INFORMATION TO USERS

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

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

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

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

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

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 1355190

Transfer of hypothesis testing strategy in fault diagnosis

Dammon, Charles T., M.A.

Rice University, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

RICE UNIVERSITY

TRANSFER OF HYPOTHESIS TESTING STRATEGY IN FAULT DIAGNOSIS

by

CHARLES T. DAMMON

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

APPROVED, THESIS COMMITTEE:

David M. Lane
Associate Professor of Psychology,
Director

Kenneth R. Laughery
Professor of Psychology

Randi C. Martin
Professor of Psychology

Houston, Texas

April, 1993

ABSTRACT

TRANSFER OF HYPOTHESIS TESTING STRATEGY IN FAULT DIAGNOSIS

by

CHARLES T. DAMMON

Transfer of training is the degree to which a skill or principle learned in one environment can be applied in another environment. Most research that demonstrates transfer relies on the use of hints or explicit instructions identifying the applicability of the information learned in training to the test task. Critics charge that this is not really transfer at all, but simply following instructions. The research reported herein describes an efficient means for testing hypotheses in a fault diagnosis task that, although it would appear to be an obvious strategy, requires an extremely simple training task for subjects to detect. Subjects in Experiment 1 apply the learned principle to a slightly more complex but similar problem, demonstrating near transfer. Subjects in Experiment 2 apply the principle in a completely dissimilar task, exhibiting far transfer.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

People are involved with complex systems on a daily basis. In fact, they employ such systems so regularly that they have become almost second nature. People drive to work in automobiles without thinking of the complex control and power components that are necessary to make the system function properly. They turn on the light switch in offices and expect the room to be illuminated, but never consider the incredibly complex systems necessary to produce electricity and deliver it in a usable form.

The only time people consider the individual components of many of these complex systems is when their function is impaired. Then they apply diagnostic reasoning, a special instance of problem solving, to define, locate, and repair the malfunction. Often, though, there are several possible causes of a system malfunction and one must search through these possible causes until one locates the "bug," or error. During this search for the error, people are plagued by all the typical biases that accompany cognition (i.e. recency, availability, etc.). Sometimes the search is hindered because people look for a very complex reason for system malfunction while the real answer is quite simple.

A computer user, for example, may experience difficulty in printing a document. The user issues the command and nothing happens. Perhaps the user then refers to the manual to make sure she used the correct command sequence. She might then examine the system files on the computer to try to determine why the document will not print. Finally, the user contacts the computer company's help desk by telephone. The company consultants retrace the steps in creating and printing the document and find nothing wrong with the sequence of steps that the user followed. In disgust the user decides to break for lunch. Upon her return, she notices that the printer is unplugged. Checking to make sure the printer is plugged in is a very inexpensive measure (in terms of time and effort), and ignoring such a simple hypothesis can prove quite costly.

Even more striking is what can happen when easy to test, simple error hypotheses are ignored in an environment where the cost may be more than just a few wasted hours. Modern aircraft engines used in small (single-engine) airplanes rarely fail due to mechanical problems. Each year, though, several "engine out" incidents occur when pilots fail to test simple or easy to disconfirm hypotheses but instead look for complex possible errors when engine problems occur. For example, the fuel tanks in small airplanes are in the wings and fuel is supplied to the engine by gravity (in high wing aircraft), fuel pumps (in low wing aircraft), or some combination of both. Located in the cabin is a three position selector lever to select either the right fuel tank, the left fuel tank, or to shut off fuel (in case of fire or for service procedures). When a particular tank runs low, the pilot must manually select the other tank. Pilots are trained extensively in fuel management procedures that involve proper planning to insure the correct amount of fuel necessary for a specific flight and selection of the proper fuel tank at the right time. Many accidents, however, occur because when the engine stops pilots test all sorts of complex hypotheses about the ignition system, ice in the carburetor, etc. and forget to simply check which fuel tank is selected. Indeed, some aircraft involved in accidents caused by fuel starvation where pilot and passenger are killed have been found to have one tank empty and one full and the fuel selector lever firmly pointing to the empty tank.

The research reported here examines how people select error hypotheses for testing in a diagnostic task. The emphasis is on developing and training an efficient strategy for selecting error hypotheses for testing and on transferring that strategy to different types of diagnostic problems. To introduce the research, relevant literature in problem solving (including the Gestalt approach and the information processing approach characterized by the research of Newell and Simon), hypothesis testing, and transfer of training is reviewed. The application (or lack thereof) of some of the findings in these various traditions are also examined. Two experiments are presented which indicate that efficient hypothesis selection behavior can be trained and that this training can be transferred to new and different domains are presented. Finally, the findings of these two experiments are discussed in the light of the challenges inherent in applying diagnostic problem solving.

Problem Solving

The Gestalt Tradition.

One of the earliest influences on problem solving theory was that of the Gestalt psychologists. The approach of the Gestaltists was characterized in two ways. First, Gestalt psychologists saw problem solving as a discontinuous, sequential process with distinctly defined components. These include the components of preparation, where the solver has recognized that a problem exists and some preliminary attempts at understanding and solving the problem have been made, incubation, a period where the problem is set aside for awhile and is not examined consciously (work proceeds at some unconscious level), illumination, that famous flash of insight that ends the unconscious work and brings the solution to the forefront of consciousness, and verification, in which the solution discovered at insight is tested and confirmed (Wallas, 1926). Secondly, the Gestaltists insisted that the problem must be seen as a whole, rather than as individual segments. Just as they believed that perception involved arranging the separate elements of the visual field into a coherent image, so problem solving involves rearranging and recombining the various elements of a problem until a stable configuration, or Gestalt, emerges.

Wolfgang Kohler (1927) studied the problem solving skills of apes. Particularly adept was an ape called Sultan, who learned to use a long pole stuck between the bars of his cage to reach a banana that was placed at a distance. In order to make the problem more difficult, Kohler gave Sultan two shorter poles, neither of which could reach the banana alone. After sulking at his new failure to reach the banana, Sultan suddenly joined the two poles into one longer one and reached the banana. Kohler asserts that this was an exhibition of insight, a sudden solution to a problem by discovery. According to Gestaltists this "insight" occurred after the period of incubation (while Sultan was sulking). Several studies (Fulgosi & Guilford, 1968; Silveira, 1971) found evidence of this incubation period and flash of insight in humans, as well. On the other hand, some studies (Dominowski & Jenrick, 1972; Murray & Denny, 1969) actually showed decrements in problem solving following an interruption.

Anderson (Neves & Anderson, 1981) offers some help in interpreting these discrepant findings in the context of another Gestalt idea known as "set." Set refers to the tendency of people to perceive objects and events in their environment in ways that are greatly influenced by prior experience and the expectations that result from that experience. In other words, our perceptions are predetermined (or "set") by our prior experiences. It may be that set is an accurate explanation for what Gestaltists called Einstellung, or "psychic blindness," the inability to see a solution that is incredibly obvious or simple. According to Anderson, when we begin to solve a problem, our prior knowledge is a resource that can be activated or called upon to propose at least a framework of procedures useful in solving the problem. If our set is appropriate to the problem at hand, we will produce effective behaviors that propel us toward a solution. On the other hand, if our set is inappropriate, we may produce ineffective behaviors that do not lead to a solution. In regards to incubation, then, Anderson would say that if we had appropriate set to begin with, incubation may impair performance because our set may change during incubation. However, if our set was inappropriate in the first place, incubation may enhance our performance by allowing us to change to an appropriate set for the problem at hand.

The phenomenon of set was examined extensively by Luchins (1942). He studied the effect using several tasks (maze tracing, geometry proofs), but the most well-known task is the water jugs task. Luchins showed subjects three different sized jugs with clearly marked capacities. The problem is to measure out an amount of water using some or all of the jugs. Subjects were not given any other measuring device besides the jugs themselves. There were two methods shown to subjects for measuring out the correct quantity of water, a long and general method that takes four steps and uses three jugs, and a shorter method that takes two steps and uses only two jugs.

After being shown both methods, one group of subjects was given practice problems in which they could only use the long method. They were then presented with a problem that could be solved by using either method (the long or the short). Most subjects in this condition used the long method. Another group of subjects was also shown both methods but did not recieve the practice problems using the long method. When they were presented with the problem that can be solved using either method, most of them use the short method.

Lewis (1978) showed further evidence for the Einstellung-like effect of set by giving subjects practice with rules in a symbol replacement task. Later, he presented a new rule that would allow subjects to save time in some cases. Lewis' results indicate that the more practice subjects had with the original rule set, the less likely they were to take advantage of the new "shortcut" rule.

The Gestalt view of problem solving as a discontinuous process is based, in part, on the idea that true creative or insightful thought (such as occurs when the solution to a problem becomes suddenly clear) is not predictable from previous behavior or thought, because the insightful solution comes as a break in any continuous stream of problem solving. Katona (1940), however, shows clearly that subjects who are given hints on solving problems which help them to understand the structure of the problem are more likely to solve similar problems in the future than are subjects who only memorize the sequence of steps needed to solve the problem. In other words, previous behavior (learning something meaningful about the structure of the problem) predicts problem solving ability.

In an earlier study of reasoning, Maier (1931) suspended two strings from the ceiling and had subjects attempt to tie them together. The strings were sufficiently far apart that subjects could not grasp one and reach the other, but were sufficiently long that they could, indeed, be tied together. In the room with the strings were several objects, one of which was a pair of pliers. The proper solution to the problem is to tie the pliers to one of the strings, use it as a pendulum to induce regular swinging, grasp the other string, and catch the swinging string on its nearest swing, and tie them together. This novel use of the pliers, though, was only found by 39% of Maier's subjects in a ten minute period.

Maier found that some hints would increase subjects' ability to solve the problem. One hint involved a confederate who walked through the room and subtly brushed one of the hanging strings, making it swing back and forth. If this hint failed to produce a correct solution, the subject was handed a pair of pliers and told "With the aid of this and no other object, there is another way of solving the problem." Maier divided the subjects who successfully solved the problem into two groups. One group consisted of those who solved the problem as a whole ("The solution just came to me. . . "). This group failed to report that the hint was useful in finding a solution to the problem. The other group were those who went through a sequential series of steps to solve the problem. This group reported (with one exception) that the hint was quite useful. More interesting, though, is the finding that even the group who indicated that the hint was useless were more likely to solve the problem than subjects that did not receive a hint at all. Obviously, then, a discontinous process in solving this problem is unlikely, for information received during the solving process provided more help in discovering a solution than did some flash of insight (even though subjects were unaware of the usefulness of such information).

The Information Processing Approach.

Newell and Simon (1972) proposed new methods for the study of problem solving. They depended heavily on verbal protocol in which subjects verbalized their thoughts as they solved problems. By examining these verbalizations, cognitive processes could by identified and observed. The core of Newell and Simon's assertion was that humans are information processors. In other words, we are a system that receives input, organizes and processes that input in a symbolic manner for some objective, and produces an output based on the internal processing of this symbolic information. As is intuitively obvious to us today, this model is consistent with the digital computer and, indeed, Newell and Simon relied heavily on computer modeling for their investigation.

A major contribution of Newell and Simon to the general direction of problem solving research is their characterization of problem solving as a search for a solution across a problem space (1972, p. 809). This problem space contains conceptual nodes which are similar to nodes in semantic network theories of memory. These nodes represent *states of knowledge* which include all the information the subject knows about a problem at any moment. The solver moves between nodes or states of knowledge by means of *operators*. Each time the solver applies some operator, he or she changes the state of knowledge. Newell and Simon assert that problem solving consists of moving through the nodes of the problem space, continually searching various states of knowledge. One of the main tasks of problem solving is finding particular operators that are effectual in a given problem, in other words that refine the state of knowledge to a point at which the solution to the problem is clear.

Two major determinants of a problem solver's effectiveness are the quality or accuracy of the problem space (similar to the Gestalt psychologists emphasis on proper representation of the problem), and the mode of search. The problem solver is conceptualized to move through the problem space by searching for a solution path or series of knowledge states that lead to a solution. The search is generally forward moving from some initial knowledge state to a goal state (solution of the problem). Two heuristics that have been identified as search strategies are subgoal analysis, which involves identifying parts of a problem that satisfy or achieve intermediate goals and, when compiled, lead to achievement of the overall goal (Wickelgren, 1974, p. 91) and means-end analysis (Newell & Simon, 1972), which is an analysis of the differences between the current and desired states and includes a suggestion of some action which will reduce the discrepancy between these states.

Hypothesis Testing Theory

Krechevsky (1932) performed early research in hypothesis testing on rats performing a maze learning task. The accepted theory of Krechevsky's day held that the rats simply were making stimulus-response connections in order to collect the food reward. Krechevsky, though, found systematic variations in the errors made by the rats, leading him to theorize that rats actually propose (or hypothesize) solutions to the maze and reject these solutions when they prove incorrect. Spence (1945), interpreted Krechevsky's theorized "rat hypotheses" in terms of associations of stimuli and response. As excitatory tendency crosses some threshold at which it overshadows inhibitory tendency, the rat exhibits the discrimination predicted by Krechevsky's hypothesis model. Harlow (1950), however, found similar behavior in monkeys while examining the errors they made in sensory discrimination tasks. Harlow's conclusion was that the errors were not made by chance or random behavior, but rather that there were systematic factors which resulted in the errors. One of these systematic factors was a tendency on the part of the monkeys to explore or test new stimuli in a discrimination problem. The systematic nature of the errors may be descriptive of hypotheses being tested and rejected.

Hypothesis testing theories were later employed to examine how people form and use concepts (Levine, 1966; Lane, McDaniel, Bleichfield, & Rabinowitz, 1976). In representative tasks, subjects are presented with stimuli which vary along several dimensions. Stimuli may be presented one at a time in some succession or simultaneously in pairs or groups. If the stimuli are presented singly, subjects are instructed to report whether the presented stimulus represents the concept under investigation. Hypothesis testing theory would suggest that people test hypotheses about the stimuli in question. In other words, subjects might hypothesize that the concept is "red images" and would respond affirmatively to stimuli if, indeed, they were red. Subjects are theorized to have some group, or universe, of hypotheses from which to sample. As hypotheses are disconfirmed by feedback received from the experimenter, this universe of hypotheses is narrowed and refined.

According to Levine (1971, 1974), subjects form domains, or subsets, of related hypotheses within a universe of hypotheses. One domain might contain hypotheses about features of the stimuli (size, shape, etc.), while a second domain might contain hypotheses about presentation of stimuli (i. e. order). Levine's concept of domain allows him to make three assumptions.

First of all is the "transfer hypothesis" (Levine, 1974, p. 271) which states that subjects "infer from the first n solutions the domain within the universe from which the (n + 1)th solution will be taken. He will start the (n + 1)th problem by sampling hypotheses from this domain." In other words, subjects who have been trained to sample from a particular domain for solutions to problems will, more than likely, sample from the same domain on the next problem. Second is the "empty-set assumption" (Levine, 1974) which indicates that a subject will continue to sample from a particular domain until it contains no more hypotheses. When every hypothesis in the domain has been tested, subjects may either transfer to a new domain or resample hypotheses from the current domain. Finally, Levine (1974) postulates the "infinite-set assumption", which implies that, since a subject will not change domains until he has sampled every hypothesis in a particular domain, if he is sampling from an incorrect domain (one which does not contain the relevant hypothesis) that is infinitely large, he will never solve the problem. Fingerman and Levine (1974) showed that subjects who were trained to solve a series of complex position-sequence problems had difficulty then solving simple discrimination problems. Lane et al. (1976) further reported that subjects who were led to sample from a domain of complex hypotheses were unlikely to transfer to a simpler domain. These findings are related to Luchins' (1942) study of the Einstellung effect.

To illustrate the assumptions of Levine's Hypothesis Testing Theory, and the paradigm used to investigate these assumptions, consider a study in which Levine (1971) prepared a deck of cards with half the cards showing the letter "A" on the left side of the face and the letter "B" on the right side. The other half of the cards were just the opposite (the "A" was on the right side of the face and the "B" was on the left). On a given trial, a series of cards was presented to the subject in random order, and the subject was to respond by verbalizing one of the letters on the card (say "A" or say "B"). For each card presented, the subject was told either "correct" or "incorrect." Subjects were presented a series of six problems at training in which the correct solution was from a domain of position-sequences (left-right, left-right, etc.). After training, a problem was then given where the solution was from a much less complex domain, either "always say A," or "always say B." Eighty-percent of Levine's subjects failed to find the correct solution to the test problem within 100 presentations. The simple problem would usually be solved immediately if given initially (without training in the complex domain).

Levine's findings bring to mind the examples presented earlier of fault diagnosis procedures in which the people involved overlook extremely simple hypotheses in favor of much more complex ones. In the case of the computer printer that was inoperative, the subject began sampling complex hypotheses of programming errors, command misuse, etc., when a very simple check of the connecting cables would have solved the problem. In the description of fuel starvation accidents in aircraft, the pilots, just like the computer operator, test complex hypotheses, failing to look at the simple solution of proper fuel tank selection.

Fault Diagnosis

One special instance of problem solving is the detection, location, and correction of faults in various types of systems. In the domain of computer programming, this process is called debugging. In technical systems such as electronic circuits, process control, etc., it is commonly known as trouble-shooting. In medicine it is, of course, diagnosis.

Dunn (cited in Ashby, 1988) stated that between 1980 and 1990, the Defense Department's annual expenditure for weapons software would grow from three billion to 30 billion dollars. While the end of the Cold War and a scaling down of military forces may reduce Defense Department software expenditures, the volume of software being produced and modified every year is stunning. Brooks' (1972) examination of the software development process indicates that debugging, in particular, comprises a major portion of the total cost of software development. Approximately one half of the time spent developing a major software project is spent debugging. Post-development maintenance costs equal fully 40% of the total development cost. Moreover, correcting a bug in software incurs a 20%-50% chance of generating another bug. The time spent in locating and correcting these additional bugs geometrically expands the time associated with debugging.

In the field of medicine, fault detection (diagnosis) procedures are often costly, usually uncomfortable, and sometimes pose physical risk to the patient. Diagnostic tests constitute a significant amount of medical spending. In Germany, 15% of all income for physicians is from laboratory testing (Schicke, 1983). Similar costs in terms of both money and time can be found in the literature for other domains of fault diagnosis.

With the cost of diagnosis so blatantly obvious, it is disturbing that programmers are seldom taught debugging as a specific, trainable skill. A survey of computer science courses listed in various college catalogs produces no evidence of specific courses on debugging. Further, examination of programming textbooks finds debugging relegated to single-chapter status if that much. Informal discussions with programmers and medical doctors indicate that debugging, like medical diagnosis is seen as a sort of "black art" that is not trained, but rather is developed from an inherent talent with experience.

Much of the research in fault diagnosis, in both the computer debugging field (Brooks, 1983; Holt et al, 1987; Nanja & Cook, 1987; Stone, et al., 1990; Vessey, 1985; Vessey, 1986) and in the medical diagnosis field (Balla, 1980; Elstein et al, 1978; Lesgold, 1984), has looked at the differences between novices and experts in an attempt to understand how people solve problems of fault diagnosis. Some of the advantages of experts include the ability to use domain-specific knowledge to generate higher quality hypotheses, clearer identification of relevant information in a program or in a patient's condition when developing hypotheses, etc. In other words, experts perform better because they are experts. Though that statement may seem quite cynical, little application has been made of problem-solving theory in the diagnosis literature. Rather, a sort of atheoretical description of behavior has become the norm.

Some information about the process of diagnosis, however, can be carefully gleaned from the research reported. Littman, Pinto, Letovsky, and Soloway (1986) found that programmers' mental representation of the program and the fault were of critical importance in predicting successful debugging procedures, just as would be suggested by Gestalt psychologists. Nanja and Cook (1987) found that programmers do generate and test sets of hypotheses about programs in order to locate the fault or bug in similar fashion to the hypothesis domains postulated by Levine. Gugerty and

Olson (1986) and Brooks (1983) reported that experts had a superior ability to begin testing appropriate, relevant hypotheses (in other words, they were more likely to sample from the correct domain). Similar findings about the quality and organization of hypotheses generated in medical diagnosis are reported by Balla (1980), Elstein et al (1978), and Lesgold (1984).

Although domain specific knowledge may help ensure the generation of high quality hypotheses and the selection of an appropriate hypothesis domain from which to sample, it can do nothing to help streamline the actual selection of hypotheses from a particular domain for testing. The most efficient method for sampling hypotheses for testing is to select the simplest (easiest to disconfirm) first and move in order of complexity to the most difficult.

It might be supposed that probability based on prior experience should weight each particular hypothesis for order of testing, and that more probable hypotheses should be tested first. However, unless the differences in probability are quite large, the strategy of simple to complex still results in more efficient hypothesis testing over several diagnosis episodes. If the hypotheses are all of high quality (as are generated by experts), then probability may be difficult to distinguish. In addition, Eddy (1982) showed that expert physicians often vastly overestimate the probability of a particular diagnosis due to violation of a fundamental axiom of Bayesian statistics by ignoring prior probabilities and false positives, instead basing their diagnosis solely on the true positive rate of diagnostic tests.

Ashby (1988) controlled probability by making it constant. Subjects were given a set of algebra equations, one of which was incorrect (to represent an error or bug). Subjects were told that each equation (or hypothesis) had an equal chance of being the bug. Subjects were presented with five equations of varying complexity (the hypothesis domain). They selected hypotheses to test and were then shown the values for each variable in the equation. From this they could determine if the equation was correct or incorrect. If subjects were using the optimal strategy, then they would select the simplest or easiest to test equation first, then move to the next simplest, and so forth. Interestingly, only one of Ashby's sixteen subjects began the task by using the optimum strategy, and, although most subjects learned the strategy in twenty trials, almost one-third of the subjects were still not using the optimum strategy by the twentieth trial.

In a second experiment, Ashby modified his task by introducing the possibility of a second buggy or incorrect equation among each group of five. Again, only one of the nine subjects in this condition started out using the optimum strategy. In this only slightly more complex condition, however, after twenty trials there was still only one subject who was using the optimum strategy (it was the same subject who began by using this strategy).

It is clear, then, that in a very simple task, people can learn to use the optimum strategy for hypothesis selection in a diagnosis task, but in an only slightly more complex task, people fail to learn the optimum strategy. Experiment 1 of the present research first attempts to replicate Ashby's (1988) findings, and then addresses the question of whether the learning exhibited by subjects in the simple one bug condition will transfer to the similar but slightly more complex two bug condition. Experiment 2 examines the transfer of this learning (should learning actually occur) to a new and different task.

Transfer of Training

One very important factor in learning new skills is the extent to which prior knowledge or skill development can be applied to learning the new skill or to applying a learned skill in a novel situation. This has been identified as transfer of training (Singley & Anderson, 1989). Transfer of training is typically measured by comparing performance between a control group which learns the target task in its normal setting and a transfer group which is given some practice in a training task. A measure called the transfer effectiveness ratio (TER) developed by Povenmire and Roscoe (1973), expresses the relative efficiency of transfer as a ratio of the amount of savings incurred by the transfer group to the time spent by the transfer group in training. A simpler measure, of course, is obtained when the question is simply whether training produces some arbitrary level of proficiency in the task.

According to Baldwin and Ford (1988), a great deal of the reported research on transfer focuses on ways of improving training programs by the application of general learning principles. Two of the most often considered principles are the concept of identical elements and the teaching of general principles. Morris and Rouse (1985) reviewed studies concerning the effects of different types of instructional strategies on transfer. One finding was that training programs that highlight the learning of procedures instead of theory are more efficient in terms of training time to learn the skill, but that the learned skill does not generalize well to other situations. On the other hand, strategies that emphasize the training of theory, while less efficient initially, provide long-term benefits in the form of greater transfer to similar tasks. The research reviewed by Morris and Rouse, however, is replete with inconsistent operational definitions, which means that many training programs reviewed mix procedural and conceptual information in unusual ways. Therefore, the results of this research must be accepted with great care.

Kieras (1987) emphasized the importance of training individuals in a deeper understanding of the technological systems with which they work, a knowledge of both functional and structural properties. He asserted that understanding both "how" and "what" about a task facilitates the development of an accurate mental model of the task which, in turn, increases the likelihood of transfer. Similarly, Anderson (1981, 1983, 1987; Singley & Anderson, 1989) proposed a two-stage theory that explained problem solving skill as beginning with declarative knowledge followed by procedural knowledge. The fact that a more general declarative knowledge base precedes a task-specific procedural knowledge base allows for a flexible underlying knowledge domain that is accessible to several specific problem solving methods.

The identical elements principle identified by Baldwin and Ford (1988) corresponds to the idea of fidelity in training situations. That skill transfer is based on a certain degree of similarity between the training task and the actual task is an idea that has prevailed since Thorndike's (1903) law of identical elements. The main idea of this law is that transfer between two tasks is, to a great degree, dependent on the common components shared by the two tasks. This idea of fidelity is often applied to justify the design and construction of elaborate training simulators and has, in fact, dominated the discussion of motion and visual systems for flight simulators (Comstock, 1984).

Cognitive justification for the high-fidelity concept is provided by the instance memory theory of skilled behavior proposed by Logan (1988). According to this theory, every occurrence of a skilled event creates an episodic memory trace. The more memory traces available for a particular skill, the more likely an individual is to accurately access a valid memory trace, and the more skilled is the performance. If, indeed, there is such a one to one correspondence between number of applications of a skill and performance of that skill, then the obvious implication is that for transfer to occur, the training task should be extremely similar to the target task. Logan does (1990, cited in Lintern, 1991) concede that some dimensions of fidelity have no effect on transfer, but rather that the components of the task that are attention-demanding are the ones which benefit by fidelity.

Another condition under which transfer has been shown to occur is in the presence of an explicit "hint." Weisberg, DiCamillo, and Phillips (1978) presented subjects with Duncker's (1945) candle problem which asks solvers to attach a candle to a wall given several unrelated objects. The solution is to empty a box full of tacks, attach the box to the wall using a tack, and then place the candle in the box. Unlike Duncker, however, Weisberg et al had subjects memorize a list of paired associates (one pair of which was candle-box) presented as words or objects prior to presentation of the problem. One group of subjects (the hint group) was told that one pair was relevant to the solution of the problem, while another group (the no-hint group) recieved the paired associates with no mention of relevance. A control group was not presented with the paired associates. The hint group performed better than the control group, but the no-hint group did not. Therefore, presentation of the paired associates was only effective in assisting a solution when accompanied by an explicit hint.

Gick and Holyoak (1980) used another of Duncker's (1945) problems, in which a cancerous tumor threatened a patient's life, but so did the massive X-ray required to destroy the tumor. The solution was to attack the tumor with numerous low-intensity X-rays from several different directions simultaneously. Prior to presentation of the problem, Gick and Holyoak's subjects were given a story comprehension task in which they read an analogous story about a general who captures a fortress without destroying it by sending small troops of his army along several different roads which converged upon the fortress. When subjects were told that the previous analogue might be useful in solving the radiation problem, they were more likely to solve it (transfer occurred) than when they were given no hint about the usefulness of the analogue. Even if experimenters explicitly stated the principle behind the analogue, subjects without the hint that one story was related to the other were unlikely to solve the radiation problem.

Detterman (1993) finds little evidence for transfer in the literature, but instead claims that studies that have found transfer are fraught with methodological problems. Much of the problem, he claims, is with the definition of transfer itself. Detterman asserts that if any explicit instruction or hint is given to subjects concerning the applicability of the training information to the test situation, then true transfer does not occur. He further asserts that systematic bias introduced by the fact that experimenters are often not blind to the conditions of their subjects contaminates results. This, of course, assumes that measures are not objective enough to preclude subjective interpretation by the experimenter.

In the present study, Detterman's concerns were addressed by eliminating any hint or instruction to subjects that might draw their attention to the relationship between training and test problems. In addition, measures obtained were created to be objective so that the experimenter did not have the latitude in recording to favor one condition over another. Even while controlling for the major factors that Detterman decries in transfer studies, this research finds that transfer can and does occur when people learn efficient hypothesis testing strategy and are given the opportunity to apply that strategy to a different problem.

Experiment 1.

Ashby's (1988) finding that subjects presented with sets of five algebra statements, one of which is incorrect, learn an efficient principle for selecting hypotheses to test (test the simplest first), but fail to learn that principle in an only slightly more complex task (the addition of a second incorrect statement in the set of five) is interesting in several ways. First, the fact that this strategy or principle is a skill that is learned and not inherent (or some sort of "common sense") is underscored by the fact that only two of Ashby's subjects began applying the principle on the very first problem and, indeed, over half of the subjects used the principle less than 50% of the time. Ashby's subjects were, after all, bright undergraduate students at a very selective university. Second, while the addition of a second incorrect algebra statement does not seem to be a very complex manipulation, it is sufficiently complex to prevent subjects from ever learning to apply the principle. On almost half of the problems (9 of 20) in the two-bug condition, only one subject applied the "simple hypotheses first" principle.

One goal of Experiment 1 in the current study, then, was first of all to replicate Ashby's rather unexpected finding. Is such a seemingly obvious strategy as starting with the simple hypotheses and moving to the more complex really that hard to learn? Another goal is to determine if the principle, once learned in the one-bug condition, can be transferred to the slightly more complex two-bug condition. The hypothesis examined by this experiment, therefore, is that subjects will learn to apply the "simple first" principle in a one-bug condition, but not in a two-bug condition, and that after learning the strategy in the one-bug condition, they will transfer it to the two-bug condition.

METHOD

Subjects.

Subjects were 24 Rice University undergraduate students who participated for required credit in undergraduate psychology courses.

Materials.

The experiment was programmed in Hypercard and run on an Apple Macintosh LC.

Procedure.

Twenty four subjects were randomly assigned to either an experimental or a control group. The experimental group received 40 sets of algebra statements with five statements in each set to represent five hypotheses in a specific hypothesis domain (see Fig. 1). They were asked to find the one statement in each set which was incorrect. Subjects were told that each statement had an equal probability of being the incorrect one (indeed, the "bug" was placed randomly in each set by the computer). A statement was selected for testing by clicking the mouse button on the "test" button to the left of each statement. The selected statement was then presented with a set of values for the variables and with a "correct" and "incorrect" button (see Figure 2). Subjects clicked on the appropriate button and were either transferred to the next set of five statements (if this was indeed the "bug") or returned to the original set to continue sampling. A check mark was placed by previously sampled statements (hypotheses) as a bookeeping aid so that subjects would not feel compelled to test sequentially to avoid retesting any particular hypothesis. After 20 sets, the subjects in the experimental group were given a ten-minute break during which they were encouraged to get a

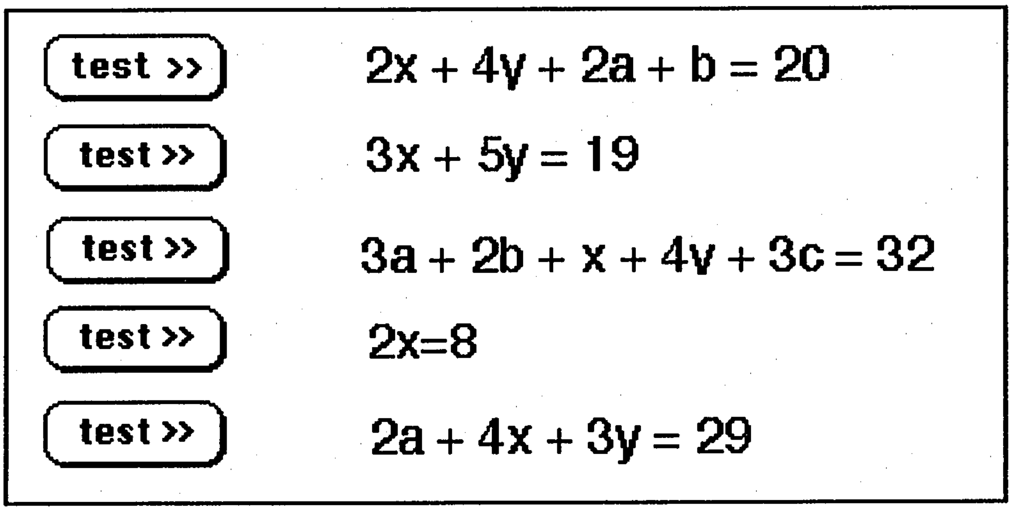

Figure 1. Hypothesis selection screen.

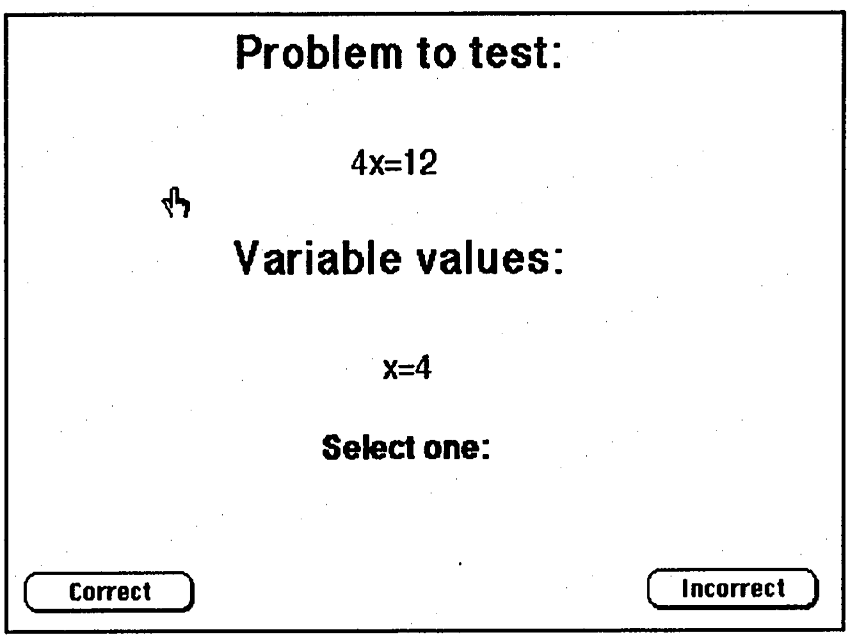

Figure 2. Hypothesis testing screen.

drink, go to the rest room, go outside and relax, etc. At the end of the break they were presented with another 20 sets of five statements each, only this time they were told that there was a possibility that two incorrect statements were in each set and that the program would only transfer them to the next set after both "bugs" had been found. The control group went through a similar procedure, only both groups of 20 (before and after the break) contained two "bugs" in each set of five statements. So training for the experimental group consisted of sets of statements with just one incorrect statement, while test consisted of sets of statements with two incorrect statements. Training and test for the control group both consisted of sets of statements with two bugs.

Results.

To simplify scoring, each set of 20 problems was divided into five blocks with four problems in each block and a point was given each time the subject used the optimal strategy. There were, therefore, four possible points to be earned in each block. To be considered an optimal strategy, a subject must test the statements in order from least to most complex until the "bug" was found.

The main findings are graphically represented by the box plot in Figure 3. Ashby's results were indeed replicated. As can be seen from the difference between Experimental and Control groups in the training problems, subjects in the two-bug condition were poorer at choosing the optimal strategy than were subjects in the one-bug condition. Further replicating Ashby and Lane, subjects in the two-bug condition showed no signs of learning to use the optimal strategy even over a large number of problems.

Figure 3 also reveals that experimental subjects had no trouble trans-

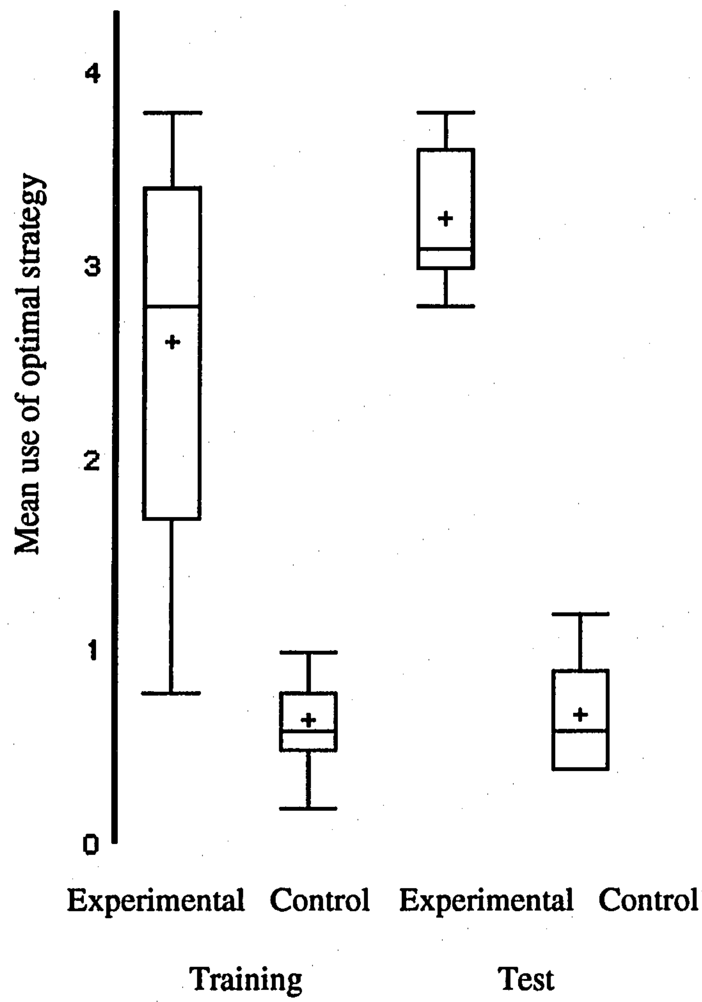

Figure 3. Box plot identifying mean number of times subjects used the optimal strategy (per block of five trials) as a function of group and task.

ferring their use of the optimal strategy from the one-bug to the two-bug conditions. Their increased score on the test problems means that optimal strategies were more common in the two-bug condition than in the one-bug condition. The results of statistical analyses are consistent with the above description of the findings. The experimental group (which received the one-bug treatment at training and the two-bug treatment at test) performed significantly better overall than the control group (which received the two-bug treatment in both sets of twenty problems), $F_{(1,22)} = 159.44$, $p < .01$. The interaction between Group (control vs. experimental) and task (training vs. test) was also significant, $F_{(1,22)} = 4.96$, $p < .05$. The means for both conditions and both sets of twenty problems can also be seen in Figure 3.

In order to validate the assertion that using the strategy of selecting the least complex hypothesis for testing first is, indeed, the most efficient strategy, time measures to complete the task were also obtained. The experimental group upon transferring the optimum strategy acquired in training to the test problems, completed test problems significantly faster than the control group, who never developed the optimal strategy, $F_{(1,22)} = 12.98$, $p < .01$.

A second time measure, average time to solve a three term algebra statement in the training problems, was taken to explore the effect of individual problem solving proficiency on overall performance. Examination of the regression plot (Figure 4) obtained by predicting total time to complete the test problems by this intermediate time measure reveals an interesting difference between the slopes of the regression lines for the groups. Figure 4 indicates that as computational speed (as defined by subjects' average time to complete a three-term problem) decreased, so did time to com-

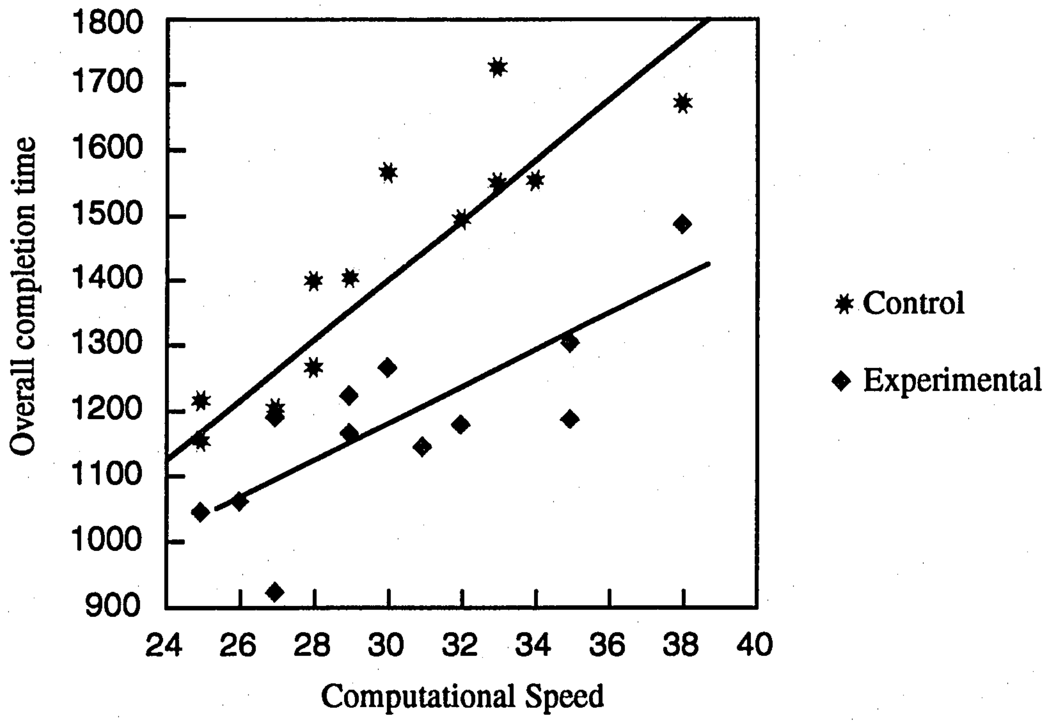

Figure 4. Overall time to complete task as a function of computational speed.

plete the entire set of problems. That overall performance speed is closely related to computational speed is not surprising. However, this increase in performance occurred more sharply for the control group. This occurs because the control group, using the inefficient strategy, tested more hypotheses before finding the bugs, so computational speed becomes more valuable in the same way that a runner's superior speed opens a wider lead for him in a long race than in a short one.

The scatterplots in Figure 4 indicate the high degree to which performance in terms of total time to complete the task can be predicted by computational speed and group membership. Indeed, the effect of these two variables individually and interactively accounts for 83.6 percent of the variance in overall performance speed.

Discussion

Again, very bright college students failed to apply the seemingly obvious principle of testing simple hypotheses first in solving a diagnostic problem. Figure 5 shows the number of subjects in each group who used the optimal strategy on each training (one-bug) problem. Somewhat in contrast to Ashby's findings, nine subjects in the experimental group started out applying the principle. Still surprising, though, is the fact than on not one problem did all twelve one-bug subjects use the simple to complex strategy. The control group's abysmal performance is just as clear. Seven control subjects applied the optimal strategy on the first problem (it turns out that due to random placement of the algebra statements on the screen, four of these seven received a presentation that was both optimal and sequential), but this quickly drops off, and on only one other problem did as many as four subjects test the simpler hypotheses first. Figure 6 shows the same pattern for the test problems, differing only in the fact that there were some problems on which all experimental subjects applied the optimal principle.

One might think that subjects who fail to apply the optimum strategy are testing hypotheses in one of two alternative ways. First of all, subjects might consider the more complex (longer) algebra statements to be more likely to contain the bug. Subjects were explicitly told, however, that the bug was equally likely to be in any statement, and an examination of individual subjects' testing order shows no clear tendency to test in order of more complex to less complex. Indeed, not one single subject tested an entire set of hypotheses in this reverse-optimal strategy. The second alternative strategy might be a sequential strategy where subjects test from top to

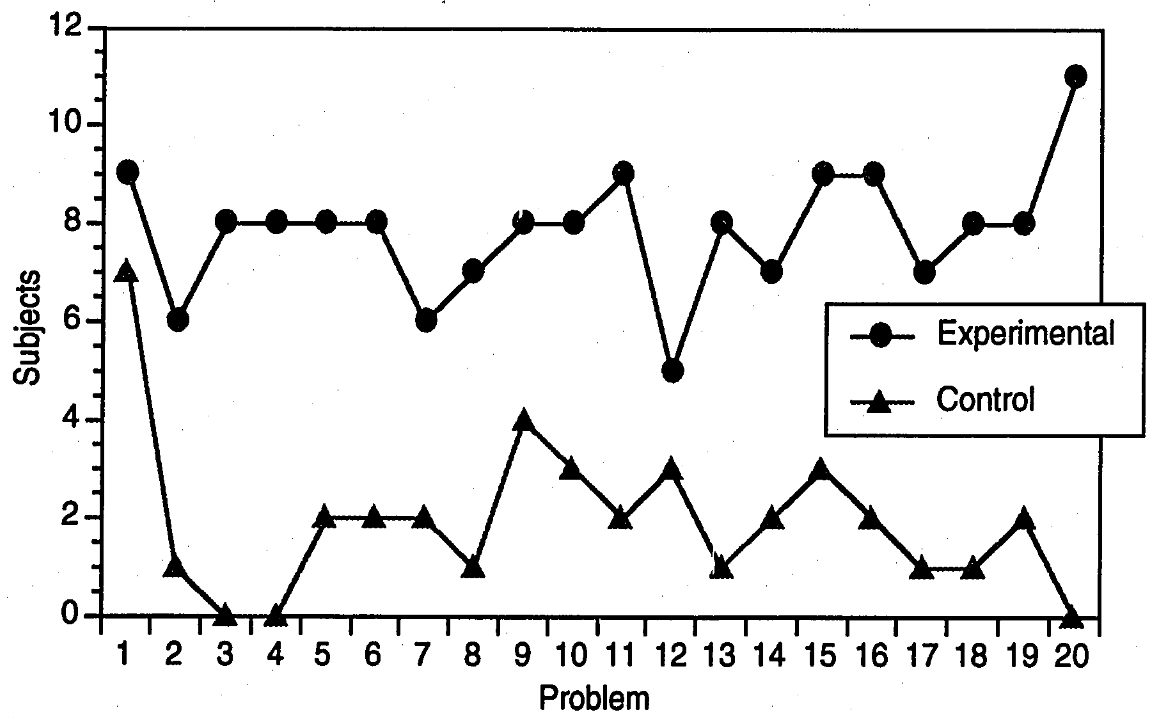

Figure 5. Number of subjects using optimum strategy on each problem set in the training problems.

bottom (or vice versa) as a bookkeeping strategy. Since a check mark was placed by previously tested hypotheses, the motivation for such a strategy was reduced, and an examination of the data indicates that an average of less than one-sixth of the subjects used a sequential strategy on any given hypothesis set.

When the subjects who learned to apply the principle in the one-bug condition continued to apply it in the two-bug condition, they were exhibiting what Detterman (1993) identifies as *near transfer.* The training task and the test task shared surface features (algebra problems as hypotheses, etc.) and were differentiated only by the fact that the training problems contained only one incorrect statement while the test problems contained two. As

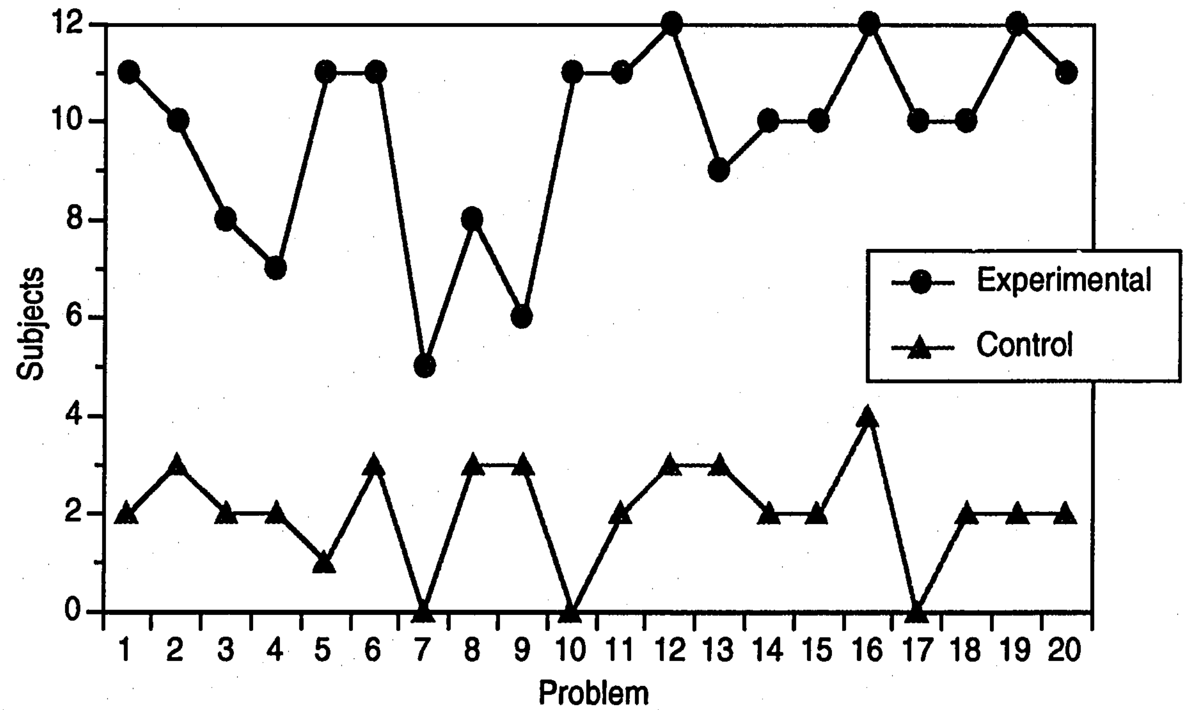

Figure 6. Number of subjects using optimum strategy on each problem set in the test problems.

Detterman suggests, an important prerequisite to this near transfer may be that subjects' attention should be drawn to similar features of the training and test problems. In this case, the equal probability of bug location in any given statement was, indeed, a key factor in encouraging application of the optimum hypothesis selection principle, and in both training and test problems, subjects were reminded of this equal probability.

Another way to describe this near transfer is as a sort of reverse practice. Reverse practice occurs when practice in some task A produces more improvement in task B than does practice in that task B. In this experiment, practice in the one-bug condition facilitated performance in the two-bug condition (the significant interaction indicates that performance in the experimental group actually improved in the test problems), while practice in the two-bug condition (control subjects) had no effect on performance in the test problems.

Experiment 2.

Experiment 1 showed near transfer of a particular principle; that of selecting simple hypotheses for testing . Experiment 2 was designed to determine whether a more general transfer or far transfer might be obtained. Once again, hypothesis sets constructed of algebra statements served as a framework for training the "starting simple" principle. A simple checkbook balancing task with which almost all adults are familiar served as the task toward which transfer was targeted.

The two tasks share no common surface structure or identical elements which, according to Detterman's (1993) definition is a requirement for far transfer. Neither were any hints (explicit or otherwise) given as has been the case in many studies which claim to find transfer (Detterman, 1993). Subjects were not told of any relationship between the two problems. Indeed, the only commonality in the instructions given the subjects in the two tasks was the explanation that errors were equally likely to occur in any of the components of the problem (algebra statements in training, clerical or logical errors in test).

It was hypothesized, first of all, that Ashby (1987) would again be replicated and subjects would be able to apply the "simplest first" principle in the experimental training condition (one-bug), but not in the control training condition (two-bug). Secondly, if (as hypothesized) far transfer can occur, experimental (one-bug) subjects should show a greater application of the "simplest first" principle in the test condition than do control subjects.

METHOD

Subjects.

Subjects were forty students enrolled in an introductory psychology course at a Houston area community college who volunteered in order to learn more about experimental design and technique.

Materials.

The computer portion of the experiment used the same Hypercard program as in Experiment 1. The checkbook balancing portion of the experiment consisted of simulated checks and checking account statements printed with an Apple Laserwriter. For the checkbook portion, subjects were given a simple calculator with large keys suitable for basic mathematical functions.

Procedure.

Subjects were randomly assigned to either a control or experimental condition. In the control condition, subjects were given the same 2-bug problems as in Experiment 1 (in which the optimal strategy was never learned) and then were presented with five checkbook balancing problems which included one of two possible types of bug. The first type of bug included clerical mistakes that could be quickly and easily tested for by comparing the amounts on the checks with those listed in the bank statement. The second type of bug was a logical problem (i.e. a deposit was subtracted rather than added, a service charge was overlooked, etc.) which was a great deal more difficult to test. Refer to Appendix A for examples of checkbook materials. In the experimental condition, subjects were give the same 1-bug problems as in Experiment 1 (in which the optimum strategy was learned) and then presented with five checkbook type problems

Table 1. Means for Experiment 2.

Means for algebra task (training)

Condition	N	Mean	Standard Deviation	Standard Error
1-Bug	20	11.25	1.713	.383
2-Bug	20	8.75	1.964	.401

Means for bank task (transfer)

Condition	N	Mean	Standard Deviation	Standard Error
1-Bug	20	8.4	1.046	.234
2-Bug	20	6.6	1.095	.245

as in the control condition. Location of bug was random both in the algebra problems and in the checkbook problems. The independent variable, then, was training condition in the algebra problems (one-bug vs. two-bug) and the dependent variable was performance on the bug location task in the checkbook problems; in particular, which bug type (simple or complex) the subject chose to test first. In addition, time to complete individual tasks were recorded to test (as in Experiment 1) the assertion that testing simple hypotheses first is actually the most efficient strategy.

Results and Discussion.

Again, the findings of Ashby were supported as subjects in the one bug group learned the optimal strategy, while those in the two bug group did not, $F_{(1, 38)} = 21.3$, $p < .001$. Also, the experimental group carried this learning over into the bank account task, where they tested the easier to test hypotheses more often than did the control group, $F_{(1,38)} = 28.239$, $p < .001$ (see Table 1). The control group again failed to find the optimum strategy in

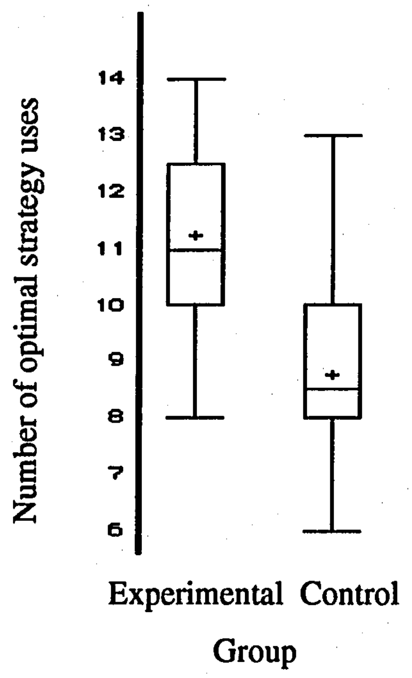

Figure 7. Box plot indicating number of times subjects in each group used the optimal strategy for training (algebra) task.

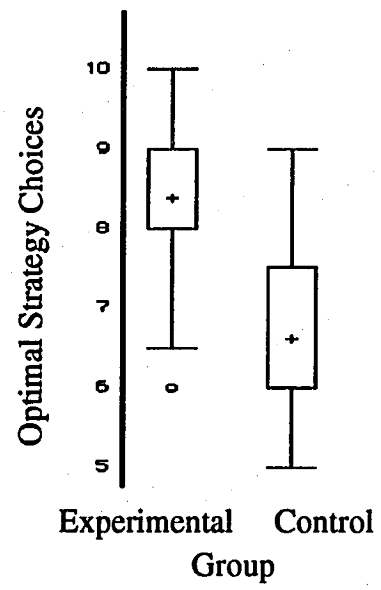

Figure 10. Box plot indicating how many times subjects in each group used the optimal strategy for bank task performance.

either half of the experiment.

The time to complete a particular bank task was significantly lower in the experimental group than in the control group, $F_{(1,38)} = 11.794$, $p = .0015$. This underscores the simple to complex method of hypothesis testing as the most efficient strategy.

Finally, evidence that subjects who performed well in the training (algebra) task were the same subjects who performed well on the transfer (bank statement) task is provided by a correlation coefficient of .617 between these two tasks. The coefficient is statistically significant, $t_{(38)} = 4.833$, $p < .01$.

Looking at the data in a finer grain, it turns out that the correlation between the training (algebra) task and the transfer (bank statement) task for the 1-bug group (those who experienced successful transfer) was .646, yielding a $t_{(18)}$ of 3.591, $p<.01$, again portraying a strong relationship between subjects who learn the proper strategy in the training task and those who utilize that strategy in the transfer task. However, when the scores for the 2-bug group (which never learned the principle) are examined, the corresponding correlation is only .112 ($t_{(18)} = .479$, $p=.638$). In other words, since the strategy was not learned in the training task, it could not be utilized in the transfer task.

GENERAL DISCUSSION

That testing simple hypotheses first is the most efficient strategy in the hypothesis selection component of diagnostic problem solving is certainly validated by the analysis of time measures presented in this research. This research suggests that once that principle is learned, though, it can be transferred to other tasks that are similar (near transfer) and to tasks that are dissimilar (far transfer). Even though cases where hypotheses vary greatly in probability may encourage problem solvers to sample hypotheses based on probability rather than on complexity, the ineptitude with which most problem solvers approach questions of probability and the likelihood of probability being relatively equivalent among the final generated set of hypotheses about a particular fault clearly support this “simple first” principle. While this strategy might seem obvious, it is clear that few people learn to apply it except in extremely simple diagnostic problems.

Previous research on transfer of training indicates that certain characteristics of the training and transfer tasks must be observed if transfer is to occur. In some experiments in which transfer was obtained, the training task and transfer task share identical elements, supporting the idea that high fidelity between tasks is a requirement for transfer. In others, hints given to subjects were required to generate transfer of an abstract principle from a training task to target task. Detterman (1993) questioned research claiming to find transfer. He asserted that identical elements of task, hints, etc. make the findings not ones of transfer, but rather of simple learning or of following instructions that come in the form of explicit hints. The research reported here finds both near and far transfer of a principle that is useful in diagnostic problem solving with little fidelity (at least in the case of far transfer) and with no hints or instruction.

Experiment 2 indicates that one does not need to perform the training task in the identical domain of the target task for this training to transfer. In other words, some general problem solving skills may be taught at a rather fundamental level without concern for fidelity between training and transfer tasks. Later, specialization in study and experience will develop within the solver a knowledge domain which will generate the high quality hypotheses characteristic of experts. Without the fundamental skill of efficient hypothesis selection, however, even the highest quality hypothesis domains may not generate solutions to the problem.

Viewed in the light of Detterman's (1993) vehement argument against transfer, the findings presented here seem extremely novel. Indeed, many studies of problem solving (for example Gick & Holyoak, 1980; Weisberg, DiCamillo, & Phillips, 1978) find transfer only in the presence of explicit information given to ensure that the solver connects the training problem with the test problem, thus supporting Detterman's reluctance to identify these effects as transfer at all. Nisbett, Fong, Lehman, and Cheng (1987), however, assert that a long tradition of formal discipline (teaching abstract rule systems by teaching logic and classical languages such as Latin, etc. not for the subject matter itself, but for the reasoning skills learned coincidentally) is valid because of transfer of training. In other words, students can apply the abstract rules of reasoning learned in one subject to a completely different subject. They provide evidence that abstract instruction in such rules as the law of large numbers can enable subjects to apply these rules to a wide range of problem content. Similarly, the research presented here shows that subjects can learn an abstract rule in one type of problem and apply it in an entirely different type.

The identification of a particular component of diagnostic problem solving and, more importantly, the exhibition of transfer in the training of this skill, lends hope to the training of fault diagnosis in many fields. Rather than focusing solely on domain-specific knowledge as a framework for diagnosis, actual diagnostic skill may be trained in safe, inexpensive environments. Doctors might be trained to test hypotheses efficiently in an artificial setting rather than risking the health of their patients on their decisions during training. Pilots may be prepared to test hypotheses efficiently before the actual need to go through an emergency checklist arises. Computer programmers may cut software development costs drastically by being taught efficient hypothesis selection along with programming languages, etc. While many questions about diagnosis and the best way to train diagnostic skill remain unanswered, the prospects for transfer of training appear to be brightened by these findings.

REFERENCES

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

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

Anderson, J. R. (Ed.). (1981). Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Ashby, A. B. (1988). Hypothesis selection strategies in a debugging task. Unpublished master's thesis, Rice University, Houston.

Ashby, A. B. and Lane, D. M. (1988). Hypothesis selection strategies in a debugging task. Unpublished manuscript, Rice University, Houston.

Baldwin, T.T., and Ford, J.K. (1988). Transfer of training: A review and directions for future research. Personnel Psychology, 41, 63-105.

Balla, J. I. (1980). Logical thinking and the diagnostic process. Methods of Information in Medicine, 19, 88-92.

Brooks, F. (1972). The mythical man-month: Essays on software engineering. Reading: Addison-Wesley.

Brooks, R. (1983). Towards a theory of the comprehension of computer programs. International Journal of Man-Machine Studies, 18, 543-554.

Comstock, J.R., Jr. (1984). The effects of simulator and aircraft motion on eye scan behavior. In Proceedings of the Human Factors Society 28th Annual Meeting, 128-132. Sanat Monica, CA: Human Factors Society.

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphemonenon. In D. K. Detterman and R. J. Sternberg (Eds.). Transfer on trial: Intelligence, cognition, and instruction. Norwood, NJ: Ablex.

Dominowski, R. L., and Jenrick, R. (1972). Effects of hints and interpolated activity on solution of an insight problem. Psychonomic Science, 26, 335-338.

Duncker, K. (1945). On problem solving. Psychological Monographs, 58 (Whole No. 270).

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). Medical problem solving: An analysis of clinical reasoning. Cambridge, MA: Harvard University Press.

Fingerman, P. and Levine, M. (1974). Non-learning: The completeness of the blindness. Journal of Experimental Psychology, 102(4), 720-721.

Fulgosi, A., and Guilford, J. P. (1968). Short-term incubation in divergent production. American Journal of Psychology, 81, 241-246.

Gick, M. L., and Holyoak, K. J. (1980). Analogical problem solving. Cognitive Psychology, 12, 306-355.

Gugerty, L. & Olson, G. (1986). Comprehension differences in debugging by skilled and novice programmers. In E. Soloway & S. Iyengar (Eds.) Empirical Studies of Programmers (pp. 13-27). Norwood, NJ: Ablex.

Holt, R., Boehm-Davis, D., & Schultz, A. (1987). Mental representations of programs for student and professional programmers. In G. Olson, S. Sheppard, & E. Soloway (Eds.) Empirical Studies of Programmers: Second Workshop (pp. 33-46). Norwood, NJ: Ablex.

Katona, G., (1940). Organizing and memorizing. New York: Columbia University Press.

Kessler, C. M. and Anderson, J. R. (1986). A model of novice debugging in Lisp. In E. Soloway and S. Iyengar (Ed.), Empirical Studies of Programmers. Norwood, New Jersey: Ablex.

Kieras, D. (1987). The role of cognitive simulation models in the development of advanced training and testing systems. Technical Report -87. Office of Naval Research - 23. Ann Arbor: University of Michigan.

Kohler, W. (1927). The mentality of apes. New York: Harcourt, Brace.

Krechevsky, I. (1932). "Hypotheses" in rats. Psychological Review, 39, 516-532.

Lane, D. M., McDaniel, J. R., Bleichfield, B. E., & Rabinowitz, F. M. (1976). Journal of Experimental Psychology: Human Learning and Memory, 2(4), 489-496.

Lesgold, A. M. (1984). Acquiring expertise. In Jr. R. Anderson & S. M. Kosslyn (eds.), Tutorials in learning and memory: Essays in honor of Gordon Bower. San Francisco: Freeman.

Levine, M. (1966). Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 71, 331-338.

Levine, M. (1971). Hypothesis theory and nonlearning despite ideal S-R reinforcement contingencies. Psychological Review, 78, 130-140.

Levine, M. (1975). A cognitive theory of learning. Hillsdale, NJ: Erlbaum.

Lewis, C.H., (1978). Production system models of practice effects. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Lintern, G. (1991). An informational perspective on skill transfer in human-machine systems. Human Factors, 33, 251-266.

Littman, D., Pinto, J., Letovsky, S., & Soloway, E. (1986). Mental models and software maintenance. In E. Soloway & S. Iyengar (Eds.) Empirical

Studies of Programmers (pp. 80-98). Norwood, NJ: Ablex.

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

Luchins, A. S. (1942). Mechanization in problem solving. Psychological Monographs, 54, (Whole No. 248).

Maier, M. (1931). Reasoning in humans: II The solution of a problem and its appearance in consciousness. Journal of Comparative Psychology, 12, 181-194.

Morris, N.M., and Rouse, W.B. (1985). Review and evaluation of empirical research in troubleshooting. Human Factors, 27, 503-530.

Murray, H. G., and Denny, J. P. (1969). Interaction of ability level and interpolated activity (opportunity for incubation) in human problem solving. Psychological Reports, 24, 271-276.

Nanja, M. & Cook, C. (1987). An analysis of the on-line debugging process. In G. Olson, S. Sheppard, & E. Soloway (Eds.) Empirical Studies of Programmers: Second Workshop (pp. 172-184.). Norwood, NJ: Ablex.

Neves, D.M., and Anderson, J.R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In Anderson, J.R. (Ed.). Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Newell A., & Simon, H. (1972). Human problem solving. Englewood Cliffs: Prentice-Hall.

Nisbett, R. E., Fong, G. T., Lehman, D., and Cheng, P. (1987). Teaching reasoning. Science, 238, 625-631.

Povenmire, H.K., and Roscoe, S.N. (1973). Incremental transfer effectiveness of a ground-based general aviation trainer. Human Factors, 15, 534-542.

Schicke, R. K. Economic aspects of diagnostic services in health care. Methods of Information in Medicine 1983; 22:1-33.

Silveira, J. (1971). The effect of interruption timing and length on problem solution and quality of problem processing. Unpublished doctoral dissertation, University of Oregon, Eugene, OR

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

Spence, K. (1945). An experimental test of the continuity and non-continuity theories of discrimination learning. Journal of Experimental Psychology, 35, 253-266.

Stone, D., Jordan, E., & Wright, M. (1990). The impact of Pascal education on debugging skill. International Journal of Man-Machine Studies, 33(1), 81-95.

Thorndike, E.L. (1903). Educational Psychology. New York: Lemcke and Buechner.

Vessey, I. (1985). Expertise in debugging computer systems: A process analysis. International Journal of Man-Machine Studies, 23(5), 459-494.

Vessey, I. (1986). Expertise in debugging computer programs: An analysis of the contents of verbal protocols. IEEE Transactions on Systems, Man, and Cybernetics, 16(5), 621-637.

Wallas, G. (1926). The art of thought. New York: Harcourt Brace Jovanovich.

Weisberg, R., DiCamillo, M. and Phillips, D. (1978). Transferring old associations to new problems: A nonautomatic process. Journal of Verbal Learning and Berbal Behavior, 17, 219-228.

Wickelgren, W. A. (1974). How to solve problems. San Francisco: Freeman.

Wiedenbeck, S. (1986). Processes in computer program comprehension. In E. Soloway & S. Iyengar (Eds.) Empirical Studies of Programmers (pp. 48-53). Norwood, NJ: Ablex.

APPENDIX A

TYPICAL SAMPLE CHECK

Check # 1005

Mr. & Mrs. John Doe
1224 Anystreet
Anytown, USA

Pay to the Order of Houston Lighting & Power $ $\boxed{143.56}$

One Hundred forty three and 56/100 dollars

Sample Checkbook Statement

Beginning Balance: 458.62

Check No.	Amount	Payee
1001	34.68	Kroger's
1002	56.32	Bookstop
1005	396.42	Citibank
1006	142.36	H. L. & P.
1007	26.35	Sportstown
1008	42.56	Warner Cable
1009	33.25	Entex
1010	66.58	Exxon Credit
1011	55.64	State Farm Ins.
Deposit #1	1524.35	
1012	12.50	Walden Books
1013	36.54	Houston Chronicle
1014	100.00	Rice Alumni
1015	23.32	Macy's
1016	47.36	Bill's Automotive
Deposit #2	565.42	
1017	48.99	Walmart

Total Checks: 2012.76
Total Deposits: 2184.56
Service Charges: 8.50
New Balance: 212.45