$0567$

**S**DTIC
**E**LECTE
OCT 3 1995
**C** **D**

# ERRORS IN SKILLED PERFORMANCE

Dan J. Woltz & Michael K. Gardner

University of Utah

August, 1995

Final Report

Grant F49620-93-0094

**19951002 021**

# REPORT DOCUMENTATION PAGE

| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 31 Aug 95 | Final Report  1 Jan 93 - 30 June 95 |

**4. TITLE AND SUBTITLE**

Errors in Skilled Performance

**5. FUNDING NUMBERS**

F49620-93-1-0094
61102F
23B/BS

**6. AUTHOR(S)**

Dan J. Woltz & Michael K. Gardner

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

University of Utah
Salt Lake City, UT  84112

**8. PERFORMING ORGANIZATION REPORT NUMBER**

Technical Report 2

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

AFOSR/NL
Bolling AFB, DC  20332-6448

Dr John F. Tangney

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT (Maximum 200 words)**

First, we discuss Anderson's (1983) ACT* theory as the basis of our work on skilled performance errors. Second, we outline conditions we believe promote errors -- *long-term priming* (training on only a subset of possible problem solution types), *short-term priming* (presenting multiple surface structure instantiations of a single, deep structure problem type in succession), and *working memory load* (presenting a concurrent secondary task requiring working memory capacity). Third, we describe our methodology for "detecting" undetected errors. Fourth, we present our empirical work. Twelve studies are presented on long-term priming. These found general support for the existence of two memory mechanisms, composition and proceduralization, and their respective roles in skilled performance errors. Five studies are presented on short-term priming. These found no support for short-term priming as a process underlying errors, despite its popularity among theorists. One study is presented on working memory which found an increase in latency, but not error rate, due to load (a surprising finding). Finally, two studies investigated individual differences variables related to undetected errors. Self-report questionnaires of error proneness did not correlate with performance errors, but working memory capacity, as measured in performance tests did. Directions for future research are discussed.

**14. SUBJECT TERMS**

Cognition, human error, skilled performance, human memory

**15. NUMBER OF PAGES**
37

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| unclassified | unclassified | unclassified | UL |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std 239-18

5 SEP 1995

Errors in Skilled Performance
Final Technical Report
for the Period 1 January 1993 through 30 June 1995
for Grant #F49620-93-1-0094 (Cognition Program)
Dan J. Woltz and Michael K. Gardner
University of Utah

Submitted to:
US Air Force Office of Scientific Research
Air Force Systems Command, USAF
Bollling AFB, DC 20332-6448

## Summary

First, we discuss Anderson's (1983) ACT* theory as the basis of our work on skilled performance errors. Second, we outline conditions we believe promote errors -- *long-term priming* (training on only a subset of possible problem solution types), *short-term priming* (presenting multiple surface structure instantiations of a single, deep structure problem type in succession), and *working memory load* (presenting a concurrent secondary task requiring working memory capacity). Third, we describe our methodology for "detecting" undetected errors. Fourth, we present our empirical work. Twelve studies are presented on long-term priming. These found general support for the existence of two memory mechanisms, composition and proceduralization, and their respective roles in skilled performance errors. Five studies are presented on short-term priming. These found no support for short-term priming as a process underlying errors, despite its popularity among theorists. One study is presented on working memory which found an increase in latency, but not error rate, due to load (a surprising finding). Finally, two studies investigated individual differences variables related to undetected errors. Self-report questionnaires of error proneness did not correlate with performance errors, but working memory capacity, as measured in performance tests did. Directions for future research are discussed.

## Research Objectives

A general finding in the literature is that higher levels of skill in cognitive tasks result in faster and more error free performance (e.g., Bryan & Harter, 1899; Crossman, 1959; Fitts & Posner, 1967; LaBerge, 1973; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). While this trend is true in general, recent theoretical advances in cognitive science (e.g., Anderson, 1983, 1987) lead to the prediction that experts (i.e., highly skilled performers) will be more error prone under certain training/transfer circumstances than novices. Furthermore, these errors should be unavailable to conscious introspection -- that is, they should be undetected by the performer. A series of experiments were performed to test these predictions.

## Theoretical Background

The predictions concerning the relative quantity of errors by skilled performers, and their ability to detect these errors, derive from the ACT* theory of skill acquisition proposed by John Anderson (1983, Singley & Anderson, 1989). In Anderson's theory, cognitive skills are represented as a set of *productions*, or condition-action statements, along with a hierarchical *goal structure* that facilitates problem solving. Skill acquisition starts with the so-called weak problem solving methods (e.g., analogy, means-ends analysis, etc.) that are applicable across a wide range of problem types. The weak methods take declarative knowledge stated in the problem instructions, and organize the initial productions necessary to solve the problem. At first these productions are both numerous and general. Two processes restructure these original productions into a smaller number of more task specific ones. The first process is called *proceduralization*. This process removes variables from the production definitions, and embeds task specific constants in their place. Proceduralized productions are hypothesized to be more efficient, but less flexible, than their generalized counterparts. The second process is called *composition*. Composition takes

two or more productions which fire in sequence and collapses them into a single, more complex production. Again, this process leads to greater efficiency by reducing the number of productions necessary to represent task performance, but at the cost of flexibility. There may be other costs as well, such as the number of conditions that must be active in working memory at one time in order to fire the production.

Productions are assumed to be learned in an all-or-none fashion. However, productions have a "strength" associated with them. Each time a production fires, its strength is enhanced, making it more likely to fire again in the future.

<div align="center">Circumstances That Promote Errors</div>

Certain task circumstances promote error making. In one paradigm, which we have designated *long-term priming*, subjects learn a set of rule sequences for performing a task requiring the serial application of separate cognitive processing steps. The set of rule sequences learned during training, however, constitutes only a subset of the total universe of rule sequences. During transfer subjects are exposed to the entire universe of rule sequences (or, occasionally, just a larger subset of the universe). We believed that subjects will be more likely to make errors on those rule sequences that are new to the transfer session, and that they will be relatively unaware of these errors. Furthermore, this tendency will be greater among more skilled performers (i.e., performers with greater degrees of expertise in the task). This prediction derives from two factors: (1.) expert subjects will have more highly composed productions than novice subjects, thus reducing their ability to introspect on how they solve the task (therefore, the errors are undetected); and (2.) expert subjects will have composed productions for rule sequences learned during training with relatively high strength, and, therefore relatively low firing thresholds. This should lead to occasions when these *strong-but-wrong* productions fire inappropriately to new rule sequences that achieve a partial match to the old productions' conditions. Although the match is only partial, it is compensated for by the extremely low firing threshold of these highly-practiced productions.

It is also possible that seeing the exact same instances of a rule-sequence would result in a benefit to processing latency and accuracy. Such *instance effects* would be predicted as part of the process of proceduralization (described above). To calculate instance effects, one must compare latency and/or error rate during transfer for old items (that is, items with a previously practiced rule sequence and previously seen surface structure) to new instances of old items (that is, items with a previously practiced rule sequence, but with a different surface structure).

Another paradigm, which we have designated *short-term priming*, is similar to long-term priming in its rationale; however, the strengthening of productions is hypothesized to occur over a brief time frame: say several trials within a trail block. If either the repeated firing of a production temporarily strengthens it, or if the presentation of contextual information associated with the firing of a production spreads activation that temporarily lowers the firing threshold of the production, then temporary priming should be possible. Again, we would expect errors in situations where a temporarily primed production shared a partial match to the conditions of an item. Furthermore, this effect should be more pronounced in experts, and should be more likely to be undetected by experts.

A final task circumstance that should promote errors is the addition of a *working memory load*. A working memory load should promote errors for several reasons. First, for novices, the memory load should compete for working memory capacity with the productions necessary to complete the experimental task. To the extent that there is insufficient working memory capacity to complete both tasks, one or both tasks should suffer. Second, for experts, composed productions require more information to be held in working memory to satisfy their conditions. For example, if production A requires condition 1, and production B requires condition 2, the composition of these productions, production A-B, requires conditions 1 and 2 to be present for its firing. This additional working memory load is offset, to some degree, by requiring fewer productions to complete a task, and by simplifying the task's goal structure. Thus it is difficult to make specific predictions concerning experts' performance.

## Detecting Undetected Errors

A methodological problem faced in this research program was how to decide if a particular error made by a subject was a "detected" error (an error in the subject's awareness which could be corrected) or an "undetected" error (an error outside of the subject's awareness). To make the distinction, we designed our tasks in the following way: during training, subjects were instructed to perform as quickly as possible, while maintaining an accuracy rate of 90%. If a response on a trial was incorrect, the word WRONG and a low tone was presented for 2 s. At the end of blocks, subjects were given feedback on median latency and percentage of errors made during that block of trials (and all previous blocks during that day's session). Subjects were thus encouraged to decrease their latency until this pushed their error rate above 10%. During transfer, subjects were given new instructions. They were told to attempt to attain an accuracy rate of 100%. In order to achieve this difficult new goal, subjects were told they could "retake" any trial they thought they had made an error on by pressing the spacebar on the computer keyboard. By retaking the trial, only data from the "corrected" trial would count toward their performance goal. In fact, we collected data from all trials, and were consequently able to distinguish undetected from detected errors. Indeed, four possibilities exist: correct trials thought to be correct by the subject, correct trials thought to be an error by the subject, incorrect trials thought to be incorrect by the subject (i.e., detected errors), incorrect trials thought to be correct by the subject (i.e., undetected errors).

The methodology was tested in a pilot study. The study included 40 subjects. The subjects learned a task we have designated *number reduction*. On each trial, a 4-digit number was presented that had to be reduced to a single digit. This reduction was accomplished by applying some combination of the following component rules. The *same rule* stated that two identical numbers could be reduced to a single digit of that same number (e.g., 77=7). The *midpoint rule* stated that two numbers that differed by two could be reduced to their midpoint (e.g., 53=4). The *contiguous rule* stated that two numbers in either an ascending or descending sequence could be reduced to the next number in the sequence (e.g., 32=1; 67=8). Finally, the *last rule* stated that two number whose difference was greater than 2 could be reduced to the last of the two numbers (e.g., 28=8; 63=3). These rules were applied to multi-digit stimuli by parsing the stimuli pairwise left to right and carrying forward intermediate solutions to be combined with the next digit in the stimulus (e.g., 9687=7).

Subjects were also administered a brief questionnaire that asked the following questions: (1.) Did you use the spacebar to correct errors today? (2.) Did the spacebar ever fail to operate properly at any time during the session? and (3.) Were you ever aware of an error, but decided not to correct it?

Results from the pilot study yielded the following conclusions and modifications. Subjects overwhelmingly said they used the spacebar for error correction and that it operated properly. Only occasionally did a subject indicate that they were aware of an error, but decided not to correct it. When this did occur, it was often because the awareness of an error did not occur until the following trial, at which time it was no longer possible to correct the error. The error detection software was modified to allow correction of an error not only during the interstimulus interval, but also during the following trial (until a numeric response was given). This allowed subjects to correct the vast majority of detected errors.

## Long-Term Priming

Several studies investigated the effects of long-term priming on error making, error awareness, and latency. The first two studies used a simplified version of the number reduction task. In this paradigm, subjects are given two rules for reducing multi-digit stimuli containing combinations of the numbers 1, 2, and 3. The *same rule* states that if two adjacent digits are the same, they may be reduced to a single digit of that value (e.g., 11=1). The *different rule* states that if two adjacent digits are different, they can be reduced to the unused number (e.g., 13=2, 23=1). The rules are applied stepwise from left to right, carrying forward intermediate results. Stimuli consisted of three digit triplets which yielded a single digit solution.

*Study I Objectives.* Study I was designed to investigate the roles of memory for general processing sequences and memory for specific instances on latency and error rate in the simplified number reduction task.

*Study I Methods.* In study I, subjects consisted of 479 US Air Force recruits in their eleventh day of basic training at Lackland AFB, San Antonio, TX. Of these, 77 were eliminated from the final data analysis, because their data indicated lack of effort. This left a final pool of 402 subjects between the ages of 17 and 27.

The training phase consisted of eight blocks of 24 trials. Each subjects performed 12 unique instances during training. However, subjects differed with regard to whether those 12 instances came from *four* different rule sequences, or whether they came from only *two* rule sequences.

The transfer phase also consisted of eight blocks of 24 trials. However, during transfer, all subjects were exposed to all 24 possible stimuli (each of the four possible rule sequences had six separate exemplars).

*Study I Results.* Training data showed no difference between the two and four rule conditions in terms of error rate. However, the conditions did differ with respect to latency, with the two rule condition being significantly faster than the four rule condition. This result reflects the importance of memory for general processing sequences, since both groups received equal amounts of practice with each component rule, and equal numbers of unique exemplars.

In transfer, subjects saw both old trials (previously presented instances) and new trials (never before seen instances). For the four rule sequence group, all new trials represented previous used rule sequences. For the two rule sequence group, however, new trials represented both new instances and new rule sequences. Comparing new and old trials for the four sequence group gives an estimate of the instance effect, or the effects of proceduralization. Differences were relatively small (on the order of 50 ms in latency and 1% in error rate), but statistically significant. Comparing new and old trials for the two sequence group gives an estimate of the of the effects of memory for general processing sequences, or composition with variable arguments. This effect was sizable (on the order of 300 ms in latency and 5% in error rate) and also statistically significant. Thus memory for general processing sequences is an important component in learning a task such as number reduction, and is more powerful an influence than memory for individual instances. This is in accord with predictions made by some theories of skilled performance (e.g., Anderson, 1983; and MacKay, 1987), but in contrast to predictions made by some instance-based theories (e.g., Logan, 1988).

*Study II Objectives.* Study II introduced differing levels of expertise or practice to the design of Study I. This allowed us to study the time course of the development proceduralization and composition. Some evidence (e.g., Anderson, 1983; Frensch, 1991) exists to suggest that such effects develop quite quickly, although it would seem reasonable to propose that these changes in performance are gradual. A second objective was to investigate negative transfer in the simplified number reduction task. Subjects exposed to new rule sequences during transfer actually performed worse on these than they performed on new sequences at the start of training (20% errors versus 10% errors). This suggests one source of error making in skilled cognitive performance: the carrying over of previously learned general rule sequence memory to new situations in which these sequence are inappropriate.

*Study II Methods.* Methods were identical to those of Study I, except for a between subjects manipulation of amount of training. Subjects received either 1, 2, 4, or 8 blocks of practice during training. Subjects were 796 US Air Force recruits in their eleventh day of basic training at Lackland AFB, San Antonio, TX. Of these, 129 were eliminated from the final data analysis, because their data indicated lack of effort. This left 667 subjects for the final data analysis.

*Study II Results.* Latency data over training blocks were extremely similar for the four skill level groups (i.e., 1, 2, 4, or 8 blocks of trials during training). On the one block common to all groups, there was no significant difference among the groups. Transfer performance was examined both in terms of latencies and

error rates. Latency data revealed two things: (1.) skill level affected transfer latency (the highest skill group was approximately 200 ms faster than the lowest skill group), and (2.) general sequence memory effects were a function of skill level (greater degrees of practice lead to larger differences between old and new sequence trials). Error data also showed greater general sequence memory effects with higher levels of skill; that is, high skill levels lead to larger differences in error rates between old and new trials, with subjects being more error prone on the new trials. Finally, there was evidence for the growth of negative transfer with skill. Subjects with more practice during training made more errors on new trials (in an absolute sense) than subjects with less practice.

Thus Study II provided evidence for the gradual growth of composition with skill, and for the development of negative transfer in close transfer situations along with skill that facilitates correct performance on old trials. One question still remained: would the effects observed in the simplified number reduction task generalize to more complex tasks, such as the full fledged number reduction task?

*Study III Objectives.* The objective of study III was to replicate the findings of study I within the more complex task environment of full fledged number reduction (described above).

*Study III Methods.* In this version of number reduction, there were 24 possible rule sequences with each sequence being three rules long. A within subjects design was used in which each subject studied a subset of 12 of the 24 rule sequences during training. Subjects received 30 training blocks of 24 trials each, over a three day period. During the third day they also performed 10 transfer blocks which contained three separate trial types: old/old trials (previously solved rule sequences with previously viewed instances), old/new sequences (previously solved rule sequences with new instances), and new/new trials (new rule sequences with new instances). Subjects were forty undergraduate students at the University of Utah. During training, sequences were selected in such a way that each subject received equal amounts of practice with each individual rule in each serial position in the three sequence chain of rules. During training blocks, subjects were encouraged to go as fast as possible, while maintaining a 90% accuracy rate. During transfer, subjects were told to go as fast as possible, while maintaining 100% accuracy. Subjects were also instructed that they could retake any transfer trial they thought they might have responded to incorrectly by pressing the spacebar on the computer keyboard before responding to the following trial. The data on undetected errors is not reported here, as this part of Study III served as the pilot for developing the methodology for "detecting undetected error" (see above). Finally, subjects were shown a list of the 24 possible rule sequences (e.g., LAST-MIDPOINT-SAME) and asked to choose the 12 they had been exposed to during training.

*Study III Results.* Latency data from transfer established a reliable difference between old/new and new/new trials, indicative of a general sequence memory effect and composition. Latency data from transfer also revealed a non-significant trend (p=.10) toward a difference between old/old and old/new trials, likely indicative of a weak instance effect and proceduralization. Error data mirrored the latency findings. The difference between old/new and new/new trials was significant (i.e., general sequence memory and composition), and the difference between old/old and old/new trials approached significance (p=.08; i.e., instance effect and proceduralization). Thus, in the current task environment there is strong evidence for general sequence memory and composition, and some evidence for a weaker effect of instance memory and proceduralization.

Finally, subjects' performance on identifying the rule sequences seen during training was at chance levels. This is in accord with their own anecdotal reports. Subjects seemed to have no conscious access to their memory for general rule sequences, as would be predicted by the process of composition.

Studies I, II, and III have been published in the *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Woltz, D.J., Bell, B.G., Kyllonen, P.C., & Gardner, M.K. [in press]. Memory for order of operations in the acquisition and transfer of sequential cognitive skill.).

*Study IV and V Objectives.* The next pair of studies investigated in more detail subjects' ability to consciously access their general sequence memory and their instance memory. Study IV investigated

general sequence memory, while Study V investigated instance memory. We hypothesized that subjects would have no conscious access to general sequence memory, because this memory is contained in composed productions in procedural memory. Once a production is composed, the contents of it are no longer open to conscious introspection (Anderson, 1983). Likewise, we predicted that subjects would have no conscious access to their instance memory, because it also is part of procedural memory.

*Study IV Methods*. Methods for Study IV were similar to those for Study III. Subjects received 45 blocks of training (each block being 24 trials long) over four sessions. Training trials consisted of 12 of the 24 possible rule sequences used in full fledged number reduction. Training rule sequences were selected so that each rule appeared in each serial position with equal frequency. Subjects also received 12 blocks of transfer trials during the fourth session. On these trials subjects were instructed to respond "old" if a number sequence represented a sequence of rule operations they had used during training, and "new" if not. A further manipulation concerned solving the items. On half the transfer blocks, subjects solved the items prior to making "old/new" judgments. On the other half, subjects simply made "old/new" judgments without solving the items. Blocks that required solving the items were alternated with blocks that did not require solving the items. The type of block that started the transfer session was counterbalanced across subjects.

Twenty-six University of Utah students participated in Study IV.

*Study IV Results*. During transfer, recognition performance on the "old/new" judgment was close to chance, but significantly better. Subjects were correct 52.53% of the time when performing items ($t[25] = 2.48$, $p < .05$), and correct 53.04% of the time when not performing the items ($t[25] = 3.53$, $p < .01$). These findings indicate a small, but significant, recognition of processing sequences.

The data were also analyzed using signal detection theory. For each condition, $d'$ was calculated. A $d'$ significantly different from zero would also indicate that some recognition memory for old sequences of processing operations existed. For the old/new judgment only condition, $d' = 0.24$, $t(25) = 2.69$, $p < .05$; for the performance and judgment condition, $d' = 0.36$, $t(25) = 4.19$, $p < .001$. Thus, signal detection analysis also indicated slight recognition memory for processing sequences.

The important question for Study IV was whether the observed processing sequence effects of earlier studies depended on recognition of the processing sequences or not. Latency data from Study IV were able to help decide this issue. At least three possibilities existed (these are presented in the figure below): (1) Processing sequence memory facilitated performance, but did not require explicit recognition of the processing sequences. In this case, old status of the sequences facilitates performance, regardless of recognition. (2) Conscious recognition of processing sequences facilitated performance. Under this possibility, correct recognition of old sequences (i.e., calling them old) facilitates performance, but their is no recognition effect for new trials. (3) Performance fluency on trials could be the basis for labeling them old or new. In this case, memory is not facilitating performance at all; rather, performance is dictating what memory labels are given -- old to quick trials and new to slow trials.

# Possible Latency Outcomes of Study IV

## Possibility #1



## Possibility #3



## Possibility #2
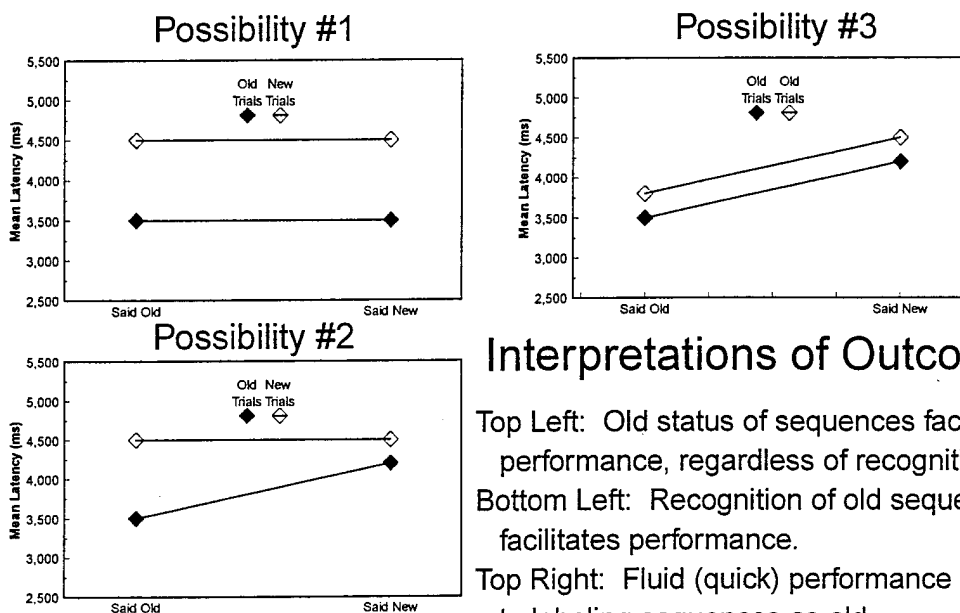


## Interpretations of Outcomes

Top Left: Old status of sequences facilitates performance, regardless of recognition.

Bottom Left: Recognition of old sequences facilitates performance.

Top Right: Fluid (quick) performance leads to labeling sequences as old.

Latency results from Study IV are presented in the figure below. They most closely match possibility three: performance fluency appeared to be the basis for calling trials old or new. Statistical analysis (repeated measures ANOVA) revealed: (1) a significant main effect of saying old versus new ($F[1,23] = 10.04$, $p < .01$), with items labeled as old being faster than those labeled new; (2) a significant main effect for old versus new processing sequence ($F[1,23] = 7.58$, $p < .05$), with old processing sequences being faster than new; and (3) a marginally significant interaction between labeling and sequence status ($F[1,23] = 4.60$, $p = .043$), with the improvement due to an old sequence being greater for trials labeled new than for trials labeled old. This interaction argues against possibility two -- that recognition of a processing sequence as old is instrumental to its performance effects. If such a possibility were tenable, the old/new difference should be greater for items labeled as old, which is exactly the opposite of the obtained results.

## Study IV Latency Results
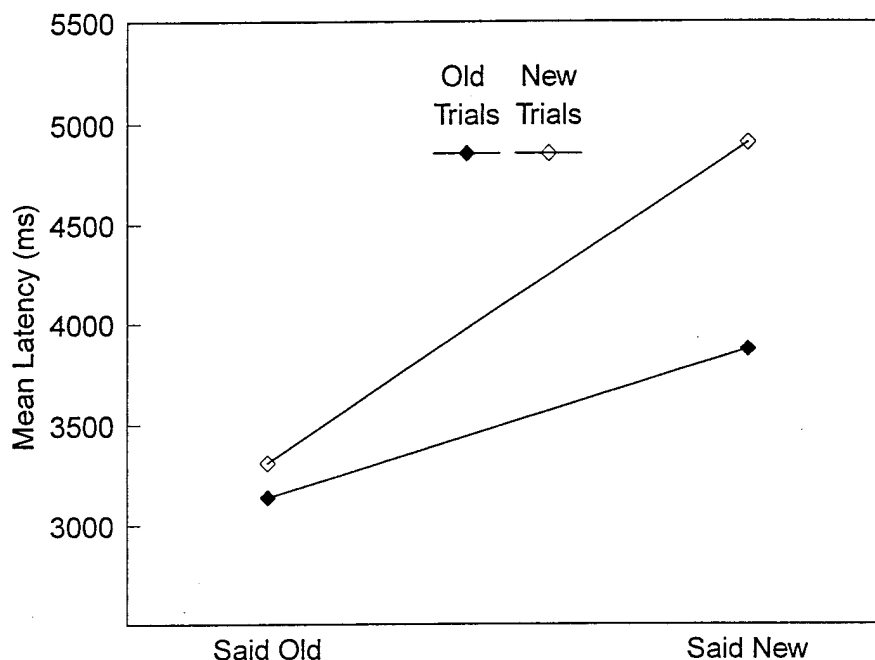


Study IV's results confirmed those of Study III: general rule sequence memory (i.e., composition) appears to primarily implicit in nature. Although some evidence for recognition exists, the size of these effects are small. Since $d'$ can be interpreted as an effect size, recognition effects appear to be on the order of a quarter to a third of a standard deviation in size. Rather than recognition of processing sequences being the basis for performance effects noted earlier, it appears that performance is the primary basis of deciding whether an item is old or new.

*Study V Methods.* Methods for Study V were similar to those for Study IV, except that the "new/old" judgment was based on whether the item was actually presented during training (i.e., was an old instance).

Thirty-one University of Utah students served as subjects in Study V.

*Study V Results.* During transfer, recognition performance on the "old/new" instance judgment was close to chance, but significantly better, just as it had been for rule sequences in Study IV. Subjects were correct 54.19% of the time when performing items ($t[30] = 5.70, p < .001$), and correct 52.23% of the time when not performing the items ($t[30] = 2.60, p < .05$). Thus, subjects demonstrated a weak but reliable ability to recognize old instances during transfer.

The data were analyzed using signal detection theory, as in Study IV. For each condition, $d'$ was calculated. For the old/new judgment only condition, $d' = 0.12, t(30) = 2.42, p < .05$; for the performance and judgment condition, $d' = 0.27, t(30) = 5.89, p < .001$. Once again there was some evidence of recognition of instances, although this effect was small.

The same three possibilities existed for the latency data from Study V as in Study IV. And just as in Study IV, the latency data from Study V (presented below) supported possibility three: that performance fluency influenced old/new judgments, but that correct recognition of old instances was not implicated in producing the weak instance effects found in previous studies. Statistical analysis (repeated measures ANOVA) confirmed this conclusion: (1) there was a significant main effect of saying old versus saying new ($F[1,29]$ = 5.81, $p < .05$); (2) there was a marginal main effect old versus new instances (i.e., a weak instance effect: $F[1,29] = 3.45, .10 > p > .05$); and (3) there was no interaction between labeling and instance status ($F[1,29] < 1, p > .05$).

## Study V Latency Results
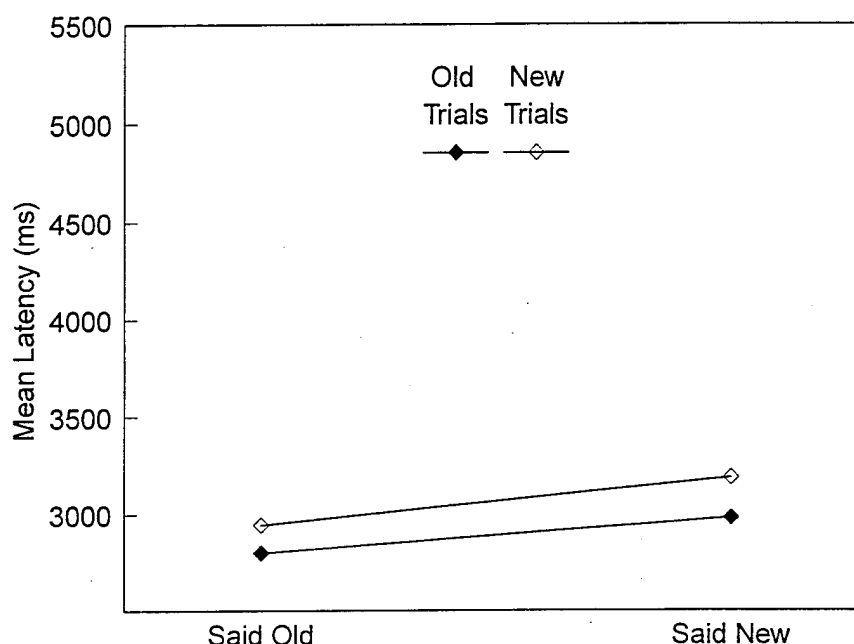


Study V's results were similar to those of Study IV: instance effects (i.e., proceduralization) appears to primarily implicit in nature. Although some evidence for recognition exists, the size of these effects are small: on the order of a quarter of a standard deviation in size or less. Rather than recognition of instances being the basis for instances effects noted in earlier studies, it appears that performance plays a role in deciding whether an item is an old or new instance.

Studies IV and V are currently being written up for submission to *Memory*.

*Study VI and VII Objectives.* Studies VI and VII were designed to generalize the findings of Studies IV and V to a new task domain, and to further investigate the distinction between declarative knowledge, which is available at the beginning of skill learning, and procedural knowledge, which dominates at later stages of skill learning. The experimental task used in Studies VI and VII was called *procedural learning*, because it involved learning a procedure (i.e., a set of rules) for classifying numbers presented on the computer screen and giving a binary response (either "L" for like or "D" for different). The procedure is taught to subjects as a hierarchical classification scheme:

When a word is presented, first determine if it is A WORD (e.g., one, two, three) or A DIGIT (e.g., 1, 2, 3).
>    If the number is A WORD (e.g., one, two, three), then determine if it is ODD or EVEN.
>    >    EVEN belongs on the top of the computer screen and ODD belongs on the bottom.
>    >    >    If it is EVEN on TOP or ODD on BOTTOM, press "L" (like), otherwise press "D" (different.
>    If the number is A DIGIT (e.g., 1, 2, 3), then determine if it is SMALL (1-9) or BIG (11-19).
>    >    SMALL belongs on the TOP of the computer screen and BIG belongs on the bottom.
>    >    >    If it is SMALL on TOP or BIG on BOTTOM, press "L" (like), otherwise press "D" (different).

Study VI investigated whether groups with greater and lesser degrees of skill learning (i.e., practice) differed with regard to their ability to access the original set of classification rules on a declarative knowledge test. Study VII explored whether access to declarative knowledge was affected by the form of the declarative knowledge test -- that is, whether knowledge was tested in manner consistent or inconsistent with its initial form at study.

*Study VI Methods.* Thirty-eight University of Utah students served as subjects in Study VI. These were divided into two groups: high skill (N = 21) and low skill (N = 17). High skill subjects received declarative knowledge training and 92 blocks (5,888 trials) of practice on the procedural learning task. Declarative knowledge training occurred during session one, while practice occurred over all five sessions. Sessions were spaced one week apart. Low skill subjects participated in an unrelated computer task for the first four sessions. During their fifth sessions, they received declarative knowledge training and 4 blocks (256 trials) of practice on the procedural learning task. At the end of session five, both groups received a true/false declarative knowledge test consisting of 100 items. The items asked questions in a fashion similar to the original training on the rules, e.g., "If a number is a digit and is small it belongs on the top?"

*Study VI Results.* The dependent variables of interest were latency and error rate on the declarative knowledge test. Low skill subjects were faster than high skill subjects on the declarative knowledge test, but high skill subjects were more accurate. Thus, both groups retained somewhat similar access to their declarative knowledge. The groups did, however, differ in their procedural knowledge. At the end of training, the high skill group had latencies approximately 400 ms faster than the low skill group.

*Study VII Methods.* Study VII extended Study VI by examining declarative knowledge in a different way. Subjects' declarative knowledge was tested by presented them with a constellation of stimulus attributes, e.g., "word-big-odd-top", to which subjects had to respond with an "L" or a "D". Procedural knowledge should not have been directly transferable to the new test situation. However, subjects could solve the new items by recompiling their original declarative knowledge. The question was whether latency and error rate would interact with skill level (high versus low) in this new test environment.

Thirty-eight University of Utah students participated in Study VII. As in Study VI, there was a high skill group (N = 19) and a low skill group (N = 19). Procedures were the same, except for the declarative knowledge test, which was described above. The test consisted of 512 items.

*Study VII Results.* High skill subjects were significantly slower than low skill subjects in the new declarative knowledge test, but there was no difference in error rate. Thus, high skill subjects encountered negative transfer when forced used their declarative knowledge in a new way. The high and low skill subjects differed in their procedural knowledge in the same way they had in Study VI -- high skill subjects were approximately 400 ms faster than low skill subjects.

The findings of Studies VI and VII are interpreted as follows: skill acquisition begins with declarative knowledge that is transformed into procedural knowledge as training progresses. This procedural knowledge is largely implicit and not open to conscious inspection. The original declarative knowledge is still accessible to high skill performers, but it can not be manipulated as flexibly. High levels of skill can lead to poorer performance on tasks that require old declarative knowledge to be used in new ways. This is a potential cost of expertise.

Studies VI and VII were presented at the annual meeting of the American Educational Research Association (April, 1995) in San Francisco, California.

The first seven studies reported investigated evidence consistent with Anderson's ACT* theory of skill acquisition: evidence for composition, proceduralization, and the implicit nature of these processes. The next trio of studies investigated the phenomenon of undetected error making.

*Study VIII Objectives.* Study VIII contrasted high and low skill individuals on the number reduction task. It was hypothesized that high skill subjects would show greater evidence of composition and proceduralization. Furthermore, high skill subjects should make more errors on trials that resemble previously practiced rule sequences. This would be due to a partial match between the conditions for a well-learned, strong, composed production from training (with an extremely low firing threshold), and a new sequence that is similar to the well-learned production in its initial rules (these productions would be weak, with a relatively high firing threshold).

*Study VIII Methods.* Seventy-two University of Utah students participated in Study VIII. Subjects were divided into two groups: high skill (n=38) and low skill (n=34). During training both groups received practice on 12 of the 24 possible rule sequences (balanced for frequency of occurrence of each rule in each serial position). The high skill group received 55 blocks of 24 trials each of training over five sessions. The low skill group practiced on an unrelated computer task for three sessions. They then received 15 blocks of training over two sessions.

During all but the last five blocks of training (i.e., training occurring on sessions one through four) subjects were encouraged to go as fast as possible while maintaining an error rate of 10%. This was done to encourage fast, skilled performance on the task. During the last session (i.e., final five blocks of training) subjects were given a new performance goal: they were told to go as fast as possible, while being entirely accurate (i.e., error rate of 0%). Because it would be difficult to achieve this goal, they were told they could "retake" any trial on which they thought they had made an error by pressing the space bar. The space bar could be pressed, and the trial retaken, any time prior to answering the following trial. It was not possible, however, to retake earlier trials. We introduced the new performance goal, and the error retake method, to allow us to distinguish detected from undetected errors.

Transfer took place during the last 10 blocks of 24 trials of the final session. These blocks were not identified to subjects as being different from the first five blocks; however, they were comprised of three different trial types: old/old (old rule sequences using old instances; 25% of each transfer block of trials), old/new (old rule sequences using new instances; 25%), and new/new (new rule sequences using new instances; 50%). By this point, subjects were well experienced using the error detection methodology.

Finally, subjects took two questionnaires. The first questionnaire asked three questions about whether or not subjects had been using the spacebar as instructed. It was identical to the questionnaire described earlier under "Detecting Undetected Errors." The second was a listing of the 24 different rule sequences (e.g., LAST-MIDPOINT-SAME). Subjects were told that only 12 of the 24 rule sequences had been practiced during previous sessions. They were asked to circle "old" or "new" for each sequence, just as subjects in Study III had.

*Study VIII Results.* Both groups latency data were well fit by the power law of learning. First session performance (practice blocks 1-10) did not differ significantly between the two groups. Performance for the first five blocks of the final session displayed differences in latency reflective of the differing levels of practice given the two groups: high skill subjects were approximately 800 ms faster than low skill subjects. During the 10 transfer blocks, high skill subjects were again significantly faster than low skill subjects. The difference here was approximately 500 ms (low skill subjects had gained additional skill during transfer). More importantly, the difference between high and low skill subjects during transfer was a function of trial types. With regard to the difference between old/old and old/new trials (i.e., the instance effect indexing proceduralization), high and low skill subjects both showed differences on the order of 100 ms., and did not

differ significantly from one another. With regard to the difference between old sequences (old/old and old/new combined) and new sequences (new/new) (i.e., general sequence effects indexing composition), high skill subjects showed a significantly greater difference than did low skill subjects. Thus, expertise did interact with skill level in the latency data, but only for general sequence memory.

Error data were divided into two types: detected errors (on which the subject had hit the space bar to retake the trail) and undetected errors (on which the subject had failed to hit the spacebar). In general, there were somewhat more undetected errors (between 4% and 6%, depending upon trial type) than detected errors (between 2% and 4%, depending upon trial type). Only one condition significantly deviated from this general finding: high skill subjects made more undetected errors (approximately 10%) on new/new trials. This confirmed our hypothesis that new sequences, which resembled old sequences, would lead to greater numbers of undetected errors -- but only for subjects with a high skill level whose productions had become composed.



High n=25; Low n=29
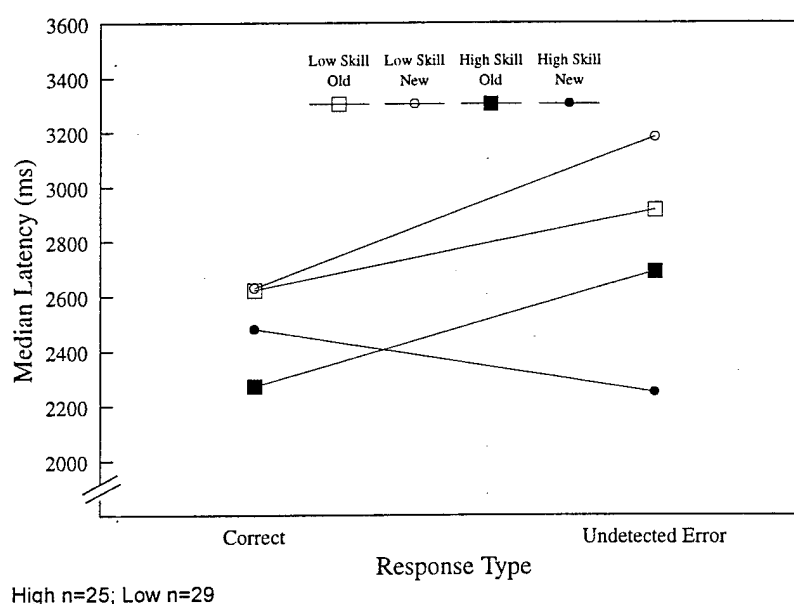
Another interpretation that could explain these data, and one proposed by Anderson (1989), is that subjects may inappropriately apply weak method solutions (e.g., mistaken analogies to previous problems), which may lead to undetected errors. To distinguish our hypothesis from this explanation, we analyzed latency to errors in Study VIII (see figure above). Our hypothesis predicts fast responses to undetected errors by high skill subjects who are misfiring composed, skilled productions due to a partial match of conditions. In contrast, Anderson's misapplication of weak methods predicts relatively slow responses leading to undetected errors, even among high skilled individuals. Latency data supported our prediction: high skill subjects responded as quickly on undetected errors on new/new trials as they did on correct trials. Low skill subjects, however, responded much slower on these errors than they did on correct trials, presumably because they had no composed productions to guide processing. This indicates that the nature of these errors differed as a function of skill level on the task. Another interesting finding from the error latencies was that high skill subjects were relatively slow on errors made to old trials (compared to correct responses). This may reflect instances in which subjects revert from using composed, skilled memory representations to older, more error prone processing.

Data from the questionnaire asking subjects if they had used the spacebar to correct errors indicated that the vast majority of subjects had conformed to the experimental instructions. In the low skill group, 81% reported using the spacebar to correct all errors they were aware of; in the high skill group, 63% reported using the spacebar to correct all errors. The difference between groups is not significant. Of those who did not use to spacebar to correct all errors, the estimated number of uncorrected errors ranged from 2 to 10 in the low skill group, and 2 to 24 in the high skill group (again, not significantly different between the groups). If one removes those subjects who failed to correct 7 or more errors (two high skill Ss and two low skill Ss) and reanalyzes the data, the results remain the same.

Data from the rule sequence recognition test showed a slight bias toward responding "old." However, the high and low skill groups did not differ from each other in their recognition performance as measured by $d'$. Thus, as in earlier studies, general processing sequence information appears to be primarily implicit in nature.

*Study IX Objectives*. Study IX was designed to test the "partial match" hypothesis in greater detail. It was argued in Study VIII that a partial match between the conditions of a new rule sequence and a strong-but-wrong old sequence resulted in the inappropriate firing of the strong-but-wrong production. Latency data from high skill errors supported this interpretation; the errors were as fast as correct trials. It is possible, however, that subjects learned performance timing information independently of the rule sequence information (see, for instance, MacKay, 1982, 1987). Since all new sequences in Study VI were matched with old sequences in their first two rules, it was not possible to rule out this competing explanation.

In Study IX subjects learned only eight rule sequences during training. During transfer, they were exposed to all 24. Eight of these were old. Of the 16 new rule sequences, eight matched the old sequences in their first two rules (just as they had in Study VIII). These were designated *partial-match new* sequences. The remaining eight began with an initial two rule sequence that had not been experienced during training. These were designated *mismatch new* sequences. If partial matching of conditions was the reason for the errors found in Study VIII, then we should find fast errors in the partial-match new condition, but not in the mismatch new condition.

*Study IX Methods*. Methods, procedures, and the experimental task were similar to those in Study VIII. Subjects were 49 University of Utah undergraduates. All subjects performed four sessions, with transfer blocks being presented during the last session. All subjects, therefore, were "high skill." Subjects received 10 training blocks during session one, 20 training blocks during session two, 20 training blocks during session three, and 5 training blocks during session four. During session four, the performance goal was changed from 90% accuracy to 100% accuracy, as in previous studies. Also during session four, retaking of trials by pressing the spacebar was introduced. The final 16 blocks of session four were transfer. Each transfer block consisted of the following: old sequences/old instances (33%), old sequences/new instances (33%), partial-match new sequences/new instances (17%), and mismatch new sequences/new instances (17%).

*Study IX Results*. Latency data showed a small, but reliable, instance effect on the order of 50 ms (difference between old/old and old/new). General sequence effects, as in previous studies, were much larger (and also reliable). When considering partial-match new trials, the sequence effect (partial match new/new versus old/new) was approximately 150 ms, with the partial match new trials being slower. When considering mismatch new trials, the sequence effect (mismatch new/new versus old/new) was approximately 400 ms, with the mismatch new trials being slower. There was a 250 ms difference between the partial-match new/new and the mismatch new/new, with the mismatch new/new being slower. Thus although previous studies showed negative transfer with regard to errors in the partial match new situation, there is also positive transfer with regard to latency, presumably due to the overlap of the first two rules.
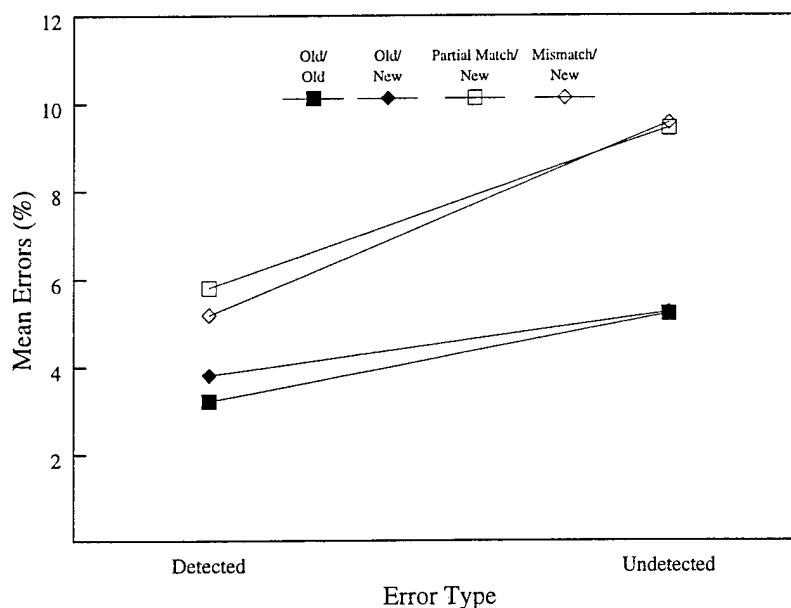
Error data are presented for the four conditions above. As can be seen from the figure, subjects made slightly more detected errors on partial-match new and mismatch new, as compared to old sequences. The partial-match and mismatch conditions did not differ from each other, however. This pattern was found in both detected and undetected errors, but the pattern was even stronger for the undetected errors.

The latency data for errors, which is the primary analysis of interest in Study IX, is displayed in the next figure below. The two old conditions (old/old and old/new) were combined in this analysis. The analysis also includes only those individuals who made at least two undetected errors in all trial conditions. First, note that the latency for partial match new undetected errors is fast, just as in Study VIII. It is as fast as correct old trials and correct partial mismatch trials. This is consistent with the interpretation made in Study VIII: undetected errors in the partial match condition appear to be due to the misfiring of strong-but-wrong composed productions that partially overlap with the rule sequences required for correct solution of these items.

Second, note that the latency for undetected errors on old sequences is quite slow. This finding was unexpected, but replicates a finding in Study VI. These errors seem consistent with what Reason (1990) has described as *overattention*. Such errors result from a performer intervening and consciously controlling performance, when, in fact, he or she would have been better off to allow skilled memory representations to guide performance. An example would be a baseball batter in a slump, who finds that his batting deteriorates even further as he tries to consciously "correct" his swing.

Studies VIII and IX have been submitted for publication in the *Journal of Experimental Psychology: General* (Woltz, D.J., Gardner, M.K., & Bell, B.G. [under revision]. Undetected mental errors during skilled performance.).

N=34
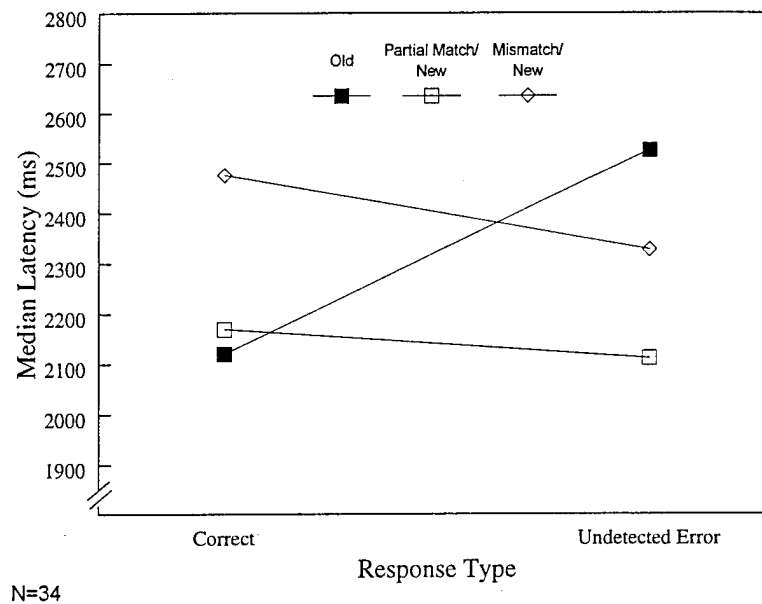
*Study X Objectives.* Study X attempted to explore the representation of sequence memory, such as that displayed in Studies VIII and IX, in greater detail. In Study IX, trials were divided into four categories: old trials with old instances (old/old), old trials with new instances (old/new), new trials that matched old trials in their first two rule sequences (partial match new), and new trials that did not match old trials at all (mismatch new). New trials in general lead to undetected errors, but partial mismatch trials lead to undetected errors with short latencies -- so-called "strong-but-wrong" errors. Study X attempted to contrast two alternative representations for sequence memory: one in which the primary unit of representation was the rule sequence triad (e.g., SAME-MIDPOINT-CONTIGUOUS), and the other in which the basic unit of representation was the rule sequence dyad (e.g., SAME-MIDPOINT; MIDPOINT-CONTIGUOUS). Anderson's (1983) composition mechanism would predict the former representation: after extended practice skills should become composed such that each of the individual productions (i.e., each rule) are fused together into a single production for the rule triad. Other theories of skill acquisition (e.g., MacKay, 1983, 1987) allow representation of skills at different levels: productions may be represented singly, in dyads, or in more complex arrangements. Study X contrasted old trials and new trials that differed in numerous ways.

*Study X Method.* The design of Study X was similar to that of Study IX. Subjects received 55 training blocks over four sessions containing multiple instances of eight rule sequences. The frequency of use of individual rules was balanced over serial positions of occurrence. During transfer, subjects received 18 blocks of 24 trials each. There were no training blocks during the transfer session (in contrast to Study IX), and error detection was introduced at the beginning of transfer. Transfer trials were divided into three types: old/new (old rule sequences with new instances); partial match trials (new rule sequences that matched the old trials on the first two rules); and mismatch trials (new rule sequences that did not match the old trials on the first two rules). An example of the three trial types is given in the table below.

Example of Three Transfer Trial Types Used in Study X

| Old Sequences | Partial-Match New Sequences | Mismatch New Sequences |
|---|---|---|
| C-L-M | C-L-S (++) | C-S-L (X+) |
| C-M-L | C-M-S (+X) | C-S-M (XX) |
| L-S-C | L-S-M (+X) | L-M-S (XX) |
| L-C-S | L-C-M (++) | L-M-C (X+) |
| M-S-L | M-S-C (++) | M-C-S (X+) |
| M-L-S | M-L-C (+X) | M-C-L (XX) |
| S-C-M | S-C-L (+X) | S-L-C (XX) |
| S-M-C | S-M-L (++) | S-L-M (X+) |

Note: C=Contiguous; L=Last; M=Midpoint; S=Same.

The trial types in Study X, however, could be classified in another way: by rule dyad and rule triad. The first and second rules comprised the first rule dyad, while the second and third rules comprised the second rule dyad. According to this classification scheme, old sequences matched training trials in both first and second rule dyad, as well as rule triad. Four other possibilities existed among the new sequences: (1) first and second dyad match (but no triad match; denoted (++) above); (2) first dyad match and second dyad mismatch (denoted (+X) above); (3) first dyad mismatch and second dyad match (denoted (X+) above); and (4) first and second dyad mismatch (denoted (XX) above). Comparing these conditions provides information regarding the representation of processing sequence memory, and the feasibility of the composition mechanism. Sixty-seven University of Utah students served as subjects in Study X.

*Study X Results.* Results concerning undetected errors replicated those of Study IX (see figure below). Both partial match and mismatch trials produced more undetected errors than old trials. Also, there were more undetected errors in general than detected errors.
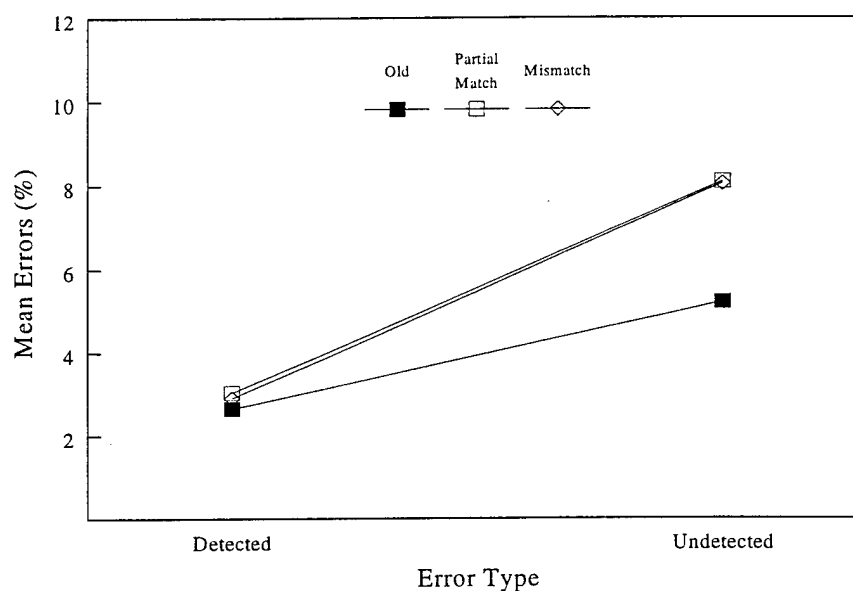
Latency results for the five dyad/triad combinations are presented in the figure below. As can be seen from the figure, trials with a triad match (also matching first and second dyads) were fastest. Trials that matched in the first dyad were next fastest (i.e., line connected by open squares), while trials that matched only in the second dyad were slowest (i.e., line connected by filled squares). These results are most consistent with an ordered hierarchical representation scheme for sequence memory. Mismatches detected in the first dyad cause a breakdown of automatic processing: older (and presumably more time consuming) productions must take over from the start. A matching first dyad allows automatic processing to begin, but it may be disrupted by a mismatch in the second dyad, or, more notably, in the triad. The results are inconsistent with composition, as Anderson has defined it. According to composition, dyad effects should not be present. The results can be accommodated by hierarchical network approaches, such as MacKay's.
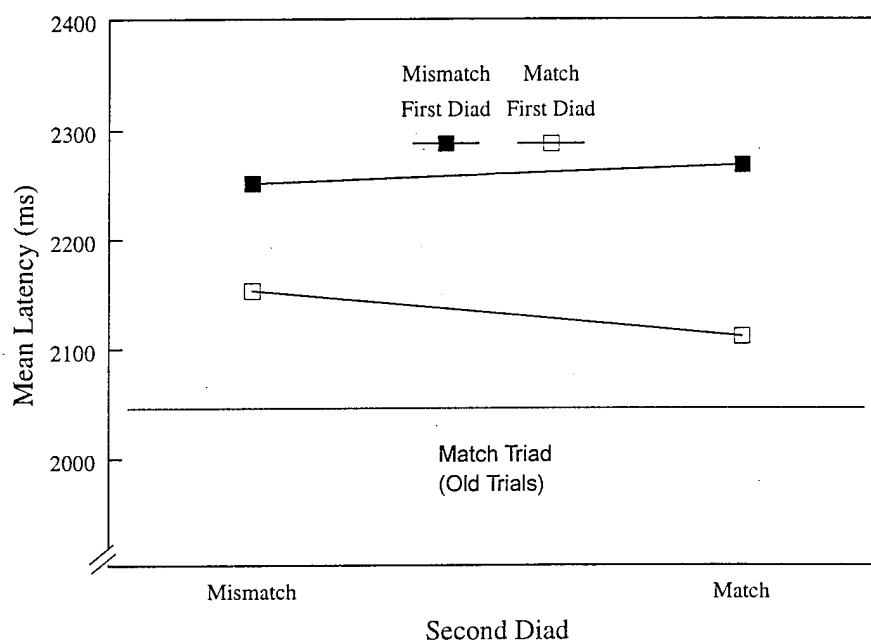


*Study XI Objectives.* Study XI attempted to generalize the finding of undetected errors due to long-term priming to a different task: procedural learning.

*Study XI Methods.* Procedural learning, used in Studies VI and VII, involved learning a procedure (i.e., a set of rules) for classifying numbers presented on the computer screen and giving a binary response (either "L" for like or "D" for different). The procedure is taught to subjects as a hierarchical classification scheme:

When a word is presented, first determine if it is A WORD (e.g., one, two, three) or A DIGIT (e.g., 1, 2, 3).

    If the number is A WORD (e.g., one, two, three), then determine if it is ODD or EVEN.

        EVEN belongs on the top of the computer screen and ODD belongs on the bottom.

            If it is EVEN on TOP or ODD on BOTTOM, press "L" (like), otherwise press "D" (different.

    If the number is A DIGIT (e.g., 1, 2, 3), then determine if it is SMALL (less than 50) or BIG (greater than 50).

        SMALL belongs on the TOP of the computer screen and BIG belongs on the bottom.

            If it is SMALL on TOP or BIG on BOTTOM, press "L" (like), otherwise press "D" (different).

The classification procedure is represented in the decision tree presented below (this tree was not presented to subjects). To induce strong-but-wrong errors, some branches of the tree were given more practice (represented by heavy arrowed paths) than others during training. The ratio of strong to weak practice was varied in two versions of Study XI. In version one, the ratio was 5:1; in version two, the ratio was 2:1.
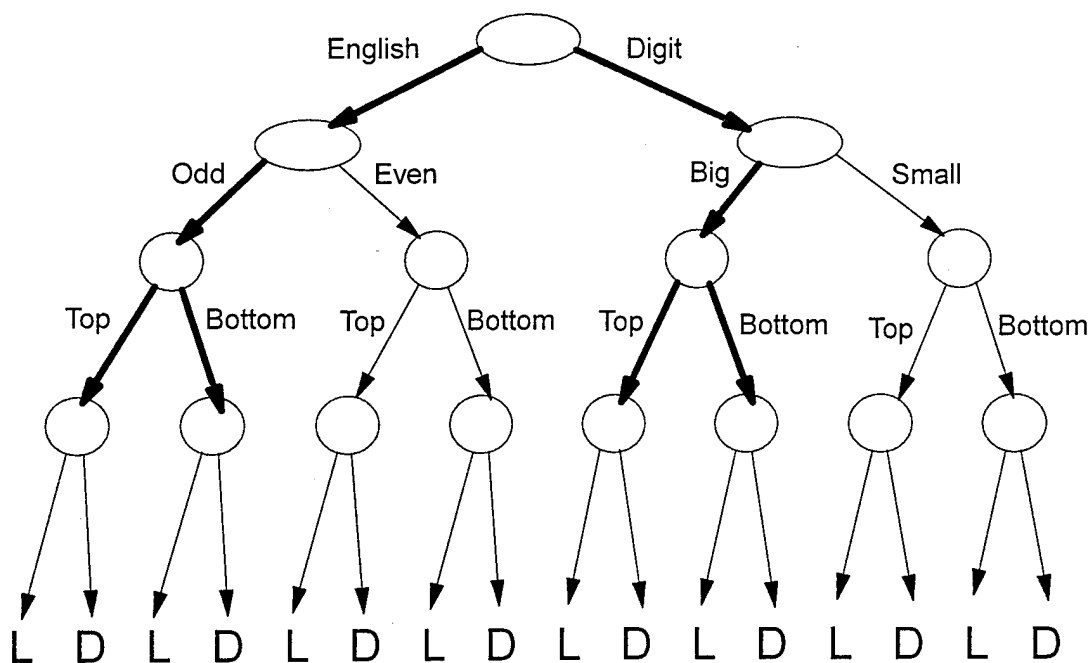


During training, subjects received 40 blocks of 48 trials each. Training took place over 3 sessions. During transfer (session 4), subjects received 15 blocks of 48 trials each. During transfer, all branches of the decision tree occurred equally often. We predicted, based on previous long-term priming studies, that subjects would make more undetected errors on rule sequences that had been practiced relatively infrequently. Furthermore, we predicted that undetected errors should be greater in the version of the task with a 5:1 training ratio between strong and weak rule sequences, since this allowed greater opportunity for composition of the strong rule sequences.

Ten University of Utah students served as subjects in Study XI.

*Study XI Results.* Study XI produced an unexpected result: almost all errors were detected (see figures below). In both versions of the task, there were fewer than 2% undetected errors, while detected errors ranged from 3% to almost 12%. Detected errors were far more numerous in version one of the task (5:1 training ratio) than in version two (2:1 training ratio). Also, there was a trend (nonsignificant) for weak rule sequences to produce more detected errors than strong rule sequences. There was no difference between weak and strong rule sequences with regard to undetected errors.
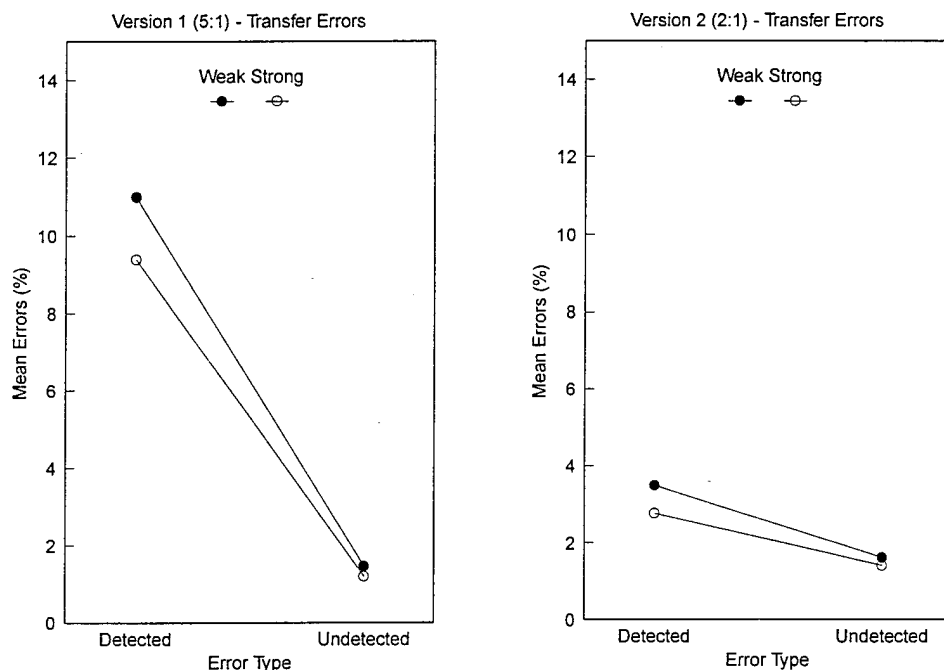


Study XI produced some important questions for future research. Why is it that number reduction produced large numbers of undetected errors and relatively few detected errors, while procedural learning produced the opposite pattern. Further, why were long-term priming effects, to the extent they were present in procedural learning, so much smaller than in number reduction tasks. The answer would seem to lie in differences between the two tasks. First, number reduction has multiple possible responses (i.e., the digits 1 through 9) while procedural learning is a binary choice task (i.e., "L" or "D"). Second, number reduction must be performed serially from left to right, in the order in which digits are presented on the screen. This is because intermediate answers must be calculated as part of the solution process. Procedural learning may not need to be learned serially. All relevant stimulus attributes are present in the presented stimulus, and they need not be classified in the manner originally taught. Finally, instance effects are small for number reduction (see discussion of experiments presented above), but they are relatively large for procedural learning (Woltz, 1991). Further research needs to address how these task differences result in performance and error making differences. If the task differences can delineate categories of tasks in which undetected errors are likely or not likely, this would be important progress.

*Study XII Objectives.* This study investigated undetected errors due to the misfiring of strong-but-wrong sequence memory in a new skill task. The task was more complex than number reduction or procedural

learning. It involved a sequence of six computation steps, some of which required output values from previous steps. In addition, we tested the role of sequence memory in skilled performance errors in a slightly different manner than previous number reduction studies. Here we tested the hypothesis that composition of processing steps would underlie undetected errors when task rules were modified following extensive practice.

*Study XII Methods.* Elio (1986) introduced a sequential computation task (we denoted it *value computation*) for calculating water quality indices. Subsequent to this, Frensch (1991) and Carlson and Lundy (1992) have used the task to study composition of processing sequences. The task requires six computations shown in the example problem below.

|  |  |  | Lime | Toxin |
|---|---|---|---|---|
|  |  | (a) | 11 | 8 |
| Solid | Algae | (b) | 11 | 9 |
| 18 | 20 | (c) | 18 | 1 |
|  |  | (d) | 14 | 6 |

1. Particulate Rating   _____   = *Max(Solid, Algae)*

2. Mineral Rating   _____   = $Min(Lime_a, Lime_b, Lime_c, Lime_d )$

3. Index 1   _____   = *Particulate Rating - Mineral Rating*

4. Marine Hazard   _____   = *(Toxin Min + Toxin Max)/2*

5. Index 2   _____   = Index1 x  Marine Hazard

6. Overall Danger   _____   = *Index1 + Index2*

During training, we varied whether subjects had the opportunity to develop sequence memory representations (i.e., composition of the steps). One group (n=37) practiced the six steps in sequence (1-6). The other group (n=24) practiced the steps in a blocked fashion (i.e., a subject would practice a block of Step 1 computations, then a block of Step 2 computations, etc.). In both conditions, answers to prior steps of the same problem were always visible. We assumed that both groups obtained equivalent training on the component computations, but that they differed in composition or other forms of memory for step sequence and transition. Both groups received three sessions of training, with 120 problems per session (each problem had six steps). During a subsequent transfer session, both groups performed the problems in the proper sequence (1-6). In addition, the task rules were modified slightly. Subjects were informed of the following conditional change involving Steps 3 and 5.

If Index1 ≥ 3, then compute Index2 as before (Index1 x Marine Hazard).
If Index1 < 3, then compute Index2 as (Marine Hazard)$^2$

Subjects performed 120 problems with the new rules and error correction instructions similar to those in previous experiments. We hypothesized that subjects with composed memory for the processing steps (or some alternate form of sequence memory) would have more difficulty inserting the conditional modification.

*Study XII Results.* Subjects who had blocked training showed a speed advantage for problem steps that did not rely on previous step solutions. As shown in the figure below, mean latency for the blocked condition was less than that for the sequential condition for Steps 1 and 2, and to a lesser extent for Step 4. In contrast, for two of the three integrative steps (3 and 6), the sequential training condition showed an advantage. It is not clear to us why Step 5, which is integrative, did not resemble Steps 3 and 6. Despite

this discrepancy, the latency data from training blocks generally supported the assumption that subjects in the sequential condition developed sequence memory which primarily facilitated integrative steps. The error data from training did not differ by condition.
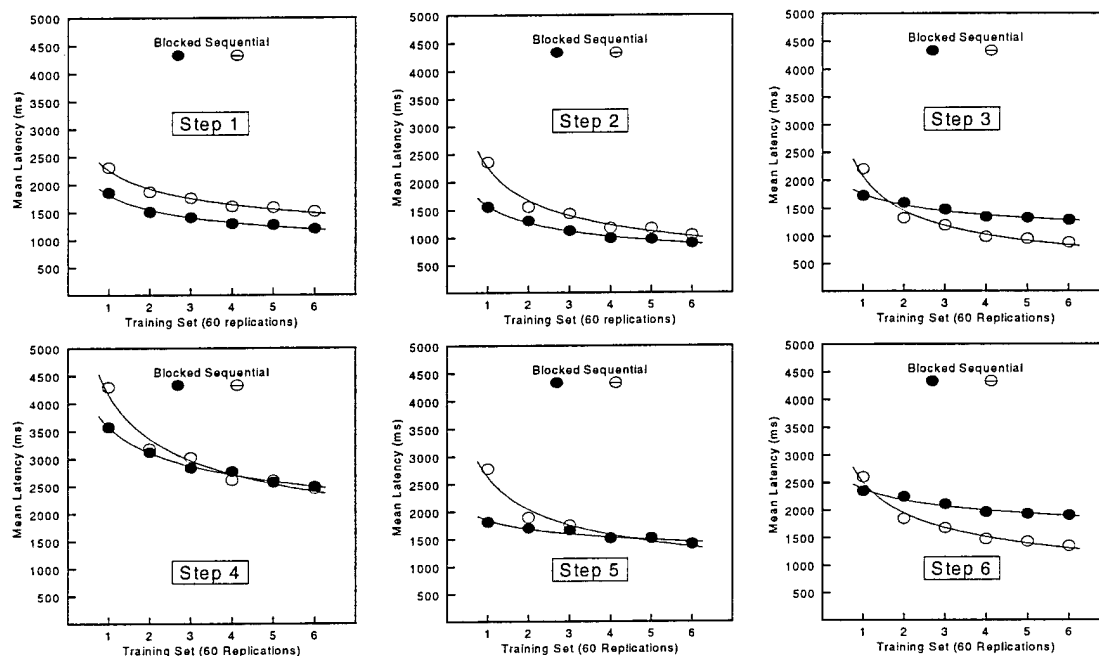


The transfer latency data also supported this conclusion. As shown in the next figure, the blocked training condition showed longer latency for all steps except Step 1 when the problems had to be solved in sequence (Steps 1-6). However, over the 12 transfer blocks, the group difference diminished substantially. That is, subjects in the blocked condition quickly began to show evidence consistent with composition of processing steps.
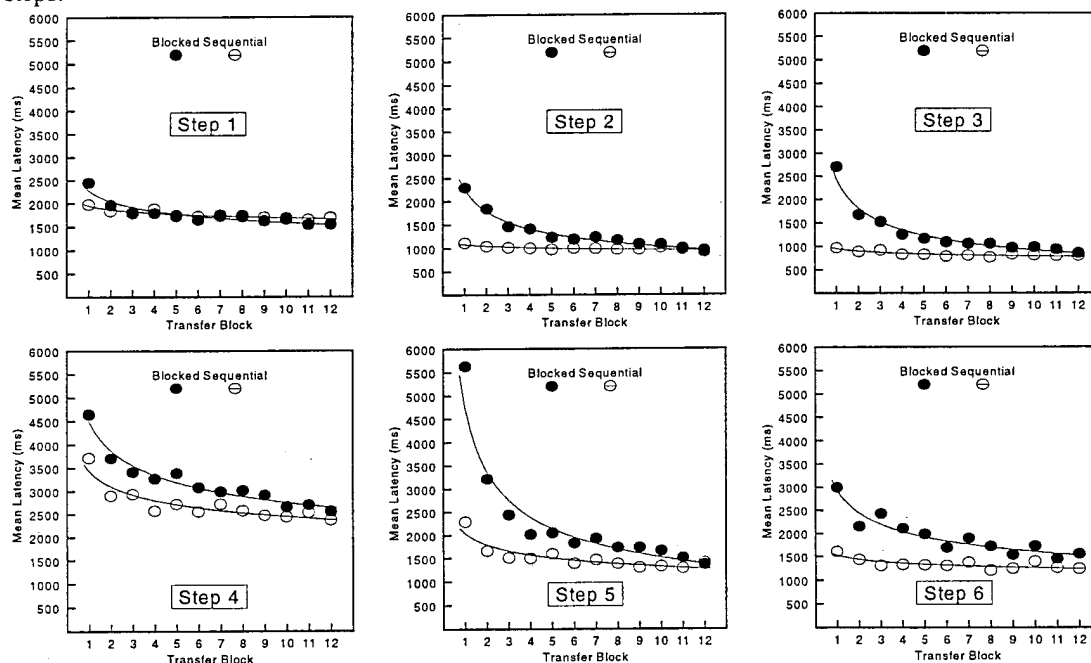
The primary hypothesis about strong-but-wrong intrusion errors was tested with error data from the transfer session. As hypothesized, subjects in the sequence training condition made more errors on critical steps (3, 5, & 6) following the introduction of the new conditional rule. The undetected error rates for each step in transfer are shown below by training condition.
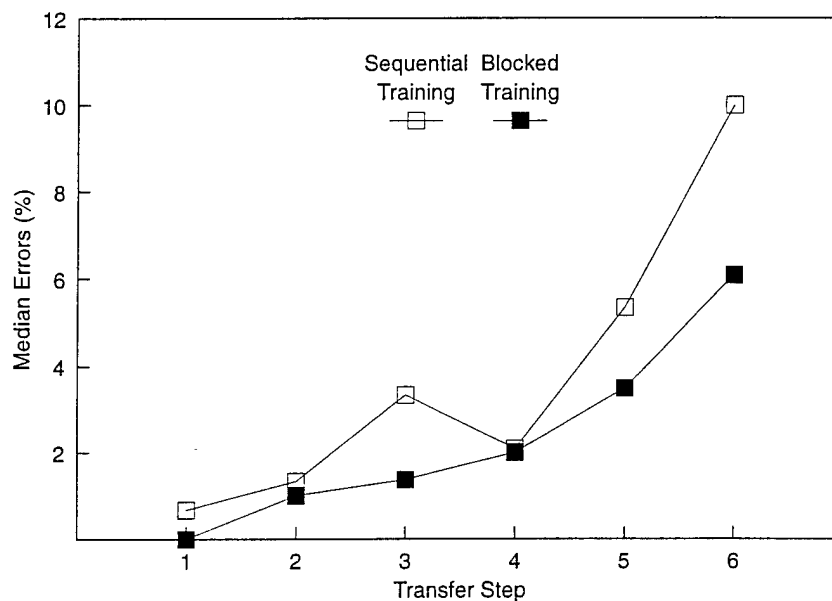


We interpreted these data as being consistent with those from previous number reduction experiments. We found evidence consistent with composition or other forms of memory for processing sequence. In this case, subjects who had been allowed to practice a standard processing sequence showed a clear performance speed advantage over those who had not. However, as was the case in earlier experiments, composition was also shown to be detrimental under some transfer conditions. When subjects had to make slight modifications to existing rules, composition appeared to hinder accurate performance. This is consistent with the notion that composed productions are inaccessible to conscious awareness, and as such, they are not easily modified.

*Conclusions*. The studies on long-term priming present a generally consistent picture of cognitive skill acquisition. Subjects retain information on both the specific instances they have seen, and the order of processing operations that are general to many instances. Although instance effects in learning have been demonstrated in many previous studies, it has not been well-established that memory for processing sequences is an important determinant of performance. Furthermore, our findings that processing sequence memory can lead to undetected errors in highly skilled individuals is novel. Finally, it is of theoretical importance that sequence memory, and also instance memory, in the paradigms we studied, were primarily implicit in nature. That is, subjects did not have conscious access to these representations. Our evidence was, for the most part, consistent with Anderson's ACT* theory of skill learning and the processes of composition and proceduralization. One study on the basic unit of sequence representation, however, was more consistent with a hierarchical network theory, such as MacKay's.

Short-Term Priming

The first set studies established undetected errors in a training/transfer paradigm where training was restricted to a subset of the total universe of possible rule sequences in the number reduction task. Short-term priming studies were aimed at investigating the possibility that undetected errors due to strong-but-wrong composed productions could be induced over a small number of trials within a single block. The

rationale was similar to that for long-term priming: if either the repeated firing of a production temporarily strengthens it, or if the presentation of contextual information associated with the firing of a production spreads activation that temporarily lowers the firing threshold of the production, then temporary priming should be possible. Predictions were similar to those for long-term priming: we would expect errors in situations where a temporarily primed production shared a partial match to the conditions of an item.

*Study XIII Objectives.* Study XIII investigated whether temporary priming could be found in the complex version of the number reduction task. Subjects became highly practiced on all 24 of the possible three rule sequences (using different instances). During transfer, they encountered groups of one, two, and three consecutive trials using the same rule sequence, and then were switched to a partial match sequence that differed only in the final rule. The question of interest is whether such "short-term primed" targets would show larger numbers of undetected errors than matched trials that were unprimed.

*Study XIII Methods.* Thirty-two University of Utah students served as subjects in Study XIII. Three were eliminated because their data indicated lack of effort, leaving 29 students in the final data analysis. Subjects received five training sessions of 12 blocks each, with each block containing 24 trials. This allowed each of the 24 possible rule sequences to appear once in each block. Each presentation of a rule sequence used a unique instance of that sequence. The final (sixth) session was a transfer session containing 24 blocks of 30 trials each. The first six trials in each transfer block served as "warm-up" trials, and contained no priming pattern. The final 24 trials were divided into six groups of four trials each. Half of these were "short-term primed" groups, and half were "short-term unprimed" controls. In a "short-term primed" group, whichever rule sequence was assigned as a target (e.g., SAME-MIDPOINT-LAST) would be preceded by either one, two, or three partial match rule sequences (e.g., SAME-MIDPOINT-CONTIGUOUS). Each transfer block contained one one-prime, one two-prime, and one three-prime grouping of four trials. Thus three rule sequences received short-term priming in each transfer block. Three other rule sequences were assigned as "short-term unprimed" targets for the other three groups of four trials. These rule sequences were not preceded by partial match priming trials. We tested our hypotheses about short-term priming by comparing target trial errors following primed versus unprimed sequences. Primed and unprimed groups of four trials alternated during transfer blocks. As in previous studies, subjects could use the spacebar to correct errors during transfer.

*Study XIII Results.* Latency data revealed no significant differences between short-term primed and unprimed trials. Likewise, number of priming trials preceding a target was not related to latency. Similar results were found for number of undetected errors: priming and number of priming trials were not related to number of undetected errors.

*Study XIV Objectives.* In Study XIII, subjects received practice on all 24 possible rule sequences during each training block. It is possible that subjects composed not only the three rules in each sequence during training, but also the process of switching from one rule sequence to another (e.g., see Carlson and Yaure, 1990). This would have minimized the likelihood of undetected errors during transfer, as subjects had become skilled at switching from one rule sequence to another. To minimize this learning, Study XIV replicated Study XIII, except that training was blocked by rule sequence. Thus, in any given training block, subjects practiced only one rule sequence.

*Study XIV Methods.* Methods for Study XIV were similar to those for Study XIII, except for the following changes: (a) training was blocked by rule sequence, so that only a single rule sequence was practiced per training block; (b) the number of training sessions was reduced from five to four, and the number of transfer blocks was increased from one to two (this was done to increase the number of opportunities for making undetected errors); and (c) the last rule, which stated that if two digits differed by more than two, they could be reduced to the last of the two digits, was replaced by the first rule, which stated that if two digits differed by more than two, they could be reduced to the first of the two digits. This last change was required by the blocking procedure. When the last rule occurred in the final position (e.g., SAME-CONTIGUOUS-LAST), problems could be solved by simply reporting the final digit of the problem. With the first rule, subjects could not use this strategy, since the first digit of the final two could be modified by intermediate solutions to the first two rule applications within a problem.

Nineteen University of Utah students served as subjects in Study XIV. None were eliminated due to lack of effort; thus, 19 subjects were included in the final data analysis.

*Study XIV Results.* Once again, both latency and error data showed no effect of short-term priming.

*Study XV Objectives.* Studies XIII and XIV found no evidence of undetected errors due to short-term priming. The rationale for these experiments was straight forward, and the priming of target rule sequences was direct. In Study XV, we explored short-term priming of an indirect nature. We reasoned that it might be possible to lower the firing threshold for a composed production by presenting contextual cues present during the composition of that production. If this contextual information was later (i.e., during transfer) presented inappropriately with a partially matching rule sequence, it could lead to an undetected error due to firing of the strong-but-wrong production associated with the contextual cue.

*Study XV Methods.* The task in Study XV was the complex number reduction task. Subjects practiced eight rule sequences during training: four randomly chosen sequences and four corresponding partial match sequences (matching their mates in the first two rules). The experiment took place over four experimental sessions. During session one, subjects received 12 training blocks; during session two, 18 training blocks; during session three, 18 training blocks; and during session four, 3 training blocks. The remaining 15 blocks of session four were transfer blocks. Each block in training and transfer was 32 trials long. As in previous studies, error detection was introduced during the final session.

A major difference between Study XV and previous studies concerned the presentation of the digit strings to be reduced. In previous studies these had been presented in the center of the computer screen against a black background. In Study XV, each of the eight rule sequences was presented in a different spatial location (at either 0, 45, 90, 135, 180, 225, 270, 315 degrees of orientation from vertical, approximately 2 inches from the center of the screen) and in a different color. Sequences were oriented so that their partial match mates were opposite them in terms of presentation position (i.e., sequence A at 0 degrees, and partial match A at 180 degrees) and presented in the complementary color. Presentation position and color cues were perfectly correlated with rule sequences during training. During transfer, two types of trials were possible: (a) "switched" trials, in which the presentation position and color of a rule sequence were switched to its partial match mate (25% of trials), and (b) consistent trials, which maintained the mapping of presentation position and color to rule sequence learned during training (75% of trials). Each rule of the eight rule sequences was switched once during each transfer block. We hypothesized that undetected errors would be more numerous on switch trials, due to short-term contextual priming.

Eleven University of Utah students served as subjects in Study XV.

*Study XV Results.* Error data did not support an increase in number of undetected errors on "switch" trials. Thus, there was no evidence in favor of short-term priming due to indirect or contextual cueing.

*Study XVI Objectives.* Previous short-term priming studies had stressed high levels of performance skill among subjects. Study XVI examined the possibility that short-term priming occurs primarily among subjects in the earlier stages of skill acquisition (i.e., low skill subjects). We predicted short-term priming of the sort employed in Studies XIII and XIV would produce undetected errors among subjects with relatively little training on the number reduction task.

*Study XVI Methods.* Study XVI was a replication of Study XIV, except that subjects received no training and there were two transfer session as opposed to one. Error detection and short-term priming were the same as in Study XIV.

Fourteen University of Utah students served as subjects in Study XVI.

*Study XVI Results.* As in Study XIV, both latency and error data showed no effect of short-term priming.

*Study XVII Objectives.* It is possible that the lack of short-term priming was a function of the number reduction task itself. Study XVII studied short-term priming in a different task -- procedural learning.

Study XVII Methods. The procedural learning task used in Study XVII was described in detail in Study XI, and will not be repeated here. However, the following differences between Study XI and Study XVII are noted:

1.  All branches of the decision tree were presented equally often.
2.  No training was given to subjects.
3.  Transfer consisted of 18 blocks of 48 trials each presented in a single session.

Short-term priming was accomplished as follows: each set of 48 trials consisted of two subsets of 24 trials each. The first eight of these trials served as warm-ups. The next 16 consisted of four sequences of four trials each: two unprimed and two unprimed. Priming involved preceding each target by two or four trials representing the same branching sequence of the decision tree. Then the target trial switched one component (top versus bottom, English versus digit, odd versus even, or big versus small).

Thirteen University of Utah students served as subjects in Study XVII.

Study XVII Results. Both latency and errors showed no effect of short-term priming. Thus, we feel safe in concluding that the findings of the earlier studies were not solely due to the peculiarities of the number reduction task.

*Conclusions.* The failure to find evidence of undetected errors in short-term priming situations was surprising, both in light of the success of producing undetected errors through long-term priming and anecdotal reports of errors due to such short-term priming phenomena. Almost all theories of error making (e.g., Heckhausen & Beckman, 1990; Norman, 1981; Reason, 1990) postulate a role for short-term priming. Future research needs to determine if the failure to find short-term priming is due to task and method variables peculiar to the current research, or whether current theories need to be revised in light of our findings.

## Working Memory Load

A consistent theme in the writings of theorists concerned with error making is that a working memory load leads to errors (e.g., Heckhausen & Beckman, 1990; Norman, 1981; Reason, 1977, 1990). In general, the theories say the following: if an individual is engaged in a task that they have routinized (i.e., are highly skilled at), and she or he is also engaged in second activity that is not routinized (so that it consumes working memory), this individual is susceptible to the intrusion of strong-but-wrong processes that are competing with the routinized activity. An example would be making tea while holding a conversation. One might inadvertently put instant coffee in the cup rather than a tea bag, resulting in a cup of coffee rather than tea. The strong-but-wrong coffee process intruded on the weaker tea processes, due, in part, to the load put on working memory by holding a conversation simultaneously with making a hot beverage.

*Study XVIII Objectives.* A study was devised to investigate to effect of a working memory load on undetected errors due to the intrusion of strong-but-wrong productions. Two groups of subjects were practiced on the complex version of the number reduction task. One group of subjects received a small amount of practice on the task, while the other group was highly practiced. Transfer blocks alternated between those requiring subjects to engage in a secondary task (i.e., memory load condition) and those without a secondary task (i.e., no memory load condition). We hypothesized that subjects would make greater numbers of undetected errors with a memory load, and that high skill subjects would show a larger effect for memory load than low skill subjects. This is because only high skill subjects have composed productions that can partially match strong-but-wrong productions during transfer.

*Study XVIII Methods.* Seventy-six University of Utah students served as subjects for Study XVIII. Thirty-eight subjects were randomly assigned to the high skill group, and 38 subjects were randomly assigned to the low skill group. For the high skill group, session one contained 10 training blocks on 12 of the 24

possible rule sequences. Sessions two, three, and four contained fifteen training blocks each. For the low skill subjects, sessions one, two, and three involved participation on an unrelated computerized task. Session four contain 10 training blocks on 12 of the 24 possible rule sequences. For both groups, session five consisted of 5 more training blocks, and 10 transfer blocks. Also during session five, the performance criterion was changed from 90% correct to 100% correct, with the possibility of retaking error trials. The 10 transfer blocks contained three types of trials: old rule sequences/old instances (25% of trials); old rule sequences/new instances (25% of trials); and new partial match sequences/new instances (50% of trials). For half of the 10 transfer blocks, subjects were also required to perform a secondary task designed to produce a working memory load. This task consisted of keeping track of how many times two randomly selected digits had served as responses to the number reduction problems within that block. New digits were selected at random for each working memory load block. Working memory load blocks alternated with blocks that did not require the secondary task (i.e., no working memory load blocks).
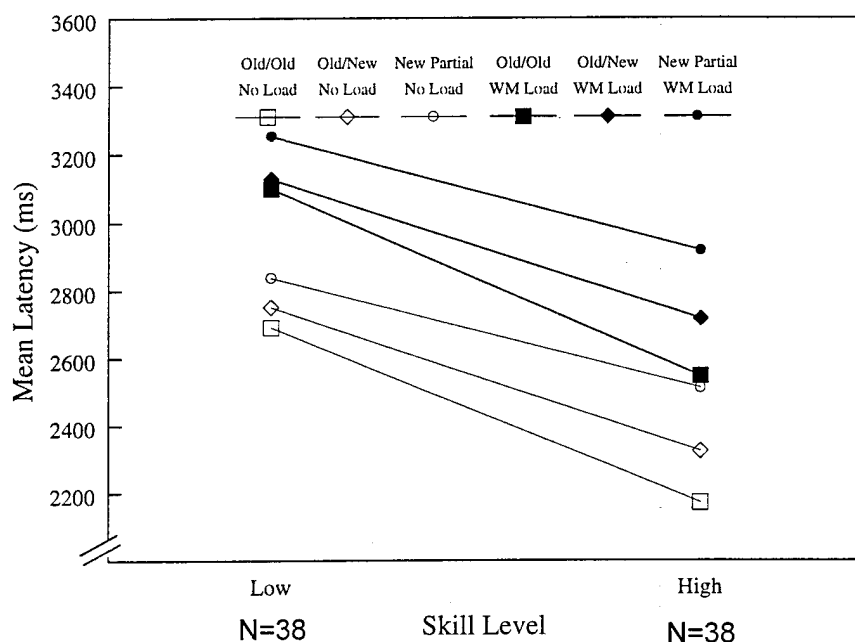


*Study XVIII Results.* Latency results for Study XVIII are presented in the figure above. First, there was a significant effect of skill level ($F[1,74] = 8.25, p < .01$), with high skill subjects performing more quickly than low skill subjects. Second, there was a significant effect for old (old and new instances combined) versus partial match sequences, ($F[1,74] = 64.10, p < .001$), with partial match sequences being slower than old sequences. Third, there was a significant interaction between skill level and sequence type ($F[1,74] = 8.63, p < .01$). The benefit due to a sequence being old was greater for high skill subjects than for low skill subjects. This is consistent with high skill subjects having general processing sequence memory representations, e.g., composition. Fourth, there was a significant effect of working memory load ($F[1,74] = 117.83, p < .001$). Having a working memory load slowed subjects performance considerably. However, working memory load did not significantly interact with skill level ($F[1,74] < 1, p > .10$) as we had predicted. Working memory load slowed both high and low skill subjects approximately equally.
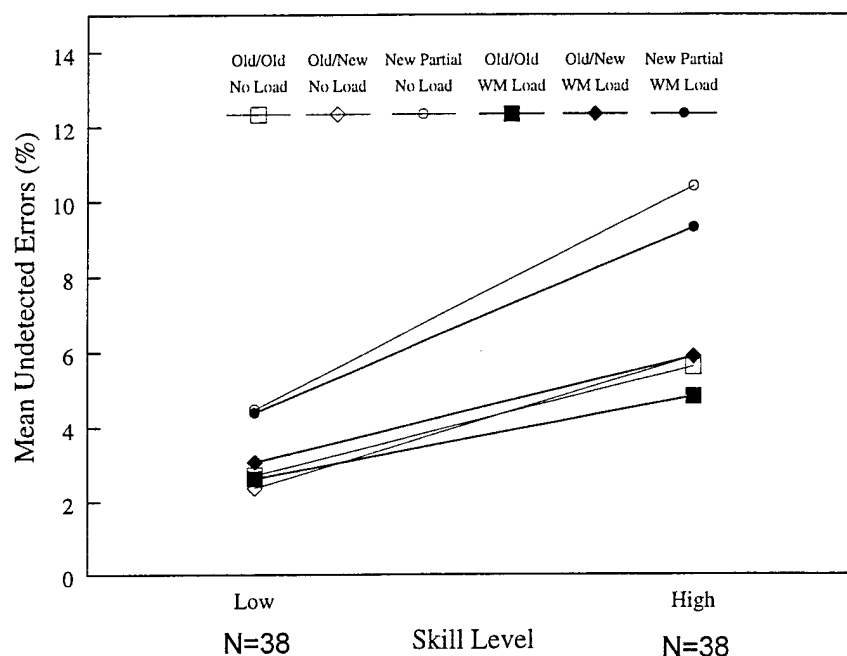
Error data for Study XVIII are presented in the figure above. First, there was a significant effect of skill level ($F[1,74] = 9.19, p < .01$), with high skill subjects making more undetected errors than low skill, consistent with composition. Second, there was a significant effect of sequence type (old combined versus partial match; $F[1,74] = 18.35, p < .001$), with partial match sequences producing greater numbers of undetected errors. Third, the interaction of skill level with sequence type approached significance ($F[1,74] = 3.32, p < .09$), suggesting a greater tendency toward undetected errors on partial match sequences in high skill subjects. Fourth, there was no effect for working memory load ($F[1,74] < 1, p < .10$), which was a surprise. While memory load affected latency, it did not affect error rate. Finally, there was no significant interaction between working memory load and skill level ($F[1,74] =< 1, p > .10$). Working memory load did not affect error rate, and this finding was constant for high and low skill subjects.

We were surprised to find that working memory load did not have a detrimental effect on errors, especially among high skill subjects. This may have been due to the nature of the working memory load manipulation: tracking responses. Subjects may have been able to swap between the dual tasks. This strategy would have inflated latency (since two tasks take longer than one), but had no affect on error rate, since only one task was being performed at any given time.

*Conclusions.* In our initial investigation, working memory load did not foster undetected errors among highly skilled performers. This is puzzling, because all current theorists note that error making is more likely under concurrent task demands. We believe this as well. Future studies need to attempt to replicate our findings with a secondary task that cannot be swapped with the primary task. That is, a secondary task needs to be found that is so intimately entwined with the primary task (e.g., number reduction), that it must be performed concurrently. We are currently working on developing such manipulations. If these procedures are unable to produce an interaction between skill level and procedures designed to foster undetected errors, current theories may have to be modified.

Individual Differences

From both a theoretical and an applied perspective, knowing who is particularly susceptible to undetected error making is of interest. Tests, such as the Cognitive Failures Questionnaire (Broadbent, Cooper, Fitzgerald, & Parkes, 1982) and Reason & Mycielska's (1982) Absentmindedness questionnaire, exist which purport to measure an individual's likelihood of making slips. However, one must remember that such tests are paper-and-pencil measures, not performance measures. As such, these measures may be susceptible to social desirability. Furthermore, individuals may not be able to introspect on their propensity to make undetected errors. Our research on long-term priming indicated memory representations for sequence information were implicit in nature. The following studies investigated individual differences in error making among skilled performers.

*Study IXX Objectives.* Study IXX was exploratory in nature, and investigated the relationship among: (1) paper-and-pencil measures of the propensity to make errors, (2) other paper-and-pencil personality measures that potentially could be related to the tendency to make errors, and (3) error making under long-term priming conditions.

*Study IXX Methods.* A subset 44 of subjects from Study X served in Study IXX. In addition to performing the number reduction task with its long-term priming component, as described in Study X, these subject received the following paper-and-pencil tests:

Self Report Measure of Error-Proneness
> *Cognitive Failures Questionnaire* (CFQ: Broadbent, Cooper, Fitzgerald, & Parkes, 1982)
> *Absentmindedness Questionnaire* (AQ: Reason & Mycielska, 1982)

Personality Measures Potentially Related to Error Making
> *Dickman Impulsivity Scale* (Dickman, 1990) subscales:
>> Functional Impulsivity (FI)
>> Dysfunctional Impulsivity (DI)
> *Jackson Personality Inventory* and *Personality Research Form* (Jackson, 1976, 1989) subscales:
>> Achievement Scale (ACH)
>> Cognitive Structure Scale (CS)
>> Desirability Scale (DSR)
>> Endurance Scale (END)
>> Impulsivity Scale (IMP)

Number of both undetected and detected errors were also recorded for each subject.

*Study IXX Results.* The correlation table presented below summarizes the findings of Study IXX. Several results are apparent. First, self report measures of error-proneness correlate well with each other, but have little or no correlation with error making (either detected or undetected). Second, several of the personality measures correlated significantly with each other, but at best showed only moderate relationships with error making. Third, some personality measures -- desirability and achievement -- showed significant relationships with the CFQ, indicating that self report measures are influenced by social desirability, at least among some subjects. We conclude that self report and performance measures of error-proneness show a divergence. Subjects show little ability to introspect on their propensity to make errors, a finding consistent with the implicit nature of sequence memory (which we implicated in undetected errors in earlier studies).

# Study IXX Results

Correlations of Self-Report Measures of Error-Proneness,
Personality Measures, and Detected and Undetected Errors

| | CFQ | AQ | FI | DI | ACH | CS | DES | END | IMP | UNDET. ERRORS | DET. ERRORS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CFQ | 1.00 | | | | | | | | | | |
| AQ | .64*** | 1.00 | | | | | | | | | |
| FI | .08 | .24 | 1.00 | | | | | | | | |
| DI | .30 | .17 | -.02 | 1.00 | | | | | | | |
| ACH | -.31* | -.17 | .20 | -.09 | 1.00 | | | | | | |
| CS | .09 | .04 | -.07 | -.29 | -.14 | 1.00 | | | | | |
| DES | -.41** | -.22 | .06 | -.34* | .49** | -.08 | 1.00 | | | | |
| END | -.24 | -.05 | .30 | -.01 | .71*** | -.02 | .47*** | 1.00 | | | |
| IMP | .19 | .18 | .12 | .73*** | .17 | -.50*** | -.16 | .13 | 1.00 | | |
| UNDET. ERRORS | -.16 | -.03 | .14 | -.09 | .30 | -.26 | .26 | -.00 | -.01 | 1.00 | |
| DET. ERRORS | -.05 | .17 | -.18 | -.12 | -.23 | .01 | .20 | -.13 | .06 | -.00 | 1.00 |

*Study XX Objectives.* Study XX sought to investigate the relationship of two new variables -- working memory and anxiety -- to error making. Working memory was included because, unlike personality measures, it was based on a performance measure. Furthermore, working memory has been hypothesized as a major individual differences variable in the cognitive abilities area (Kyllonen & Christal, 1990). Anxiety was included because at least two theories of anxiety would predict greater numbers of errors by high anxious subjects. The first theory states that anxious thoughts consume working memory capacity, and therefore lead to errors through working memory overload (Eysenck, 1983). The second theory states that anxious subjects are less sensitive to peripheral cues than non-anxious subjects (Bacon, 1974; Leon & Revelle, 1985; Wine, 1971). Thus, high anxious subjects may make more errors due to reduced cue utilization.

To replicate the findings of Study IXX that self-report measures of error-making are unrelated to performance measures, the CFQ was retained in the present study.

*Study XX Methods.* Subjects participated in the full number reduction task, with long-term priming to promote errors. Number of detected and undetected errors served as the primary dependent variables from the number reduction task. The following measures were also collected from each subject:

Self Report Measure of Error-Proneness
        *Cognitive Failures Questionnaire* (CFQ: Broadbent, Cooper, Fitzgerald, & Parkes, 1982)

Working Memory Capacity
        *Verbal and quantitative working memory* capacity were assessed with a computerized performance task that required subjects to remember the last three (of several) stimuli while transforming them

in response to cues presented on the computer. Subjects also gave confidence ratings for each trial on a three-point scale (i.e., "I think I got it wrong"; "I am not sure"; and "I think I got it right").

Anxiety Measures

*State Trait Anxiety Inventory* (STAI; Spielberger, Gorsuch, Luschene, Vagg, & Jacobs, 1983) which yields two scores:

State Anxiety: one's current state of anxiety

Trait Anxiety: the trait of being an anxious person

*Test Anxiety Inventory* (TAI: Spielberger, Gonzalez, Taylor, Anton, Algaze, Ross, & Westberry, 1980) which yields two scores:

Worry: the cognitive aspects of anxiety

Emotionality: the physiological aspects of anxiety

Seventy-eight US Air Force recruits from Lackland AFB served as subjects in Study XX.

*Study XX Results.* The correlation table presented below summarizes the results of Study XX. Since almost all errors were undetected, only results for undetected errors are reported in the table. First, as in Study IXX, the CFQ was unrelated to undetected errors. Thus, individuals self reports of error making propensity were not related to their actual error rates on skilled task. Working memory capacity (both verbal and quantitative), however, did show significant and relatively strong relationships with undetected errors.

# Study XX Results

## Correlations of Cognitive Failures Questionnaire, Working Memory Measures, Anxiety Measures, and Undetected Errors

| | CFQ | WM Verbal | WM Quant. | State Anxiety | Trait Anxiety | Worry | Emotionality | Undetected Errors |
|---|---|---|---|---|---|---|---|---|
| CFQ | 1.00 | | | | | | | |
| WM Verbal | -.09 | 1.00 | | | | | | |
| WM Quant. | -.25* | .54*** | 1.00 | | | | | |
| State Anxiety | -.21 | .03 | .23 | 1.00 | | | | |
| Trait Anxiety | -.29* | .01 | .04 | .14 | 1.00 | | | |
| Worry | .13 | .19 | .22 | .15 | .09 | 1.00 | | |
| Emotionality | .22 | .00 | .14 | .09 | .00 | .85*** | 1.00 | |
| Undetected Errors | -.03 | .43*** | .45*** | .05 | -.28* | .25* | .16 | 1.00 |

Subjects with greater working memory capacity made fewer undetected errors on the number reduction task (here working memory was scored in terms of the number of errors made, thus the positive correlation). Finally, anxiety showed a mixed pattern of results. Trait anxiety was negatively correlated: higher trait anxiety led to fewer undetected errors. While this seems puzzling, Eysenck (1983) has hypothesized that trait anxious subjects expend more effort in their performance. This could account for the result. Worry, on the other hand, was positively correlated: higher worry was related to more undetected errors. This is consistent with the notion of negative thoughts consuming working memory capacity (indeed, the

correlation between the working memory variables and worry is in compatible with this interpretation). The relationships between anxiety variables and undetected errors need to be explored in greater detail to test these interpretations. We are currently pursuing further research in this area.

*Conclusion.* Individual differences in the propensity to make errors in skilled cognitive tasks, especially undetected errors, is an important research area for both theoretical and practical reasons. Theoretically, understanding individual differences in undetected error making may shed light on how processes such as composition and proceduralization occur, and what cognitive resources are necessary to accomplish them. Practically, it would be of great importance to be able to decide -- before costly training procedures -- who will, and who will not, be a reliable expert performer at a skilled cognitive task. Air traffic control, medicine, and military applications would all benefit from such screening procedures. Our initial work is hopeful. We have shown that paper-and-pencil tests of the error making propensity do not relate to actual performance. We have also shown that certain cognitive resource measures (i.e., working memory capacity) do relate to performance. Personality measures seem to be helpful only to the extent that they are theoretically implicated, such as anxiety through its potential consumption of working memory capacity. It may be possible to bypass personality measures entirely, if the correct ability measures are chosen.

## General Conclusions

Current theories of error making (Heckhausen & Beckman, 1990; Norman, 1981; Reason, 1990) are stated at a high level of generality, and do not clearly explicate underlying psychological mechanisms. They rely heavily on anecdotal data, and often use the database which served to generate the theory to validate it. We have attempted to rely on current cognitive theory to develop a set of explicit mechanisms for error making -- especially undetected error making -- in skilled cognitive tasks. We have then attempted to validate these using laboratory studies of induced errors. While we have certainly not answered all the questions that could be raised concerning how skilled performance can go awry, we have begun the process of systematically exploring this topic.

Future study in this area needs to address how task demands can affect the likelihood of undetected errors. In addition, research needs to more clearly contrast different forms of memory representation for general processing sequence information. Skill acquisition theories may need to be modified or elaborated to accommodate new evidence on this topic. Individual differences studies need to determine if other ability variables are related to the processes that create undetected errors. Finally, competing theories of skill representation that related to skilled performance errors need to be made explicit (e.g., in computer simulations) so that diverging predictions can be made and tested.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94,* 192-210.

Anderson, J. R. (1989). The analogical origins of errors in problem solving. In D. Klahr & K Kotovsky (Eds.), *Complex information processing* (pp. 343-372). Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Broadbent, D. E., Cooper, P.F., Fitzgerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21,* 1-16.

Bacon, S. J. (1974). Arousal and the range of cue utilization. *Journal of Experimental Psychology, 102,* 81-87.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review, 6,* 345-375.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 484-496.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics, 2,* 153-156.

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Journal of Personality and Social Psychology, 58,* 95-102.

Elio, R. (1986). Representation of similar well-learned cognitive procedures. *Cognitive Science, 10,* 41-73.

Eysenck, M. W. (1983). Anxiety and individual differences. In G. R. J. Hockey (Ed.), *Stress and fatigue in human performance* (pp. 273-298). New York: John Wiley and Sons.

Fitts, P. M., & Posner, M. I. (1967). *Human Performance.* Belmont, CA: Brooks Cole.

Frensch, P. A. (1991). Transfer of composed knowledge in a multi-step serial task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 997-1016.

Heckhausen, H., & Beckman, J. (1990). Intentional action and action slips. *Psychological Review, 97,* 36-48.

Jackson, D. N. (1976). *Jackson Personality Inventory Test Manual.* London, Ontario: Sigma Assessment Systems.

Jackson, D. N. (1989). *Personality Research Form Manual.* London, Ontario: Sigma Assessment Systems

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389-433.

LaBerge, D. (1973). Attention and the measurement of perceptual learning. *Memory & Cognition, 1,* 268-276.

Leon, M. R. & Revelle, W. (1985). Effects of anxiety on analogical reasoning: A test of three theoretical models. *Journal of Personality and Social Psychology, 49,* 1302-1315.

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

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review, 89,* 483-506.

MacKay, D. G. (1987). *The organization of perception and action.* New York: Springer-Verlag.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88,* 1-15.

Reason, J. T. (1977). Skill and error in everyday life. In M. Howe (Ed.), *Adult learning: Psychological research and applications.* London: Wiley.

Reason, J. T., & Mycielska, K. (1982). *Absentmindedness? The psychology of mental lapses and everyday errors.* Englewood Cliffs, NJ: Prentice-Hall.

Reason, J. T. (1990). *Human error.* Cambridge, UK: Cambridge University Press.

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

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.

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

Spielberger, C. D., Gonzalez, H.P., Taylor, C. J., Anton, W.D., Algaze, B., Ross, G.R., & Westberry, L. G. (1980). *Manual for the Test Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P.R., & Jacobs, G. A. (1983). *Manual for the Stait-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76,* 92-104.

Woltz, D. J. (1991). [Instance effects in the procedural skill task]. Unpublished raw data.

Woltz, D. J., Bell, B. G., Kyllonen, P. C., & Gardner, M. K. (in press). Memory for order of operations in the acquisition and transfer of sequential cognitive skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Woltz, D. J., Gardner, M. K., & Bell, B. G. (1995). *Undetected errors in sequential cognitive skills.* Manuscript submitted for publication.

Appendix A
Publications and Submissions

Woltz, D. J., & Madsen, J. G. (under review). Independence in conceptually-driven implicit and explicit memory measures. Manuscript submitted to *Memory & Cognition*.

Larkin, A. A., Woltz, D. J., Reynolds, R. E., & Clark, E. (1995). Relationships of semantic and repetition priming to reading ability differences. Manuscript submitted to *Contemporary Educational Psychology*.

Woltz, D. J., Gardner, M. K., & Bell, B. G. (under revision). Undetected errors in sequential cognitive skills. Manuscript under revision for *Journal of Experimental Psychology: General*.

Woltz, D. J., & Christal, R. E. (under revision). *Detected and undetected errors in working memory.* Manuscript under revision for *Memory & Cognition*.

Woltz, D. J. (under revision). Direct priming of conceptually-driven mental operations. Manuscript under revision for *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

Woltz, D. J., Bell, B. G., Kyllonen, P.C., & Gardner, M. K. (in press). Memory for order of operations in the acquisition and transfer of sequential cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Woltz, D. J. (in press). Perceptual and conceptual priming in a semantic reprocessing task. *Memory & Cognition*.

Woltz, D. J., & Shute, V. J. (1995). The time course of forgetting exhibited in repetition priming of semantic comparisons. *American Journal of Psychology, 108*, 499-525.

Ackerman, P.A., & Woltz, D. J. (1994). Determinants of learning and performance in an associative memory/substitution task: Constraints, individual differences, and volition. *Journal of Educational Psychology, 86*, 487-515.

Gardner, M. K., & Sternberg, R. J. (1994). Novelty and intelligence. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence*. Cambridge, UK: Cambridge University Press.

Gawlick-Grendall, L. A., & Woltz, D. J. (1994). Dominance of meanings for 120 homographs. *Behavior Research Methods, Instruments, & Computers, 26*, 5-25.

Woltz, D. J., & Shute, V. J. (1993). Individual difference in repetition priming and its relationship to declarative knowledge acquisition. *Intelligence, 17*, 333-359.

Appendix B
Participating Professionals

Dan J. Woltz, Ph.D., Assistant Professor, Co-Principal Investigator
Michael K. Gardner, Ph.D., Associate Professor, Co-Principal Investigator

Janet G. Madsen, Graduate Student (Ph.D.)
    University of Utah Graduate Research Fellowship, 1993-1994
    Dissertation Title: Changes in Access to Declarative Knowledge as Skill Level Increases
    Defense Date: March 31, 1995
Brian G. Bell, Graduate Student (Ph.D.)
    WPA Student Research Award, 1993
James Farnham, Graduate Student (Ph.D.)
    University of Utah Graduate Research Fellowship, 1993-1994
Susan Curtis, Graduate Student (M.S.)
Angelique Scharine, Graduate Student (M.S.)
Leslie Bailey, Undergraduate Student (B.A.)
Stacey Lundberg, Undergraduate Student (B.A.)

## Appendix C
## Papers and Presentations

Gardner, M.K., Bell, B.G., & Woltz, D.J. (June, 1995). *Individual differences in undetected error making in skilled performance.* Paper presented at the annual meeting of the American Psychological Society, New York, NY.

Woltz, D.J., Larkin, A.A., Reynolds, R.E., & Clark, E. (April, 1995). *Individual differences in semantic and repetition priming and their relationship to children's reading ability.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Madsen, J. G., & Woltz, D. J. (April, 1995). *Changes in access to verbal knowledge as procedural skill develops.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Woltz, D. J., Gardner, M. K., Bell, B. G., & Farnham, J. M. (July, 1994). *Memory for processing sequences can produce undetected errors in skilled performance.* Paper presented at the annual meeting of the American Psychological Society, Washington, D.C.

Woltz, D. J., Gardner, M. K., & Bell, B. G. (April, 1994). *Knowledge compilation and undetected errors by skilled performers.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Woltz, D.J. (April, 1994). *Implicit memory for semantic processes.* Invited address to the Psychological Services Division, Salt Lake Veterans Administration Medical Center, Salt Lake City, UT.

Gardner, M. K., Woltz, D. J., & Bell, B. G. (August, 1993). *The costs of expertise: Undetected errors in cognitive skill acquisition.* Paper presented to US Air Force Armstrong Laboratories, Brooks AFB, San Antonio, TX.

Woltz, D. J. (June, 1993). *Individual differences in implicit and explicit memory for procedural sequences in skill acquisition.* Paper presented at the annual meeting of the American Psychological Society, Chicago, IL.

Bell, B. G., Woltz, D. J., & Madsen, J. G. (April, 1993). *The role of memory for procedural sequences in skill acquisition.* Paper presented at the annual meeting of the Western Psychological Association, Phoenix, AZ.

Appendix D
New Discoveries, Inventions, and Patents

None