown differences between CM and VM in the first practice session (e.g., Ackerman, 1988; Fisk & Schneider, 1983; Kramer & Strayer, 1988; Schneider & Fisk, 1982, 1983). The low level of practice is important to our ability to distinguish strategy shifts from automaticity: Strategy shifts should occur at levels of practice that are too low for automaticity to have developed. The early experiments examined catch-trial performance at the end of a 768-trial session; later experiments (3 and 5) examined catch-trial performance earlier in practice. Experiment 5 presented a catch trial on the very first trial. Performance in these latter experiments is more likely to reflect strategy shifts than automaticity.

Experiment 1

Experiment 1 involved three different conditions. The first was a *categorical CM* condition, in which the positive items were letters and the negative items were digits (or vice versa). This follows Schneider and Shiffrin's (1977) initial experiments and represents the memory-search analog of Gleitman and Jonides's (1976) catch-trial study. The second was a *letter CM* condition, in which a set of letters was divided into positive and negative supersets, which were then mapped consistently onto responses, following experiments in which subjects learn to search through arbitrary sets automatically. This may be more typical of memory-search studies of automatization. The third condition was a *letter VM* condition, which used the same set of letters as the letter CM condition but did not divide them into positive and negative supersets. Each letter was a target as often as it was a distractor. The letter VM condition served as a control against which to evaluate performance changes in the CM condition and the tendency to false alarm on catch trials. In view of previous research, VM performance should not change much with practice, and there should be no false alarms on catch trials.

Each subject received 768 training trials followed by a single catch trial. A memory set and a probe were presented on each trial. During training, memory-set size varied from one to four, and each set size was presented equally often. On the catch trial, set size was one, two, or three, and the probe was the fourth member of the positive superset in the (letter and categorical) CM conditions. The VM condition was the same except that there was no positive superset, so the probe was simply another item that was not in the current memory set.

Method

Subjects. There were 72 subjects in all; 24 in each (letter CM, categorical CM, and letter VM) condition. Eight subjects in each condition received catch trials at each of the three memory-set sizes (one, two, and three). All of the subjects were introductory psychology students at Purdue University who served to fulfill course requirements. All of them reported English as their native language and had normal or corrected vision.

Apparatus and stimuli. The experiments were controlled by an IBM XT computer. Stimuli were displayed on an Amdek monitor (model 300) using the standard character set of the computer. Each letter subtended 0.35° × 0.56° of visual angle at a viewing distance of 50 cm. The viewing distance and the visual angles are only approximate because viewing distance was not fixed (e.g., by a headrest). The stimuli were the digits *1*, *2*, *3*, *4*, *5*, *6*, *7*, and *8* and the letters *D*, *F*, *H*, *K*, *N*, *R*, *T*, and *W*. Subjects made their responses by pressing the "\" and "/" keys, which were the leftmost and rightmost keys on the bottom row of the XT keyboard.

In the CM conditions, subjects were paired (though tested individually) so that one used a set of stimuli for the positive superset while another used the same stimuli for the negative superset, and vice versa. The stimuli were drawn from their respective pools at random for the 12 pairs of subjects.

Categorical CM condition. In this condition, the positive superset contained members of one category (e.g., letters) and the negative superset contained members of the other (e.g., digits). Twelve different sets of four letters and four digits were constructed, and each set was assigned to 2 subjects, one with letters in the positive superset and one with digits.

Subjects received 768 trials of CM memory search. To facilitate analysis of practice effects, the 768 trials were divided into two blocks of 384 in which all experimental factors were completely balanced. Half of the trials in each block used probes from the positive superset, and half used probes from the negative superset. There was an equal number of positive and negative probes (48 per block, 96 total) at each set size, one through four. Within these conditions, each possible combination of memory set and probe occurred an equal number of times. The assignment of items to positions in the memory sets as they were displayed on each trial was random.

Catch trials were constructed by selecting the probe item at random from the positive superset. For 8 of the subjects, one of the other items was drawn at random to be the memory set, for another 8 two of the items were drawn, and for the last 8 all three remaining items were used in the memory set.

Letter CM condition. In this condition, positive and negative supersets contained letters. The population of eight letters was divided into two supersets of four in 12 different ways. Each pair of supersets was assigned to 2 subjects such that the positive superset for one subject was the negative superset for the other (and vice versa). Otherwise, the 768 experimental trials and the single catch trial were constructed in the same way as in the categorical CM condition.

Letter VM condition. The VM condition used the same number of trials as the CM conditions (i.e., 768 divided into two balanced blocks of 384), and again, there was an equal number of positive and negative probes at each set size. The eight letters were used equally often for each of the combinations of response set size and response type (which makes it varied mapping). The items were ordered randomly in the memory-set displays.

An additional negative trial was included at the end of the session to allow comparison to the catch trial in the CM conditions. Eight of the subjects had one item in the memory set for this last trial, 8 had two items, and 8 had three items.

Procedure. A trial began with the presentation of a fixation point ("+") for 500 ms in the center of the screen. The items in the memory set were presented simultaneously immediately after the fixation point was removed. The first item was positioned one line above and one character to the left of fixation. If there was more than one item in the memory set, the second item occupied the position above and to the right of fixation, the third was positioned below and to the left, and the fourth was below and to the right. The memory set remained on the screen for 2,000 ms. A question mark was then displayed for 500 ms in the position that had been occupied by the fixation point. It was immediately replaced by the probe, which appeared in the same position as the fixation point. The probe remained on the screen for 1,500 ms or until the subject responded, whichever came first. Each subsequent trial began 5 s after the onset of the fixation point for the preceding trial. Subjects were given seven sets of 96 trials and an eighth block of 97 trials (which included the catch trial). Subjects were allowed short breaks between 96-trial sets.

At the beginning of the session, subjects were given general instructions about the nature of the task. They were told to respond as quickly as possible while maintaining at least 90% accuracy. Half of the subjects were instructed to respond by pressing the "\" key if the probe was in the memory set and "/" if it was not. The other half of the subjects received the opposite instruction.

Results

The primary focus of this article is on subjects' behavior on catch trials. It is also important, however, to examine their performance on the 768 trials that led up to the catch trial. The experiment should replicate standard effects in the memory-search literature. Consequently, the results are divided into two sections, a *training* section dealing with the initial 768 trials and a *catch-trial* section dealing with the catch trial.

Training. Mean reaction time and error rates were calculated for each subject in each set-size and response-type condition in each of the 384-trial blocks in the experiment. The mean reaction times across subjects are presented in Figure 1.

The reaction time results replicate several standard effects in the literature. The VM condition produced the linear increase with set size that is characteristic of memory search (Sternberg, 1975); *yes* responses were faster than *no* responses, but the slopes were roughly the same (52 ms per item vs. 45 ms per item). The consistent mapping conditions produced the flatter slopes that are characteristic (Schneider & Shiffrin, 1977) of CM search (12 ms per item in categorical CM and 23 ms per item in letter CM). The slope difference between CM and VM appeared in the first block of trials (first block slopes were 12 ms per item for categorical CM, 25 ms per item for letter CM, and 52 ms per item for letter VM). Performance improved from the first block to the second, but the intercept changed more than the slope.

These conclusions were confirmed in an analysis of variance (ANOVA) on the reaction time data (Group × Set Size × Response Type × Practice Block, with subjects nested in group). There were significant main effects of set size, $F(3, 207) = 197.28$, $p < .01$, $MS_e = 1{,}871.00$; response type, $F(1, 69) = 47.12$, $p < .01$, $MS_e = 4{,}896.38$; and practice, $F(1, 69) = 12.90$, $p < .01$, $MS_e = 14{,}433.70$. There were significant interactions between set size and practice, $F(3, 207) = 4.05$, $p < .01$, $MS_e = 788.85$; set size and group, $F(6, 207) = 30.91$, $p < .01$, $MS_e = 1{,}871.00$; and group and response type, $F(2, 69) = 7.38$, $p < .01$, $MS_e = 4{,}896.38$. The interaction between group, set size, and response type approached significance, $F(6, 207) = 1.96$, $p = .08$, $MS_e = 906.65$.

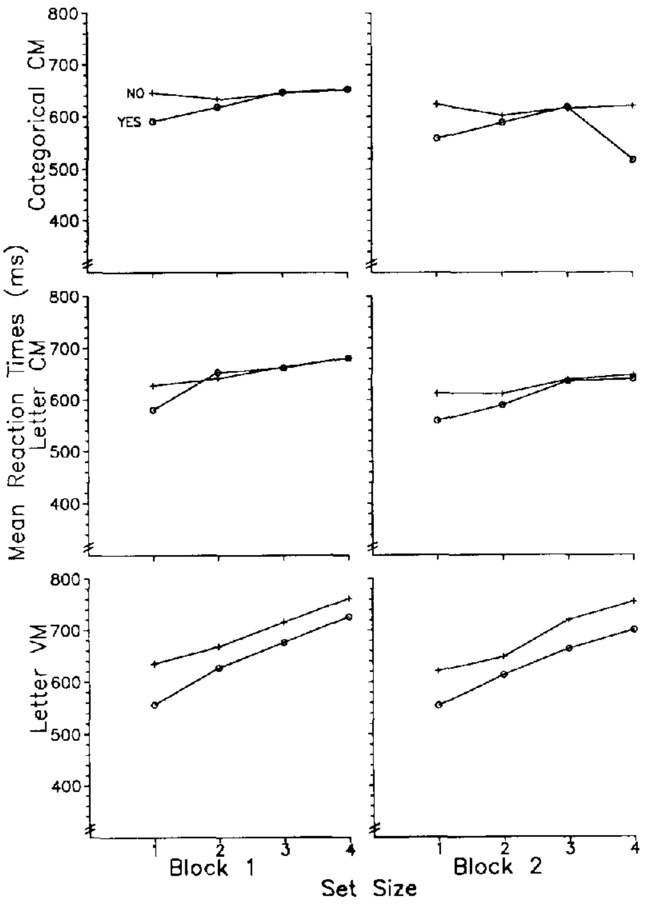

Figure 1. Mean reaction times for each condition of Experiment 1 as a function of set size, response type, and practice block. (CM = consistent mapping; VM = varied mapping.)

The interaction between set size, practice, and group was not significant, $F(6, 207) = 1.02$, $MS_e = 788.85$, but separate ANOVAs on each group revealed a significant Set Size × Practice interaction in the letter CM condition, $F(3, 69) = 4.23$, $p < .01$, $MS_e = 848.49$, but not in the categorical CM condition, $F(3, 69) < 1$, $MS_e = 639.57$, or the letter VM condition, $F(3, 69) = 1.23$, $MS_e = 878.50$.

The error rates for each combination of conditions are presented in Table 1. Errors were relatively rare, and there were no effects that would compromise the conclusions drawn from the reaction time results.

The difference in slope between CM and VM conditions in the first block was surprising because, in theory, slope differ-

Table 1
Mean Proportion Error in Training as a Function of Set Size, Response Type, and Practice Block for Each Condition of Experiment 1

	Yes responses				No responses			
Condition	1	2	3	4	1	2	3	4
Categorical CM								
Block 1	0.02	0.03	0.02	0.02	0.02	0.01	0.02	0.01
Block 2	0.02	0.03	0.03	0.02	0.02	0.02	0.02	0.02
Letter CM								
Block 1	0.04	0.05	0.06	0.04	0.04	0.02	0.03	0.03
Block 2	0.04	0.04	0.05	0.05	0.03	0.03	0.03	0.03
Letter VM								
Block 1	0.06	0.05	0.08	0.08	0.04	0.03	0.03	0.04
Block 2	0.03	0.05	0.05	0.06	0.03	0.04	0.04	0.03

Note. CM = consistent mapping; VM = varied mapping.

ences emerge with practice. We conducted a finer grained analysis of practice effects to see whether we could find equivalent CM and VM slopes early in practice and changes in CM slopes over practice. We broke the first 192 trials down into blocks of 24 trials and analyzed reaction times as a function of memory-set size. We collapsed the data over *yes* and *no* responses to increase the stability in the data. (The experiment was designed so that each memory set occurred equally often with *yes* and *no* responses in each set of 96 trials. Thus, there was no guarantee that every subject would have trials in each cell of the Set Size × Response Type design in blocks of 24 trials. This became possible when we collapsed over response type.) This analysis showed changes in the CM conditions over practice, but CM slopes were shallower than the VM slope even in the first block of trials. For the first 24 trials, the slope of the set-size function was 22 ms per item in the categorical CM condition, 36 ms per item in the letter CM condition, and 52 ms per item in the letter CM condition. By the eighth block (Trials 169–192), the slopes decreased to 9, 22, and 44 ms per item, respectively.

The significance of these differences was tested with planned comparisons, which assessed differences in linear trend (all MS_es = 4,091.38). In the first block, the slope in the categorical CM condition was shallower than that in the letter VM condition, $F(1, 1449) = 40.04$, $p < .01$, and the slope in the letter CM condition was shallower than that in the letter VM condition, $F(1, 1449) = 11.24$, $p < .01$. There were significant differences in slope between the first and eighth blocks in the categorical CM condition, $F(1, 1449) = 7.09$, $p < .01$, and in the letter CM condition, $F(1, 1449) = 8.26$, $p < .01$, but not in the letter VM condition, $F(1, 1449) = 2.76$, $p < .10$. Contrasts comparing the practice effects in between the CM and VM conditions (i.e., comparing between-conditions differences in differences in linear trend between Blocks 1 and 8) were not significant (both Fs < 1.0).

As a further check for differences between VM and CM conditions early in practice, we calculated the slopes of linear functions relating reaction time to memory-set size for the first presentation of each set size. The slopes, averaged across subjects, were 26, 38, and 38 ms per item for categorical CM, letter CM, and letter VM, respectively. For the letter conditions, at least, there was no difference in performance on the first exposure to each set size.

Catch trials. The catch trial required subjects to respond to a probe from the positive superset that was not a member of the current memory set. The appropriate response was *no*—a correct rejection. The three process-switching mechanisms (item-based learning, category comparison, and superset strategy) predict a strong tendency to respond *yes*—to false alarm—on catch trials, because subjects do not compare the probe with the current memory set. The one process improvement mechanism (process-based learning) predicts no tendency to false alarm; the probe is not a member of the current memory set, so it should be rejected as easily as any other negative item.

False-alarm rates from the catch trials in each condition are presented in Table 2. Thirteen of the 24 categorical CM subjects and 8 of the 24 letter CM subjects false alarmed on the catch trial. These figures can be compared with the false-alarm rates for the last negative trial that preceded the catch trial, also presented in Table 2, which represents a comparable level of practice. Only 1 subject in each of the CM groups false alarmed on the last negative trial. The CM catch-trial data can also be compared with data from VM catch trials; only 1 of the 24 VM subjects false alarmed on the catch trial.

The size of the memory set on the catch trial affected the false-alarm rate as well. Subjects in the CM groups were more likely to false alarm the larger the set size, whereas subjects in the VM groups were not. Also, false-alarm rates for the last negative trial were relatively unaffected by set size.

These conclusions were confirmed in binomial probability tests, which compared the observed number of false alarms with the number expected given the number of observations and an expected probability of false alarm, taken either from the last negative trial or from the false-alarm rate in the second block of 384 trials. Averaging across set size, there were significantly more false alarms on the catch trial than on the last negative trial ($p < .01$) or in the second block ($p < .01$) for the categorical CM and letter CM groups, but not for the letter VM group ($p > .05$). There were significantly more false alarms in categorical CM than in letter CM ($p <$

Table 2
False Alarm Rates (as Proportions) for the Catch Trial and the Last Negative Trial as a Function of Set Size for Each Condition of Experiment 1

	Set size			
Condition	1	2	3	*M*
Categorical CM				
Catch trial	0.25	0.63	0.75	0.54
Last negative trial	0.13	0.00	0.00	0.04
Letter CM				
Catch trial	0.13	0.38	0.50	0.33
Last negative trial	0.00	0.13	0.00	0.04
Letter VM				
Catch trial	0.00	0.00	0.13	0.04
Last negative trial	0.00	0.13	0.00	0.04

Note. CM = consistent mapping; VM = varied mapping.

.05), using either the categorical or letter CM false-alarm rate to generate binomial probabilities.

The effect of the catch trial was assessed at each set size, using the second-block false-alarm rates to generate binomial probabilities. In these analyses, catch-trial false-alarm rate was elevated significantly for each set size ($p < .05$ for a set size of 1; $p < .01$ for set sizes of 2 and 3) in categorical CM and for set sizes of 2 and 3 in letter CM ($p < .01$).

The increase in catch-trial false-alarm rate with set size was tested by generating binomial probabilities in two ways. One analysis used false-alarm rates from Set Size 1 to generate binomial probabilities; the other used false-alarm rates from Set Size 3. In the categorical CM condition, Set Size 2 and 3 both differed significantly ($p < .05$) from Set Size 1, but only Set Size 1 differed significantly ($p < .01$) from Set Size 3. In the letter CM condition, Set Size 3 differed significantly from Set Size 1 ($p < .01$), and Set Size 2 was marginally different from Set Size 1 ($p < .07$). Set Size 1 but not Set Size 2 differed significantly ($p < .05$) from Set Size 3.

The tendency to false alarm can also appear in reaction time data. Subjects may be able to suppress or inhibit the tendency and respond correctly on the catch trial, but the suppression or inhibition should increase the time taken to respond. Thus for CM subjects, process-switching mechanisms predict slower correct rejections on the catch trial than on the last negative trial, whereas process improvement mechanisms predict no difference.

Reaction times for catch trials and last negative trials were averaged over set size for stability. Categorical CM subjects took 1,120 ms to correctly reject the catch trial versus 636 ms to correctly reject the last negative trial. For letter CM subjects, the figures were 979 ms versus 642 ms. Compared with these large differences, the difference in the VM condition was quite small (760 ms for correct rejections on the catch trial vs. 694 for correct rejections on the last negative trial).

The significance of the difference between the catch trial and the last negative trial was tested with correlated *t* tests. It was significant in the categorical CM group, $t(10) = 2.63$, $p < .05$, $MS_e = 176.63$, and the letter CM group, $t(15) = 4.24$, $p < .01$, $MS_e = 81.06$, but not the VM group, $t(22) = 1.38$, $MS_e = 57.18$.

The same point can be made by comparing catch-trial reaction times for CM and VM subjects. Process switching predicts slower correct rejections for CM subjects; process improvement predicts no difference. Again the data were clear. Categorical CM subjects took 360 ms longer then VM subjects to correctly reject the catch trial; letter CM subjects took 219 ms longer than VM subjects. We compared the catch-trial reaction times from subjects who correctly rejected in a one-way ANOVA. The overall *F* was significant, $F(2, 47) = 5.84$, $p < .01$, $MS_e = 90{,}746.56$. Fisher's least significant difference test showed significant differences ($p < .01$) between categorical CM and VM subjects and between letter CM and VM subjects.

Could the tendency to false alarm on the catch trial have been predicted from the training data? In the categorical CM condition, subjects who false alarmed were faster than those who correctly rejected (569 vs. 672 ms), $F(1, 22) = 6.94$, $p < .05$, $MS_e = 147{,}779.49$, but there were no other differences. In particular, there was no interaction between set size and false alarm versus correctly reject on the catch trial. In the letter CM condition, subjects who false alarmed were slightly faster than subjects who correctly rejected (621 vs. 638 ms), but the difference was not significant, $F(1, 22) < 1$. There were no significant interactions involving the nature of the catch-trial response.

It could be argued that catch-trial false alarms might be produced by process-based learning if subjects misperceived or misremembered the memory set and substituted superset items for the misperceived or forgotten memory-set items.[3] If the substituted items were presented on the catch trial, subjects would compare them with what they believed to be the memory set and decide (incorrectly) that they warranted a *yes* response. This hypothesis could account for some of the false alarms that we observed on catch trials, making it difficult to distinguish process-based learning from the other alternatives. However, the misperception–misremembering hypothesis could not easily account for the slow correct rejections we observed, and it may not account for the large number of false alarms we observed. Misperception or misremembering of the memory set should produce errors on regular trials, and as the data in Table 1 indicate, errors on regular trials were relatively rare. Nevertheless, the hypothesis must be taken seriously, so we constructed what we considered to be a worst case version of it that would predict the largest false-alarm rates on catch trials.

Predictions about false-alarm rates depend entirely on assumptions about the nature of misperception and misremembering. If there were no restrictions on the items subjects substituted for misperceived or misremembered memory-set items (i.e., if any letter in the alphabet could be substituted), false-alarm rates would be very low (because the probability of probing the substituted item would be very low). However, if subjects substituted only items from the positive superset for misperceived or misremembered items, then false-alarm rates could be high enough to take seriously. To be specific, we assumed that all subjects misperceived or misremembered one item from the memory set and substituted another item from the positive superset in its place.[4] On that assumption, the probability of a false-alarm on the catch trial would be .33, .5, and 1.0 for Set Sizes 1, 2, and 3, respectively, because those are the probabilities that the subject would substitute the item that we chose for the catch trial from those remaining in the superset (there are three such items for Set Size 1, two for Set Size 2, and one for Set Size 3). If about half of the subjects misperceived or misremembered (or both) one item in the memory set, this hypothesis could account for our observed false-alarm rates. Moreover, the tendency to false alarm would increase with set size, just as we observed.

---

[3] We thank an anonymous reviewer for pointing out this possibility.

[4] Misperceiving or misremembering one item is a worst case scenario in that it is the only possibility for memory sets of one and three. Only one item can be misconstrued with Set Size 1, and there is only one item remaining in the superset to be substituted for Set Size 3. The worst case scenario for Set Size 2 would be to misconstrue both memory-set items and substitute the two remaining superset items. We considered the one-item case for Set Size 2 to make it more comparable to our analyses of Set Sizes 1 and 3.

Fortunately, the misperception–misremembering hypothesis makes testable predictions for errors on regular (noncatch) trials: If all subjects misperceived or misremembered one item of the current memory set, they should miss (fail to respond *yes*) on positive trials with probability 1.0, 0.5, and .33 for Set Sizes 1, 2, and 3, respectively, because those are the probabilities that the positive probe would be the misperceived or forgotten item. These probabilities would be smaller if some subjects perceived and remembered the memory sets accurately. However, the miss rates should be linked to the catch-trial–false-alarm rates because the probability of misperception or misremembering (or both) should be the same on positive trials as on catch trials (because subjects should not be able to anticipate the nature of the upcoming trial). Regardless of the probability of misperception or misremembering (or both), miss rate averaged across set size should be equal to catch-trial–false-alarm rate averaged across set size. Within set sizes, miss rates should be three times as large as false-alarm rates for Set Size 1, equal to false-alarm rates for Set Size 2, and one third the size of false-alarm rates for Set Size 3.

We tested this hypothesis in two ways. First, the error rates in Table 1 were used to generate the probability of misperception or misremembering the memory sets in the second block. Miss rates were low and unaffected by set size (contrary to the hypothesis), averaging 0.027 for Set Sizes 1–3 in the categorical CM condition and 0.043 for Set Sizes 1–3 in the letter CM condition. These values were significantly lower than the observed false alarm rates on catch trials ($p < .01$). The predicted false-alarm rates were 0.007, 0.03, and 0.09 for Set Sizes 1–3 in the categorical CM condition and 0.013, 0.04, and 0.15 for Set Sizes 1–3 in the letter CM condition. The values for Set Sizes 2 and 3 were significantly smaller than the observed false-alarm rate ($p < .05$).

Second, we calculated the probability of a miss for the last positive trial before the catch trial for Set Sizes 1, 2, and 3 for the categorical and letter CM conditions. Again, the miss rates were relatively low, averaging 0.04 in the categorical CM condition and 0.08 in the letter CM condition, which was significantly lower than the observed false-alarm rates on catch trials ($p < .01$). Moreover, miss rates tended to increase rather than decrease with set size, contrary to prediction. In the categorical CM condition, the miss probabilities were 0.0, 0.0, and 0.13 for Set Sizes 1–3, respectively, leading to predicted false-alarm rates of 0.0, 0.0, 0.375. The values for Set Sizes 2 and 3 were significantly lower than the observed false-alarm rates ($p < .05$). In the letter CM condition, the miss probabilities were 0.0, 0.0, and 0.25, respectively, leading to predicted false-alarm rates of 0.0, 0.0, and 0.75. The value for Set Size 2 was significantly lower than the observed false-alarm rate ($p < .05$).

On the whole, the miss rates suggest that subjects did not misperceive or misremember memory-set items very often. Consequently, we can discriminate process-based learning from the alternative hypotheses and reject it as an interpretation of our results. Note that the alternative (process-switching) hypotheses predict that subjects should respond positively to all items in the positive superset, which would lead to low miss rates on regular trials and high false-alarm rates on catch trials, just as we observed.

Discussion

The training data replicated a standard effect in the memory-search literature: CM conditions produced shallower slopes than VM conditions. These slope differences were apparent in the first block of 384 trials and did not change much in the second block. Further analysis revealed differences in the initial 24 trials in each condition. The differences disappeared only when slopes were calculated for the first exposure to each set size. The fact that CM slopes were different from VM slopes so early in practice is puzzling because, in theory, CM and VM should be equivalent initially and diverge with practice. The differences cannot be attributed to items because letter CM and letter VM groups used the same set of eight letters. They must have been produced by experience, and apparently not much experience was required. This turns out to be consistent with the literature: In many studies that compare CM and VM memory-search tasks, differences in slope were apparent in the very first session and grew with practice (see Ackerman, 1988; Fisk & Schneider, 1983; Kramer & Strayer, 1988; Schneider & Fisk, 1982, 1983). We suspect the differences we observed would become larger with more extensive practice. Thus, some performance improvement was under way, although it had not yet reached asymptote.

The catch trials were intended to shed light on the mechanism underlying performance improvement in CM tasks. The data suggest that probes were not compared with explicit representations of the memory set in short-term memory. Twenty-one of the 48 CM subjects (43%) false alarmed on the catch trial, and those who correctly rejected on the catch trial were slower than they were on their last negative trial and slower than VM subjects. These data are more consistent with process-switching mechanisms than process improvement mechanisms. Process improvement mechanisms cannot explain the elevated false-alarm rate without invoking some other mechanism, such as misperceiving or misremembering the memory set. As we saw, however, that mechanism predicts an elevated miss rate on positive trials, which was not observed. Moreover, process improvement mechanisms cannot explain the slow correct rejections on the catch trial. They predict correct rejections that are as fast as those on regular (noncatch) trials. By contrast, process-switching mechanisms readily explain the data: Subjects switched to a process that responds to the superset rather than the memory set. False alarms occurred because subjects largely ignored the current memory set. Miss rates on positive trials were low because subjects respond readily to superset items. Correct rejections were slow on catch trials because process switching was not yet complete: Some subjects engaged both processes on catch trials. Comparisons between probes and memory sets either ran in parallel with superset-based responses or followed them.

The idea that some subjects engaged both processes on some trials explains why false alarms were less common with smaller memory sets: If there were a race between memory search and the superset-based process (e.g., if memory search raced with item-based learning; Logan, 1988), memory search would win more often the smaller the memory set because comparison time would be shorter overall. If memory search and superset-based processing alternated strategically, subjects might invoke memory search more often with smaller set sizes and superset-based processing more often with larger set sizes because memory search is easy with small set sizes and hard with large set sizes. The idea that some subjects engaged both processes also explains why CM slopes were not zero: Performance was not yet asymptotic. From the perspective of process-switching mechanisms, the processes were still switching; performance was a mixture of preswitch and postswitch strategies.

False-alarm rates were higher in the categorical CM condition than in the letter CM condition. Possibly, categorical CM subjects used the category comparison strategy in addition to item-based learning or the superset strategy used by letter CM subjects. Alternatively, categorical CM subjects may have used item-based learning or the superset strategy just as letter CM subjects did, but learned associations to items or induced the superset more easily than letter CM subjects. Experiment 4 distinguished these alternatives, comparing item-based learning and the superset strategy directly with the category comparison strategy.

Experiment 2

The second experiment was a replication of the VM and CM conditions in Experiment 1, with words rather than letters as stimuli. A number of search studies of automaticity have used words (e.g., Fisk & Schneider, 1983; Schneider & Fisk, 1984), so it is important to generalize our findings from letters and digits to words.

Method

The method was identical to that used in the first experiment, except for the stimuli. As in the first experiment, stimuli were assigned to positive and negative supersets and used for a pair of subjects in CM conditions. However, in this experiment, stimuli were drawn from a pool of 340 five-letter words that were the most frequent five-letter nouns in the Kucera and Francis (1967) norms. The frequencies ranged from 8 to 787 per million with a mean of 75. Each subject saw a different sample of 8 words.

The words were arranged on the screen in much the same way as the letters and digits in Experiment 1. The first and second words were positioned one row above the fixation point so that the last letter of the first word was one column to the left of the fixation point and the first letter of the second word was one column to the right of the fixation point. The third and fourth words were positioned similarly one row below the fixation point. The probe words appeared so that their middle letter occupied the same position as the fixation point.

There were 48 subjects in all, 24 in the CM condition and 24 in the VM condition. Eight subjects in each condition received catch trials at each of the three memory-set sizes (1, 2, and 3). All subjects were from introductory psychology classes at Purdue University and served to fulfill course requirements. All were native speakers of English and had normal or corrected vision.

Results

Training. Mean reaction times and error rates were calculated for each subject in each set size and response-type condition in each of the 384-trial blocks in the experiment. The mean reaction times across subjects are presented in Figure 2.

As with letters and digits, the VM condition produced the characteristic linear increase in reaction time, with set size with roughly equal slopes for positive and negative responses (55 vs. 46 ms per item, respectively). The CM condition produced a flatter slope even in the first practice block (27 ms per item); there was not much change in the slope with practice (26 ms per item in the second practice block).

These conclusions were confirmed in a Group (CM vs. VM) × Set Size × Response Type × Practice Block ANOVA with subjects nested in groups. There were significant main effects of group, $F(1, 46) = 4.51$, $p < .05$, $MS_e = 150{,}357.71$; set size, $F(3, 138) = 307.47$, $p < .01$, $MS_e = 1{,}660.98$; response type, $F(1, 46) = 71.75$, $p < .01$, $MS_e = 5{,}492.10$; and block, $F(1, 46) = 7.41$, $p < .01$, $MS_e = 8{,}009.76$. The interaction between group and set size was significant, $F(3, 138) = 29.25$, $p < .01$, $MS_e = 1{,}660.98$, indicating steeper slopes for the VM subjects. The interaction between set size and practice block approached significance, $F(3, 138) = 2.48$, $p < .07$, $MS_e = 655.47$, but the three-way interaction between group, set size, and practice block was not significant ($F < 1$), indicating no differential slope change in the CM and VM groups. Furthermore, in an ANOVA on the CM data by themselves, the interaction between set size and block was not significant ($F < 1$), indicating no slope reduction in the CM condition.

In addition, there were significant interactions between group and response, $F(1, 46) = 5.80$, $p < .05$, $MS_e = 5{,}492.10$,

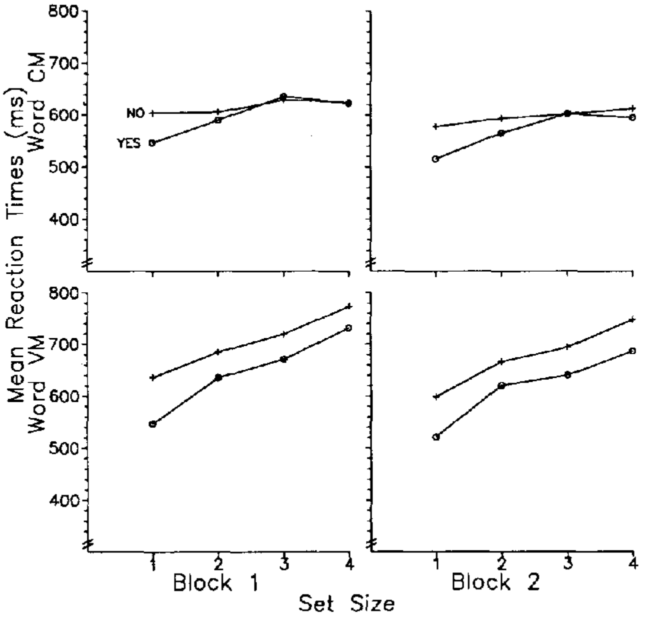

Figure 2. Mean reaction times for each condition of Experiment 2 as a function of set size, response type, and practice block. (CM = consistent mapping; VM = varied mapping.)

and response and set size, $F(3, 138) = 18.79$, $p < .01$, MS_e = 1,252.56, and a marginal interaction between group and block, $F(1, 46) = 2.65$, $p < .12$, MS_e = 8,009.76.

The error rates for each combination of conditions are presented in Table 3. Error rates were low again, and there were no effects that would compromise the interpretation of the reaction time results.

Again, we conducted a finer grained analysis of practice effects to see if CM and VM slopes were equivalent early on. Mean reaction times were calculated for each subject for each set size, collapsing over response type, for the first eight blocks of 24 trials, as in Experiment 1. The slopes were different even in the first 24 trials (33 ms per item in CM; 49 ms per item in VM), though the difference grew larger by the eighth block (Trials 169–192; 25 ms per item in CM; 53 ms per item in VM). Planned comparisons assessing differences in linear trend revealed a significant difference between CM and VM in the first block, $F(1, 966) = 9.56$, $p < .01$, MS_e = 3,254.11, and no other significant differences (i.e., no change in CM or VM slopes from Block 1 to Block 8).

As in Experiment 1, we calculated slopes of linear functions relating reaction time to memory-set size for the first trial of each set size in each condition, averaging across positive and negative trials. The average slopes across subjects were 34 ms and 45 ms per item for CM and VM conditions, respectively. These differences were about as large as those averaged over the first 24 trials.

Catch trials. False-alarm rates from the catch trial and last negative trial in each condition are presented in Table 4. Twelve of the 24 CM subjects false alarmed on the catch trial, though none of them false alarmed on the last negative trial. Only 2 of the 24 VM subjects false alarmed on the catch trial; 4 false alarmed on the last negative trial. For CM subjects, false-alarm rate on the catch trial increased with memory-set size once again, though it was relatively unaffected by set size in the other conditions.

These conclusions were confirmed in binomial probability tests, which compared the observed number of false alarms with the number expected given the number of observations and an expected probability of false alarm taken from the second block of trials. Averaging across set size, the number of false alarms on the catch trial was significantly higher than what would be expected from the false-alarm rate in the second block ($p < .01$) for CM group, but not for the VM group. Analyses within set sizes showed that catch-trial false-alarm rates were elevated significantly for each set size ($p < .01$) in the CM group and for Set Size 3 in the VM group. Analyses based on the last negative trial were not possible for the CM group because there were no false alarms. For the VM group, at Set Size 3, the false-alarm rate on the catch trial was not significantly greater than the one on the last negative trial. Another analysis, using the catch-trial false-alarm rate in the CM group to generate binomial probabilities, showed that for Set Size 3, the CM false-alarm rate was significantly higher than the VM false-alarm rate ($p < .01$).

Table 3
Mean Proportion Error in Training as a Function of Set Size, Response Type, and Practice Block for Each Condition of Experiment 2

	Yes responses				No responses			
Condition	1	2	3	4	1	2	3	4
Word CM								
Block 1	0.03	0.04	0.03	0.03	0.03	0.02	0.02	0.02
Block 2	0.01	0.03	0.03	0.02	0.03	0.02	0.02	0.03
Word VM								
Block 1	0.04	0.05	0.07	0.06	0.05	0.04	0.03	0.05
Block 2	0.03	0.06	0.05	0.06	0.04	0.04	0.03	0.04

Note. CM = consistent mapping; VM = varied mapping.

Table 4
False Alarm Rates (as Proportions) for the Catch Trial and the Last Negative Trial as a Function of Set Size for Each Condition of Experiment 2

	Set size			
Condition	1	2	3	*M*
Word CM				
Catch trial	0.25	0.38	0.88	0.50
Last negative trial	0.00	0.00	0.00	0.00
Word VM				
Catch trial	0.00	0.00	0.25	0.08
Last negative trial	0.25	0.13	0.13	0.17

Note. CM = consistent mapping; VM = varied mapping.

In the CM condition, the increase in catch-trial–false-alarm rate with set size was tested, by using false-alarm rates from Set Sizes 1 and 3 to generate binomial probabilities. Set Size 3 but not Set Size 2 differed significantly ($p < .01$) from Set Size 1, and both Set Sizes 1 and 2 differed significantly ($p < .01$) from Set Size 3.

The tendency to false alarm also appeared in catch-trial reaction times. For CM subjects, correct rejections on the catch trial took longer than correct rejections on the last negative trial (873 ms vs. 618 ms), and CM subjects took longer to correctly reject the catch trial than VM subjects (873 ms vs. 580 ms). These differences were tested with t tests: CM catch trial versus VM catch trial was significant, $t(32) = 2.53$, $p < .05$, MS_e = 114.62. Consistent mapping CM catch trial versus CM last negative trial was not significant, $t(11) = 1.49$, $p = .16$, MS_e = 178.77.

During training, the reaction times of subjects who false alarmed on the catch trial averaged 584 ms; those of subjects who correctly rejected averaged 618 ms. The difference was not significant, $F(1, 22) = 1.03$, MS_e = 110,392.14. The nature of the response on the catch trial did not interact with set size or response type.

Again, we tested the suggestion that subjects may have misperceived or misremembered (or both) memory-set items and compared a misconstrued memory set against the catch trial probe. First, the miss rate for positive trials with Set Sizes 1–3 in the second block averaged 0.02, which was significantly lower than the average false-alarm rate on catch trials ($p < .01$). Moreover, miss rate did not decrease with set size as the misperception–misremembering hypothesis predicted. The miss rates in Table 3 led to predicted false-alarm rates of 0.003, 0.03, and 0.09. The values for Set Sizes 2 and 3 were significantly lower than the corresponding false-alarm rates observed on catch trials ($p < .05$). Second, the miss rate for the last positive trial before the catch trial averaged 0.08, which was significantly lower than the average false-alarm rate on the catch trial ($p < .01$). Miss rates were 0.0, 0.25, and 0.0 for Set Sizes 1–3, respectively. These did not decrease with set size, contrary to the hypothesis. They led to predicted false-alarm rates of 0.0, 0.25, and 0.0, respectively. The predicted values were less than the observed for Set Size 3 ($p < .05$). These data suggest that subjects did not misperceive or misremember memory-set items, and so process-based learning cannot account for performance on catch trials. Note as well that the reviewer's hypothesis cannot account for the slow correct rejections we observed on the catch trials.

Discussion

The major findings of Experiment 1 were replicated in this experiment: In training, CM subjects produced flatter slopes than VM subjects, suggesting that some performance improvement had occurred. Again, however, the difference between CM and VM slopes was apparent early in practice and did not change much as practice progressed. As in Experiment 1, the difference cannot be attributed to items because CM and VM subjects used exactly the same materials (words); it must have been due to experience. The initial difference is puzzling. We note that others have observed it (Ackerman, 1988; Fisk & Schneider, 1983; Kramer & Strayer, 1988; Schneider & Fisk, 1982, 1983), so it may be typical of comparisons between CM and VM.

The catch-trial data replicated Experiment 1 as well. Consistent mapping subjects false alarmed more often and correctly rejected more slowly, which is more consistent with process-switching mechanisms of performance improvement than with process improvement mechanisms. Thus, the major findings generalize beyond letters and digits to words.

The results of the first two experiments are broadly consistent with process-switching mechanisms and inconsistent with process improvement mechanisms. The data, however, do not allow a discrimination among the three process-switching mechanisms. On the one hand, the appearance of slope differences early in practice (i.e., in the first block of 24 trials) seems more consistent with a strategy switch than item-based learning, but there is no principled basis for that interpretation (i.e., strategies may take more than 384 trials to be invoked; item-based learning can take less than 24 trials, see Logan, 1988; Strayer & Kramer, 1990, Experiment 2). On the other hand, the nonzero slopes and the correct rejections on catch trials suggest that memory search went on concurrently with a superset-based process, and that seems more consistent with item-based learning than with the category comparison or superset strategies. Theories already assume that item-based learning can go on concurrently with other processes (e.g., Logan, 1988). However, it may be possible to model concurrent use of strategies, so there is no principled basis for this interpretation either. Which process-switching mechanisms are involved remains an empirical question, to which the remaining experiments are addressed.

Experiment 3

The third experiment manipulated the amount of practice before the catch trial in an attempt to discriminate item-based learning from the category comparison and superset strategies (cf. Hock et al., 1985). Four groups of subjects performed a consistent mapping memory search task with words as stimuli. One group received the catch trial after 96 training trials, another group after 192, another group after 384, and the last group after 768 trials, as in the previous experiments. The main question was whether the tendency to false alarm on the catch trial would increase with the amount of prior practice. Item-based learning should become stronger over practice, and the tendency to false alarm should increase accordingly. By contrast, the category comparison and the superset strategies can be adopted relatively quickly, so the tendency to false alarm may not increase with practice.

These predictions are limited somewhat because we cannot predict with any precision how much practice is necessary to produce a given false-alarm rate in each performance improvement mechanism. Ninety-six trials should be enough to establish category comparison and superset strategies asymptotically, because strategies, by definition, can be developed and adopted very quickly (see, e.g., Logan, 1985a). But 96 trials should not be enough for item-based learning to reach asymptote. Many studies show improvement in performance over several sessions (e.g., Logan, 1978; Strayer & Kramer, 1990). However, there are no hard data on the necessary time course of strategy acquisition versus item-based learning. Our design relied on guesswork rather than deduction from first principles, and the predictions are weakened accordingly.

The third experiment also provided an opportunity to examine the consequences of the catch trial. The previous experiments suggested there was very little difference in performance before the catch trial for subjects who false alarmed versus correctly rejected; this experiment looked for differences after the catch trial. Three of the four groups of subjects experienced four blocks of 96 trials after the catch trial, and we compared performance on the block before and the block after the catch trial.

Subjects who respond strategically (either though the category comparison or superset strategy or through a strategic choice to rely on item-based learning) may abandon their strategy after the catch trial. If they do, the slope of the memory set-size function should increase in the subsequent block. However, subjects may abandon the strategy only if they notice the catch trial. In that case, the subsequent slope should be steeper for subjects who correctly rejected the catch trial because they are likely to have noticed it, but not for subjects who false alarmed and likely did not notice it.

Method

Subjects. There were 64 subjects in all, 16 in each practice condition. Eight subjects in each practice condition received catch trials at each of the two memory-set sizes (one and three). All of the subjects were introductory psychology students at Purdue University who served to fulfill course requirements. All of them reported English as their native language and had normal or corrected vision.

Apparatus and stimuli. The stimuli were words sampled from the pool used in Experiment 2. This time there were four words in the positive superset and 48 in the negative superset. A different set of words was sampled randomly for each subject. The stimuli were displayed on the same equipment and in the same fashion as the stimuli in Experiment 2.

Procedure. The procedure was the same as in the CM conditions in previous experiments, except that catch trials occurred on the 97th, 193rd, 385th, and 769th trials. In each case, the catch trial occurred at the end of a block, so subjects received a short break before proceeding with the task. A different group of 16 subjects received each of these catch trials. Memory-set size was varied on the catch trials once again, but this time only Set Sizes 1 and 3 were used to reduce the number of subjects we required. In previous experiments, the biggest differences were between Set Sizes 1 and 3, so it seemed safe to drop Set Size 2. Also, to reduce the number of subjects required, we dropped the VM conditions. Previous experiments showed no tendency to false alarm in VM conditions, and we expected no change in that tendency over practice. Moreover, the interesting contrasts are between levels of practice in CM conditions, and VM controls were not necessary to evaluate those contrasts.

Subjects were not told about the catch trial until the end of the experiment.

Results

Training. Mean reaction times for the last 96-trial block before the catch trial are presented for each of the four practice groups in Figure 3. They represent average performance on Trials 1–96, 97–192, 289–384, and 673–768. Reaction times decreased with practice, from a mean of 649 ms in the first 96 trials to 507 ms in the last 96. The set-size function was relatively flat in the first 96 trials (14 ms per item) and remained flat throughout practice, increasing only slightly over blocks (18, 18, and 22 ms per item).

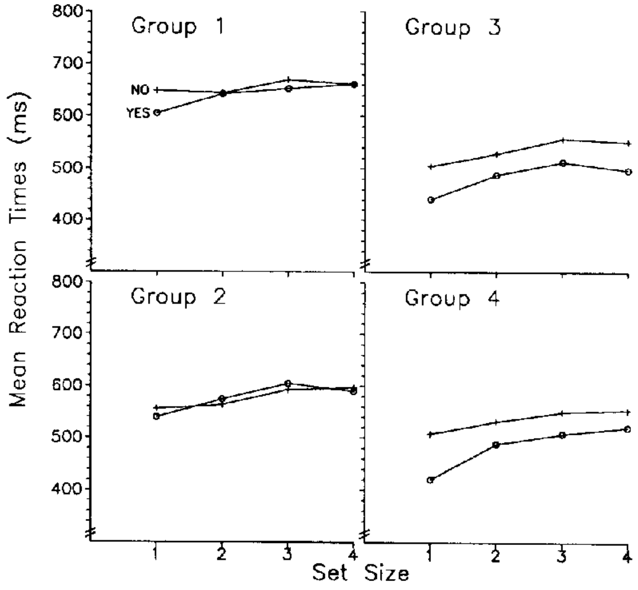

Figure 3. Mean reaction times for each practice group in Experiment 3 as a function of set size and response type (Group 1 represents Trials 1–96; Group 2, Trials 97–192; Group 3, Trials 289–384; and Group 4, Trials 673–768).

Table 5

Mean Proportion Error in Training as a Function of Set Size and Response Type for Each Practice Group in Experiment 3

Practice group	Yes responses				No responses			
	1	2	3	4	1	2	3	4
Group 1	0.07	0.06	0.05	0.06	0.11	0.08	0.02	0.04
Group 2	0.05	0.03	0.05	0.03	0.07	0.01	0.03	0.04
Group 3	0.02	0.04	0.06	0.05	0.06	0.06	0.02	0.07
Group 4	0.05	0.02	0.03	0.06	0.03	0.10	0.06	0.08

Note. Group 1 represents Trials 1–96; Group 2, Trials 97–192; Group 3, Trials 289–384; and Group 4, Trials 673–768.

The reaction times were analyzed in a Practice Group × Catch Trial Group (Set Size 1 vs. Set Size 3 on catch trial) × Set Size × Response Type (positive or negative) ANOVA. There were significant main effects of practice group, $F(3, 56) = 11.32$, $p < .01$, $MS_e = 51{,}904.58$; set size, $F(3, 168) = 28.13$, $p < .01$, $MS_e = 2{,}801.80$; and response type, $F(1, 56) = 25.14$, $p < .01$, $MS_e = 3{,}655.30$. There were no significant interactions involving practice group. The interaction between set size and response type was significant, $F(3, 168) = 3.77$, $p < .05$, $MS_e = 1{,}967.07$, reflecting a steeper slope for *yes* responses than for *no* responses (20 vs. 14 ms per item).

Error rates are presented in Table 5.

Catch trials. The false-alarm rates for the catch trials and the last negative trial that preceded them are presented in Table 6. The main results of the previous experiment were replicated at each level of practice: Subjects were more likely to false alarm on the catch trial than on the last negative trial, and the false-alarm rate on the catch trial increased with memory-set size. Practice itself had little effect on the false-alarm rates; subjects were almost as likely to false alarm on the catch trial after 96 training trials as after 768.

These conclusions were confirmed in binomial probability tests, using the false-alarm rate from the last negative trial and for the block before the catch trial to generate binomial probabilities. The total number of catch-trial false alarms (27) was significantly greater than what would be expected from the false-alarm rate on the last negative trial ($p < .01$) and the last 96-trial block ($p < .01$). The number of false alarms at each set size (9 for Set Size 1; 18 for Set Size 3) was significantly ($p < .01$) greater than expected, as was the number of false alarms at each level of practice (6, 5, 7, and 9, respectively). Analyses within each combination of set sizes and practice levels showed that the number of catch-trial false alarms was significantly ($p < .05$) greater than expectation given the false-alarm rate on the last block for the highest level of practice for Set Size 1 and for all levels of practice for Set Size 3.

The effect of set size on catch-trial–false-alarm rate was significant ($p < .01$), using either false alarm rate to generate binomial probabilities, replicating previous experiments. The effect of practice was assessed by binomial tests and by Spearman's rank order correlation. The binomial tests used either

Table 6

False Alarm Rates (as Proportions) for the Catch Trial and the Last Negative Trial as a Function of Set Size for Each Practice Group of Experiment 3

Condition	Set size 1	3	*M*
Group 1 (Trials 1–96)			
Catch trial	0.25	0.50	0.38
Last negative trial	0.12	0.00	0.06
Group 2 (Trials 97–192)			
Catch trial	0.25	0.38	0.31
Last negative trial	0.00	0.00	0.00
Group 3 (Trials 289–384)			
Catch trial	0.13	0.75	0.44
Last negative trial	0.00	0.00	0.00
Group 4 (Trials 673–768)			
Catch trial	0.50	0.63	0.56
Last negative trial	0.00	0.00	0.00

the first or the last practice block to generate binomial probabilities. The only significant difference to emerge from these analyses was between the last practice group (nine false alarms) and the second (five false alarms). Spearman's rank order correlation between false-alarm rate and practice block was not significant, $r = .80$, $t(2) = 1.89$.

The tendency to false alarm was tested in the catch-trial reaction times. Reaction times for correct rejections on the catch trials and on the last negative trials are presented for each practice group in Table 7. The data were averaged over set size to obtain sufficient stability. Overall, reaction times were slower on the catch trial than on the last negative trial, $t(36) = 2.37$, $p < .05$, $MS_e = 42.19$, replicating previous experiments. The difference was apparent in all but the 96-trial practice group but significant only for the two groups with the most practice: For the 96-trial practice group, $t(9) = 0.03$, $MS_e = 33.33$; for the 192-trial practice group, $t(10) = 0.58$, $MS_e = 108.62$; for the 384-trial practice group, $t(8)$, $= 2.86$, $p < .05$, $MS_e = 54.55$; and for the 768-trial practice group, $t(6) = 2.71$, $p = .05$, $MS_e = 84.13$.

Reaction times from the trial blocks immediately preceding and following the catch trial were analyzed to assess the effects of the catch trial on subsequent performance. In all groups, reaction times were about the same for the block before the catch trial as for the block after (mean difference = 24 ms). Notably, the set-size effect was about the same before and after the catch trial for all groups (mean slope before = 16 ms per item; mean slope after = 18 ms per item). The data were broken down further according to the subject's response on the catch trial. All subjects tended to speed up slightly after the catch trial, whether they false alarmed or correctly rejected on the catch trial (before vs. after was 572 ms and 550 ms for subjects who false alarmed and 582 ms and 564 ms for subjects who correctly rejected). Moreover, the set-size function was not different for subjects who false alarmed (19 ms per item before; 17 ms per item after) than for subjects who correctly rejected (14 ms per item before; 19 ms per item after). Thus, subjects' catch-trial behavior cannot be predicted from their performance before the catch trial, and their catch-trial behavior has no deleterious effect on their performance after the catch trial.

These conclusions were confirmed in a number of ANOVAs performed on the reaction times. One compared performance on the block before the catch trial for subjects who false alarmed as opposed to correctly rejected on the catch trial. Subjects' response on the catch trial was not related to training performance; neither main effect nor interactions with set size and response type was significant. Another compared performance before and after the catch trial for subjects who false alarmed on the catch trial, and found no significant differences. The last one compared performance before and after the catch trial for subjects who correctly rejected, and found significantly faster responding after the catch trial than before, $F(1, 7) = 10.34$, $p < .05$, $MS_e = 3{,}298.43$.

We tested the hypothesis that subjects misperceived or misremembered the memory set by assessing miss rates for positive trials in the block preceding the catch trial and for the last positive trial before the catch trial. The averages over Set Sizes 1 and 3 for the block before the catch trial were 0.06, 0.05, 0.04, and 0.04 for the 96-, 192-, 384-, and 768-trial groups, respectively. The averages for the last positive trial were 0.06, 0.06, 0.0, and 0.0, respectively. Calculated either way, the values were significantly smaller than the observed false-alarm rates ($p < .01$), which suggests that process-based learning cannot account for the tendency to false alarm.

Discussion

This experiment was designed to assess the effect of practice on the tendency to false alarm on the catch trial. The results suggested that if practice had an effect, it was not strong. Overt false-alarm rate increased with practice, but not significantly, and reaction times for correct rejections were elevated more, relative to controls, as practice progressed, but the trend was not significant. One could be impressed with the tendency for false alarms to increase with practice or with the relatively high rate of false alarms after the first 96 trials. The slight practice effect is consistent with item-based learning. The tendency for false-alarm rate to increase with memory-set size on catch trials can be interpreted similarly (i.e., as resulting from a race between item-based learning and memory search; cf. Logan, 1988; Schneider, 1985a). On the other hand, the fact that so little practice was required to induce a strong

Table 7

Reaction Times (in Milliseconds) for Correct Rejections on the Catch Trial and Last Negative Trial in Each Practice Group in Experiment 3

Trial type	Practice group 1	2	3	4	*M*
Catch trial	582	694	582	724	669
Last negative trial	581	631	426	496	569

Note. Practice Group 1 represents Trials 1–96; Group 2, Trials 97–192; Group 3, Trials 289–384; and Group 4, Trials 673–768.

tendency to false alarm suggests that subjects used the superset strategy or some kind of category comparison. Ninety-six trials may not allow much item-based learning.

The experiment replicated the previous ones in showing that reaction times before the catch trial did not distinguish subjects who correctly rejected from those who false alarmed. Analysis of reaction times after the catch trial showed that the catch trial had little impact on subsequent performance: Subjects who correctly rejected were indistinguishable from those who false alarmed.

Experiment 4

This experiment addressed the category comparison strategy, asking whether subjects would manifest any item-based learning when they used the strategy. All subjects performed 768 training trials in the categorical CM condition of Experiment 1, looking for digit targets among letter distractors. There were two kinds of catch trials, a superset catch trial and a category catch trial. On superset catch trials, the probe was a member of the superset that was not a member of the current memory set. On category catch trials, the probe was a member of the same category as the memory set (i.e., it was a digit), but it was not a member of the superset or a member of the current memory set. If item-based learning accrued to members of the superset, then superset catch trials should produce more false alarms than category catch trials. But if there was no item-based learning (i.e., if subjects evaluated only category membership and learned nothing about the identity of the probes), then superset catch trials should produce no more false alarms than category catch trials.

Method

Subjects. The subjects were 32 undergraduate students from University of Illinois who served for course credit or cash, 16 in each condition. All subjects reported English as their native language and had normal or corrected vision.

Apparatus and stimuli. These were essentially the same as in the categorical CM condition of Experiment 1, except that five digits (*2*, *3*, *5*, *6*, and *8*) and four letters (*D*, *F*, *R*, and *T*) were used. For each subject, four of the digits served as the positive superset and the fifth serves as a category catch trial, if it was required. There were five different ways to select the four superset digits from the set of five, and 4 subjects in each group received each assignment.

Procedure. The procedure was the same as in the categorical CM condition of Experiment 1. All subjects received a set size of three on the catch trial. The superset subjects received the remaining member of the positive superset as a probe. The category subjects received the fifth digit, never presented before in the experiment as a probe.

Results

Training. Mean reaction times from the 768 training trials for each group are presented in Figure 4. As in the categorical CM condition of Experiment 1, reaction times were relatively unaffected by set size (mean slope = 10 ms per item). There were no differences between the groups during training. Mean reaction times were about the same (563 ms for superset subjects; 562 ms for category subjects), as were the slopes (9 ms per item for superset subjects; 11 ms per item for category subjects).

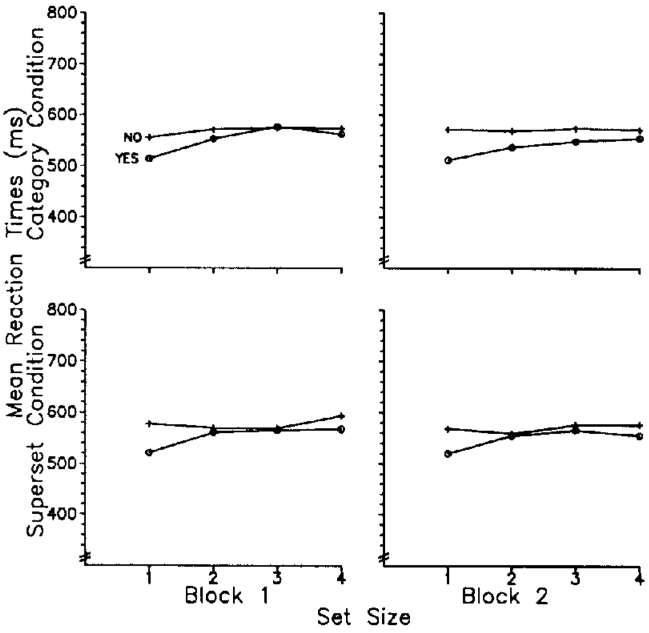

Figure 4. Mean reaction times for each condition of Experiment 4 as a function of set size, response type, and practice block.

These results were confirmed in ANOVA. There were significant main effects of set size, $F(3, 90) = 24.26$, $p < .01$, $MS_e = 846.53$, and response type, $F(1, 30) = 14.16$, $p < .01$, $MS_e = 5{,}009.34$, and a significant interaction between them, $F(3, 90) = 19.42$, $p < .01$, $MS_e = 571.51$, but there were no significant effects involving catch-trial groups, neither main effect nor interactions.

The error data appear in Table 8.

Catch trial. Twelve of the 16 superset subjects false alarmed on the catch trial; 11 of the 16 category subjects false alarmed on the catch trial. The difference between them was not significant. The tendency to false alarm more to a probe that had been presented throughout practice was no stronger than the tendency to false alarm to a completely novel probe that belonged to the same category. The false-alarm rate (72%) was about the same as that observed in the categorical CM condition of Experiment 1 at Set Size 3 (75%).

Mean reaction time was 818 ms for the 4 superset subjects who correctly rejected the catch trial (vs. 565 ms on the last negative trial). It was 838 ms for the 5 category subjects (vs. 637 ms on the last negative trial). Because of the small number of observations, neither of these differences was significant. Collapsing the two groups increased the number of observations to nine, but the difference was still not significant, $t(8) = 1.78$, $MS_e = 125.84$.

The hypothesis that subjects misperceived or misremembered the memory set was tested by calculating miss rates for Set Size 3 for the last block and for the last positive trial before the catch trial. The mean miss rates for the last block were 0.06 and 0.03 for the superset and category groups, respectively. The mean miss rates on the last positive trial

Table 8
Mean Proportion Error in Training as a Function of Set Size and Response Type for Each Catch-Trial Group in Experiment 4

Catch-trial group	Yes responses				No responses			
	1	2	3	4	1	2	3	4
Superset	0.03	0.03	0.06	0.05	0.04	0.03	0.04	0.04
Category	0.03	0.04	0.03	0.04	0.02	0.02	0.02	0.03

before the catch trial were 0.06 and 0.0 for the superset and category groups, respectively. According to the misperceive-misremember hypothesis, false-alarm rates should be about three times as large as these values (i.e., 0.18 at maximum). The observed false-alarm rates were much higher than predicted, allowing us to rule out the hypothesis and to rule out process-based learning as an interpretation of our results.

Discussion

This experiment provided no evidence for item-based learning. Apparently, subjects based their performance on the category comparison strategy, evaluating category membership and paying no attention to prior history with the items. This experiment shows that under some conditions, performance changes in memory search can be primarily strategic, namely, when membership in a well-known category distinguishes positive probes from negative ones. This suggests that other experiments comparing search for digits among letters or vice versa may also invoke the category comparison strategy (cf. Cheng, 1985) and therefore do not involve the kind of automatic processing envisioned by Shiffrin and Schneider (1977). This is important because many experiments, including the ones in the initial Schneider and Shiffrin (1977) paper, use search for letters among digits and vice versa as a way of producing automatic performance. Those experiments require reinterpretation in the light of the present results (i.e., they may provide evidence of category comparison rather than of item-based learning).

Experiment 5

This experiment sought evidence of purely strategic performance in memory search. Subjects were given instructions for a categorical CM task, in which they were told explicitly that all targets would be digits and all distractors would be letters, and then they were given one single trial. That trial was a catch trial. Subjects received a memory set of three digits and then a probe digit that was not a member of the memory set. False alarms on this trial should reflect a pure category comparison strategy because there is no prior training to establish item-based learning or even to allow subjects to induce the positive superset.

Method

Subjects. The subjects were 12 undergraduates from University of Illinois who received course credit for their participation. All were native speakers of English with normal or corrected vision.

Apparatus and stimuli. The apparatus was the same as in the previous experiment. The stimuli were the digits *3*, *5*, *7*, and *9*.

Procedure. Subjects were given standard memory-search instructions, using examples in which memory sets were digits and distractors were letters. Then they were told that all positive probes would be digits and all distractors would be letters. Then they were given one single trial with a three-digit memory set and a digit probe that was not a member of the memory set. Three subjects received each of the four possible ways of assigning three of the four superset digits to the memory set. The catch-trial item was the fourth member of the superset. Temporal and spatial parameters were the same as in Experiment 1. All subjects were told to respond *yes* by pressing the / key with the index finger of their right hand and to respond *no* by pressing the *z* key with the index finger of their left hand.

Results and Discussion

Seven of the 12 subjects (58%) false alarmed on the catch trial. To obtain control false-alarm rates against which to compare this value, we analyzed the first trial for subjects in Experiment 4, who also had digit memory sets and letter distractors. Of the 13 subjects who had negative trials on Trial 1 and who pressed appropriate keys, only 2 false alarmed ($p < .01$ by binomial test); of the 5 subjects who had negative trials with Set Size 3 on Trial 1, none false alarmed.

Because there were no prior training trials on which process-based learning could take place, the false alarms in this experiment represent purely strategic responding. Because there were no training trials to establish which digits were in the positive superset and which were not, these false alarms must represent the category comparison strategy. These results demonstrate that memory-search performance can be based entirely on the category comparison strategy. In principle, category comparison can produce flat slopes without any prior training.

General Discussion

The first 2 experiments demonstrated performance differences that are typical of memory-search tasks: CM slopes were flatter than VM slopes. The catch trials in those experiments and the subsequent ones were intended to distinguish among four mechanisms that might underlie these differences: process-based learning, item-based learning, category comparison, and the superset strategy.

Process Improvement Versus Process Switching

The experiments showed a strong tendency for subjects to false alarm on catch trials to items that were not members of the memory set but were members of the superset of items from which memory sets were sampled. The tendency to false alarm appeared as overt *yes* responses for 95 of the 188 CM subjects (50.5%) and as a tendency to respond *no* slowly for the remaining subjects, as if the catch-trial probe evoked a covert false alarm, which was suppressed or inhibited. These results rule out the possibility that practiced CM performance depends (entirely) on comparing probes with an explicit representation of the memory set in short-term memory (also see Strayer & Kramer, 1990). They rule out process improvement and support process switching as the mechanism underlying the reduction in slope. Process improvement mechanisms, such as process-based learning, predict fast correct rejections on the catch trials, not overt false alarms or slow correct rejections. The three process-switching mechanisms predict overt false alarms and slow correct rejections, as observed.

It is important to note that subjects did not false alarm all of the time. Even though correct rejections were slow, they were nevertheless correct. This suggests that subjects did not abandon the memory-search process over the course of the experiment. On some occasions, at least, they must have compared probes with explicit representations of the memory set in short-term memory, or else they would have false alarmed on every catch trial. However, we suggest that improvements in the memory-search process were not responsible for the flattened slopes in CM conditions. The false alarms suggest some process-switching mechanism used in parallel with the memory-search process, and each process-switching mechanism provides a way for CM slopes to flatten. One need not postulate process improvement to explain the flattened slopes. We suggest that the memory-search process does not change over practice in CM conditions, just as it does not change over practice in VM conditions (cf. Burrows & Murdock, 1969).

Strategy Switching Versus Item-Based Learning

Overall, the experiments seem more consistent with the strategic process-switching mechanisms (category comparison and the superset strategy) than with process switching that involves item-based learning: First, CM slopes flattened very early in practice. Second, the tendency to false alarm was present after only 96 trials of practice when there was no obvious categorical difference between positive and negative items (Experiment 3). Third, the tendency to false alarm was present even on the very first trial when subjects anticipated a categorical difference between positive and negative items (Experiment 5). And fourth, there was no evidence of item-based learning when there was a categorical difference between positive and negative items; subjects false alarmed to a new item from the target category as strongly as to an old item that had been presented many times during practice (Experiment 4).

The main evidence suggestive of item-based learning was the slight but nonsignificant tendency for overt and covert false-alarm rates to increase with practice in Experiment 3. Perhaps a more powerful experiment or a stronger manipulation of practice (i.e., over sessions) would reveal significant differences. However, it would always be possible to interpret practice as providing more opportunity for subjects to adopt a strategy, so more powerful results may remain equivocal. The tendency to false alarm more with larger set sizes is consistent with a race between item-based learning and memory search (cf. Logan, 1988; Schneider, 1985a), but even that could be interpreted as a race between a strategy and memory search. We may not be able to rule out item-based learning entirely, but we can raise serious doubts about its ability to explain practice effects uniquely.

Three Process-Switching Mechanisms or One?

The three types of process switching are hard to discriminate from each other. Item-based learning differs from category comparison depending on assumptions about how categories are represented. Item-based learning differs from the superset strategy primarily in terms of the deliberation involved in responding to the superset: In item-based learning, the superset is represented implicitly (i.e., as separate items) in long-term memory; in the superset strategy, the superset is represented explicitly in short-term memory. And the superset strategy resembles or differs from category comparison depending, for example, on assumptions about whether the superset is a category, how categories are represented, and so forth. These considerations suggest there are large gray areas within which the alternatives cannot be discriminated.

One possibility is that all areas are gray: The alternatives are not really distinct, but rather are separate manifestations of the same underlying learning mechanism. It would seem straightforward to produce all three effects in an instance-learning mechanism (cf. Hintzman, 1986; Medin & Schaffer, 1978) or a connectionist model (cf. McClelland & Rumelhart, 1985). These mechanisms assume item-based learning. They associate responses or response categories with the items presented to them. They would learn to respond to each item in the positive superset and consequently produce false alarms to probes that were in the superset but not in the current memory set. These mechanisms also emphasize distinguishing characteristics and would learn to exploit any categorical differences between the positive and negative supersets. They could produce false alarms to novel probes that share important characteristics with members of the superset.[5]

It is also possible that few areas are gray: The alternatives are distinct under some interpretations, and those interpretations may be closer to the truth than the unitary model just described. There may be redundancy of mechanism. There may be deliberate category comparison and superset strategies in addition to the unitary model. Sometimes subjects may respond to category probes by testing deliberately for category membership and other times by comparing the probes against (long-term) memory representations that happen to possess common characteristics. Strategies may operate early in practice, flattening the slope initially, and item-based learning may take over as practice progresses, flattening the slope for a different reason.

The idea behind the experiments was to find areas of higher contrast among the gray. Learning should increase gradually with practice and so can be ruled out (to some extent) by finding effects with little (Experiment 3) or no (Experiment 5) practice and failing to find stronger effects as practice progresses (Experiment 3). Item-specific learning should accrue more to items experienced in training than to novel items and so can be ruled out by finding no advantage for old over new items (Experiment 4). In light of the data, the best conclusion seems to be that there is some evidence for strategic effects beyond the effects that could be produced by item-based learning mechanisms.

---

[5] Schneider and Fisk (1984) and Schneider and Shiffrin (1985) suggested that subjects could categorize items and develop automatic associations between categories and responses. Although this may be possible in principle, it seems unlikely given recent evidence that category-to-response associations do not contribute much to automaticity (Logan, 1990).

Implications for Memory Search and Automaticity

The results challenge the zero slope criterion of automaticity in memory-search tasks (Logan, 1978), showing that zero slopes, or at least shallower slopes in CM than in VM conditions, can be produced without process improvement or item-based learning, that is, without automatization. They suggest that memory search is not a valid paradigm for investigating Shiffrin and Schneider's (1977) concept of automaticity. Consequently, aspects of Shiffrin and Schneider's theory that were developed to account for memory-search phenomena are suspect and warrant reinterpretation.[6]

These conclusions apply primarily to analyses based on the slopes of the functions relating reaction time to memory-set size: Slope differences are subject to too many alternative interpretations to be informative. There may be item-based learning in addition to strategic shifts in processing. For example, in Experiment 3, mean reaction time decreased substantially over training blocks even though slopes remained stable. It may be possible to separate this item-based learning from the strategic effects by focusing on mean reaction time and ignoring the set-size function. This is essentially what is done by current theories that focus on the power-function speedup in reaction time as the primary phenomenon to be explained (e.g., Logan, 1988; Newell & Rosenbloom, 1981; Schneider, 1985a).

The challenge to the zero slope criterion is most valid when there are obvious categorical differences between targets and distractors, such as letters versus digits. Those conditions produced the strongest tendency to false alarm (Experiment 1) and showed the clearest evidence of purely strategic responding (Experiments 4 and 5). Interestingly, Schneider and Shiffrin (1977) used letter versus digit search in the initial experiments in their article, and subsequently other researchers have used letter versus digit search as a short cut to automaticity (e.g., Hoffman, Nelson, & Houck, 1983). Those experiments require reinterpretation in the light of the present results. The evidence they provide may be more characteristic of strategic processing (i.e., category comparison) than automaticity.

The zero slope criterion is also challenged when there are not obvious categorical differences. Subjects may discover and exploit differences that were not obvious to the experimenter (Cheng, 1985) or they may respond strategically to the superset. The data from Experiment 3 are most relevant here. The tendency to false alarm on catch trials was quite strong after the first 96 trials even though there was no obvious categorical difference between positive and negative items. There was more room for item-based learning in that experiment than in the categorical CM experiments, so it is possible that performance improvement was not entirely strategic. However, the possibility that performance was at least partly strategic cannot be ruled out, so it must be treated as a serious alternative in that experiment and ones like it.

These considerations lead us to recommend that researchers should investigate tasks other than memory search in future studies of automaticity. Memory search may be a popular paradigm in which to study automaticity, but it is certainly not the only one. Problems with the zero slope criterion may be unique to memory search. Important properties of automaticity may be seen more clearly in other tasks (also see Logan, 1985b). Researchers who remain interested in the automatization of memory search should focus on properties other than the slope of the set-size function (e.g., Strayer & Kramer, 1990).

---

[6] Shiffrin and Schneider's (1977) conclusions about visual search may not be compromised by our results. Although Shiffrin and Schneider assumed a great deal of communality between the processes underlying visual search and memory search, we suspect they may be quite different. Some factors that affect memory-search performance may not appear at all in visual search tasks (see, e.g., Stadler & Logan, 1989). Specifically, the category comparison strategy may not be important in visual search (also see Cheng, 1985). However, other problems may plague visual-search studies that letter targets and digit distractors (and vice versa). Duncan (1983) and Krueger (1984) have argued that such tasks produce flat slopes because letters are easy to discriminate from digits (i.e., that flat slopes depend on perceptual differences, independent of learning).

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117,* 288–318.

Burrows, D., & Murdock, B. B., Jr. (1969) Effects of extended practice on high-speed scanning. *Journal of Experimental Psychology, 82,* 231–237.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review, 92,* 414–423.

Corballis, M. C. (1975). Access to memory: An analysis of recognition times. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (pp. 591–612): San Diego, CA: Academic Press.

Duncan, J. (1983). Category effects in visual search: A failure to replicate the "oh-zero" effect. *Perception & Psychophysics, 34,* 221–232.

Ellis, S. H., & Chase, W. G. (1971). Parallel processing in item recognition. *Perception & Psychophysics, 10,* 379–384.

Fisk, A. D., & Schneider, W. (1983). Category and word search: Generalizing search principles to complex processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 177–195.

Gleitman, H., & Jonides, J. (1976). The cost of categorization in visual search: Incomplete processing of targets and field items. *Perception & Psychophysics, 20,* 281–288.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace model. *Psychological Review, 93,* 411–428.

Hock, H. S., Rosenthal, A., & Stenquist, P. (1985). The category effect in visual search: Practice effects on catch trials. *Perception & Psychophysics, 37,* 73–80.

Hoffman, J. E., Nelson, B., & Houck, M. R. (1983). The role of attentional resources in automatic detection. *Cognitive Psychology, 15,* 379–410.

Jones, W. P., & Anderson, J. R. (1982). Semantic categorization and high-speed scanning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 237–242.

Kramer, A. F., & Strayer, D. L. (1988). Assessing the development of automatic processing: An application of dual-task and event-related brain potential methodologies. *Biological Psychology, 26,* 231–267.

Krueger, L. E. (1984). The category effect in visual search depends on physical rather than conceptual differences. *Perception & Psychophysics, 35,* 558–564.

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

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Logan, D. G. (1978) Attention in character classification: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General, 107,* 32–63.

Logan, G. D. (1985a). Executive control of thought and action. *Acta Psychologica, 60,* 193–210.

Logan, G. D. (1985b). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39,* 367–386.

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

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22,* 1–35.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114,* 159–188.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

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

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Ryan, C. (1983). Reassessing the automaticity—control distinction: Item recognition as a paradigm case. *Psychological Review, 90,* 171–178.

Schneider, W. (1985a). Toward a model of attention and the development of automatic processing. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 475–492). Hillsdale, NJ: Erlbaum.

Schneider, W. (1985b). Training high-performance skills. *Human Factors, 27,* 285–300.

Schneider, W., & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21, pp. 53–119). San Diego, CA: Academic Press.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1984). Automatic and control processing and attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 1–27). San Diego, CA: Academic Press.

Schneider, W., & Fisk, A. D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. *Perception & Psychophysics, 31,* 160–168.

Schneider, W., & Fisk, A. D. (1983). Concurrent automatic and controlled visual search: Can processing occur without resource cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 261–278.

Schneider, W., & Fisk, A. D. (1984). Automatic category search and its transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 1–15.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1–66.

Schneider, W., & Shiffrin, R. M. (1985). Categorization (restructuring) and automatization: Two separable factors. *Psychological Review, 92,* 424–428.

Shiffrin, R. M., & Czerwinski, M. P. (1988). A model of automatic attention attraction when mapping is partially consistent. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 562–569.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology, 93,* 72–83.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Shiffrin, R. M., & Schneider, W. (1984). Automatic and controlled processing revisited. *Psychological Review, 91,* 269–276.

Stadler, M. A., & Logan, G. D. (1989). Is there a search in fixed-set memory search? *Memory & Cognition, 17,* 723–728.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27,* 1–32.

Strayer, D. L., & Kramer, A. F. (1990). An analysis of memory-based theories of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 291–304.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes.* Cambridge, England: Cambridge University Press.

Received July 10, 1990
Revision received October 1, 1990
Accepted October 5, 1990 ■