Title
Transfer in Problem Solving as a Function of the Procedural Variety of Training Examples

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 9(0)

Authors
Catrambone, Richard
Holyoak, Keith J.

Publication Date
1987

Peer reviewed

Richard Catrambone Keith J. Holyoak

University of Michigan University of California at Los Angeles

Transfer in Problem Solving as a Function of the Procedural Variety of Training Examples

Abstract

Students often have difficulty solving homework assignments in quantitative courses such as physics, algebra, programming, and statistics. We hypothesize that typical example problems done in class teach students a series of mathematical operations for solving certain types of problems but fail to teach the underlying subgoals and methods which remain implicit in the examples. In the studies reported here, students in probability classes studied example problems that dealt with the Poisson distribution. In Experiment 1, the four examples all used the same solution method, although for one group the examples were superficially more dissimilar than for the other group. All subjects did well on the Near Transfer target problem that used the same subgoals and methods as the training examples. However, most did poorly on two Far Transfer target problems that had different subgoal orders and different methods. These results suggest that subjects typically learn solutions as a series of non-meaningful mathematical operations rather than conceptual methods in a subgoal hierarchy. In Experiment 2, one group studied problems that demonstrated two different subgoal orders using different methods while the other group received superficially different problems which had identical subgoal orders and methods. Both groups still had difficulty with the Far Transfer problems. Subjects who received examples with varied subgoal orders and methods seemed to isolate the subgoals, however, but not the methods. This result suggests that goals and methods may be useful ways of characterizing training problems. However, students may require explicit instruction on subgoals and methods in order to successfully solve novel problems.

Introduction

A relatively consistent finding in the analogical reasoning and transfer literature is that subjects do not seem to make use of prior information to solve new problems if the new problems differ from training examples in more than minor ways (Gick & Holyoak, 1980, 1983; Reed, Dempster, & Ettinger, 1985; Spencer & Weisberg, 1986). If similarities between training examples and target problems are pointed out to subjects or if they are encouraged to consider similarities between problems or domains, then subjects have somewhat more success at noticing and applying analogies or transferring information (Gentner & Gentner, 1983; Gick & Holyoak, 1983; Tenney & Gentner, 1984).

Card, Moran, and Newell (1983) proposed the GOMS model to account for the text-editing behavior of experts performing routine tasks. In this model, the expert knowledge representation consists of four components: Goals, Operators, Methods, and Selection rules. We would like to propose that in quantitative domains such as mathematics and physics, students acquire, or should acquire, goals, methods, and selection rules for solving problems as a function of the examples they study. Operators are simple mathematical procedures which college students typically already possess, such as calculating an average. Goals are initially quite general: solve the problem. After studying several examples, a student's goal may be more refined so that it is something like "get an answer that looks like the examples' answers." This type of goal is especially likely if the examples solve for the same unknown using the same procedure. In this case, students may simply learn that in order to achieve the goal they need to string together a series of operations. However, if the examples are varied in their givens and ultimate goal (the unknown being solved for), then students are less likely simply to string together a series of operations. Rather, they may recognize and develop subgoals which correspond to the steps in the examples. In addition, students will perceive that these subgoals can be reached by particular methods, which develop after students see a set of operations used together several times to achieve some subgoal. That is, students will compartmentalize the problems into subgoals which call on particular methods to satisfy them. A method will consist of a series of mathematical operations connected together conceptually. With experience, students may develop different methods for achieving the same subgoal. The particular method chosen will depend on the particular givens in the problem. Different problems will evoke different subgoals which will in turn evoke different methods. Students will develop selection rules for choosing which method to use. If students had only studied one type of example, then they would only have one method for solving problems in that domain. In fact, the method may really be a series of operations with no clear organizing feature except order of application. Thus, varied examples may be necessary in order to demonstrate how a series of operations can be grouped as a particular method for achieving a particular subgoal.

Reed et al. (1985) conducted several experiments using college students taking an algebra course. Subjects studied word problems dealing with traditional topics like distance, mixture, and work and then solved target problems. Reed et al. manipulated the superficial similarity of the target problems to the training problems. Their general finding was that subjects exhibited little transfer of the concepts from the training problems to the target problems except in those cases where the target problems were essentially identical in solution procedure to the examples. Reed et al. (1985) concluded that subjects were relying on a syntactic approach to the problems. This suggests that in general the subjects did not understand the goals and methods being demonstrated in the problems but rather had learned a series of operations for solving the problems.

We might suppose that if students were exposed to training examples that used different solution procedures, they would be more likely to learn the underlying subgoals and methods illustrated by the examples. This might happen because they would attempt to determine the similarities (such as the goal structure) between different series of operations which produce a value for the same final goal. A resolution process could lead to the identification of subgoals, methods, and generalizations of the methods (Anderson, 1983; VanLehn, 1985). However, if the series of operations from example to example are too different, students will fail to identify subgoals or to isolate a series of operations as a method (VanLehn, 1985). Each example will be perceived as unique.

Overview of Current Studies

We suspect that the difficulty students have in grasping subgoals and methods is due to the default reasoning of students who are still relatively unsophisticated in a particular domain. By default, students focus on superficial features of problems and the operations used to achieve an end goal because the features and operations are easier to isolate than the underlying subgoals and methods (Larkin, McDermott, Simon, & Simon, 1980; Schoenfeld & Herrmann, 1982). Students have a great deal of experience with the real world objects such as decks of cards and blocks of wood which populate the world of quantitative problems. Students in quantitative courses are also quite experienced with mathematical operations such as multiplication and addition as well as somewhat more "compiled" operations such as calculating means. Thus, it is not surprising that these students would tend to focus their attention and organize their problem solving skills around the mathematical operations with which they are most familiar (Greeno, Riley, & Gelman, 1984; Hayes, Waterman, & Robinson, 1977). We would like to begin to investigate what qualities of examples can help students go beyond their default focus and help them isolate subgoals and methods in a particular domain.

We solicited paid volunteers from three upper-level probability courses at the University of Michigan. The courses are quite similar for the first third of the semester. All students learn about counting rules (e.g., ordered and unordered sampling) and are then introduced to the notion of a random variable. Then students learn about certain basic discrete probability distributions such as the binomial, Poisson, and geometric. The courses introduce the binomial distribution first, followed by the Poisson distribution. Students participated in the present experiments, which dealt with the Poisson distribution, after learning the binomial distribution but before learning the Poisson distribution.

The Poisson Distribution and Some Examples

The Poisson distribution is often used to approximate binomial probabilities for events that occur in time or space with some small probability p. The Poisson equation is: $P(X=x)=[e^{-\lambda}(\lambda)^{x}]/x!$. It can be used to calculate probabilities for various values of X. Then the predicted frequencies of various values of X can be calculated by multiplying the probabilities by the total number of events. These steps are illustrated in Figure 1.

The example in Figure 1 deals with an event occurring randomly in time. The Poisson distribution is also used to model events occurring randomly in space. For example, one could reasonably fit a Poisson distribution to the number of fossils found in each section of a partitioned quarry. This problem is presented in Figure 2. It can be solved by the same procedure as the first problem.

It seems intuitively clear that a person could learn to solve problems of this type by memorizing the series of operations without understanding the meaning of the output from the operations. Nevertheless, the two examples do differ on the surface: one is about events in time and the other is about events in space. Thus, it is possible that students who study these examples may notice that the units in the operations are different and they may be induced to consider how the units were derived and to form a generalization about the operations. On the other hand, subgoals and methods can be identified more directly by comparing procedural differences in problems. Thus it is debatable whether superficial differences are sufficient to induce students to recognize these "deeper" aspects.

The subgoals and methods (in parentheses) for the two problems described above could be listed as follows:

1) find λ (calculate λ as a weighted average)

2) find the expected probabilities for each X (plug $X=x$ into the Poisson equation)

3) find the expected frequencies for each X (multiply each $P(X=x)$ by the total observed frequency)

A physicist observed a radioactive substance during 2608 time intervals (each 7.5 seconds long). She recorded the number of particles reaching a geiger counter for each period. Let x be the number of particles observed in each time period. Fit a Poisson distribution to x, that is, give the expected frequencies for the different values of x based on the Poisson model.

Number of Particles Observed	Observed Frequency
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11 or more	6
Total	2608

Solution:

$$E(X)=[0(57)+1(203)+2(383)+3(525)+4(532)+5(408)+6(273)+7(139)+8(45)+9(27)+10(10)+11(6)]/2608 = 10092/2608 = 3.87 = \lambda$$

= average number of particles that reached geiger counter each period

$$P(X=x) = [(e^{-3.87})(3.87)^{x}]/x! = [(.021)(3.87)^{x}]/x!$$

Fitted Poisson Distribution:

x	Expected Frequency
0	.021 x 2608 = 55
1	.081 x 2608 = 211
2	.157 x 2608 = 409
3	.203 x 2608 = 529
4	.196 x 2608 = 511
5	.152 x 2608 = 396
6	.098 x 2608 = 256
7	.054 x 2608 = 141
8	.026 x 2608 = 68
9	.0113 x 2608 = 29
10	.0044 x 2608 = 11
11 or more	.00153 x 2608 = 4

Figure 1: Example problem for event occurring in time.

A horizontal quarry surface was divided into 30 squares about 1 meter on a side. In each square the number of specimens of the extinct mammal Ditolestes motissimus was counted. The results are given in the table below. Fit a Poisson distribution to $\underline{x}$, that is, give the expected frequencies for the different values of $\underline{x}$ based on the Poisson model.

Number of Specimens per Square	Observed Frequency
0	16
1	9
2	3
3	1
4 or more	1
Total	30

Figure 2: Example problem for event occurring in space.

Experiment 1 explores how well students learn methods and subgoals from examples that differ only in the superficial ways shown above. If students studied problems like those above, they should be able to solve other superficially different problems that involve the same set of operations. It is less clear would happen if they tried to solve problems that had a different subgoal order and used modified methods. Consider the problem below.

Suppose you were making a batch of raisin cookies and you did not want more than one cookie out of 100 to be without a raisin. How many raisins will a cookie contain on the average in order to achieve this result? Use the Poisson distribution to find your answer.

Solution (not presented to subject):

$$P(X=0) = .01 = [(e^{-\lambda})(\lambda^{0})]/0!$$
$$.01 = e^{-\lambda}$$
$$\ln(.01) = \ln(e^{-\lambda})$$
$$-4.6 = -\lambda$$
$$4.6 = \lambda = \text{average number of raisins per cookie}$$

Figure 3: Cookie problem.

This "cookie" problem literally looks different than the prior ones. In this problem the student must realize that he or she is provided with the following piece of information: $P(X=0) = .01$ (i.e., only one cookie out of 100 should have zero raisins). He or she must also realize that the goal is to find λ—the expected value of the random variable which in this case is the average number of raisins that a cookie receives. If they recognize these two facts then the problem simply becomes a matter of inserting $P(X=0) = .01$ into the Poisson equation and solving for λ. It is unclear, however, how these realizations would follow from the types of practice problems to which the students have thus far been exposed. Students would only have learned a series of operations. They would not have learned that the calculation of λ is a subgoal which can be carried out by several different methods depending on the givens. One way to calculate λ is to find a weighted average as was done in the example problems. Another way is to find values for the other unknown in the Poisson equation (i.e., a value for some $P(X=x)$) and then solve for λ.

The subgoals and methods for this problem are listed below:

1) find the known value for some $P(X=x)$ (divide 1 by 100 to get $P(X=0)$)

2) find λ (plug $P(X=0)$ into Poisson equation and solve for λ)

Consider another problem:

Suppose you took a random sample of 500 people and found out their birthdays. A "success" is recorded each time a person's birthday turns out to be January 1st. Assume there are 365 days in a year, each equally likely to be a randomly chosen person's birthday. Fit a Poisson distribution to x̲ (the number of people born on January 1st) and find the predicted likelihood that exactly 3 people from the sample are born on January 1st.

Solution (not presented to subject):

$\lambda = 500/365 = 1.37 =$ average number of people born on any given day

$$P(X=3) = [(e^{-1.37})(1.37^3)]/3!$$
$$=[(.254)(2.57)]/6$$
$$=.109$$
=likelihood of exactly three people being born on January 1st (or any other given day)

Figure 4: Birthday problem.

The birthday problem requires that the student realize that λ can be calculated simply by dividing the number of days by the number of people (as opposed to being calculated as a weighted average from an observed frequency table). It also requires that the subject realize he or she was being asked to solve only for $P(X=3)$ and not to produce an expected frequency table.

The subgoals and methods for this problem are:

1) find λ (divide the number of events [birthdays] by the number of slots [days of the year])

2) find $P(X=3)$ (plug $X=3$ into the Poisson equation)

Both the cookie and birthday problems have different or modified methods compared to the training examples, yet they still have either the same subgoals (in a different order) or fewer subgoals. Students' performance on the cookie and birthday problems should indicate whether they isolated subgoals and methods during training or whether they simply learned a series of operations to achieve the single goal of producing an expected frequency table.

Experiment 1

Method

Subjects. Seventy-one students from three probability classes were recruited and were paid $7 for their participation.

Materials and Procedure. Subjects were given a booklet to study. The cover page contained a description of the relationship between the binomial and Poisson distributions and provided the Poisson equation. The next four pages contained four worked out Poisson distribution problems isomorphic to the radioactive particle and quarry problems. Subjects were told to study the problems carefully since after studying them they would be asked to solve three problems. They were also told they could refer back to the cover page but not to the examples. This was done to increase the likelihood that subjects would pay attention to the examples and how they were solved.

Subjects were randomly divided into two groups. The SAME group studied four examples which dealt with the same class of events: either four space problems or four time problems. The DIFFERENT group received problems from both classes of events: two space problems and two time problems. All problems were solved using the same procedure, which was identical to the radioactive particle and quarry problems discussed above. The example problems were picked from a pool of four space and four time problems. There was no effect in subjects' performance on the target problems as a result of the specific space or time problems a subject received, and all reported results are collapsed over this factor.

After studying the examples subjects worked on the three target problems. The first target problem is labeled the "Detroit Tiger" problem and is presented below. This problem will be called a Near Transfer problem since it embodies the same subgoals (and same subgoal order) and methods as the training examples.

In a 162-game baseball season, the Detroit Tiger infield made a total of 107 errors. The table below gives the number of games in which x errors were made. Fit a Poisson distribution to x, that is, give the expected frequencies for the different values of x based on the Poisson model.

Number of Errors x made in a game	Observed Frequency
0	85
1	52
2	20
3 or more	5
Total	162

Figure 5: Detroit Tiger Problem

The second and third target problems were the cookie and birthday problems, respectively. They will be referred to as Far Transfer problems because they involve different subgoal orders and methods than the training examples. The order of the target problems was the same for all subjects. Subjects worked at their own pace for the entire experiment. In general, subjects took about 35 minutes to complete the experiment. Subjects were asked to show all their work but could use a calculator for the basic arithmetic. The solution and error frequencies were analysed using the likelihood ratio chi-square test (G^2) which is a test of equality of proportions between rows or columns.

Results

Subjects' answers to the transfer problems were first scored as correct/incorrect. Both groups did well on the Detroit Tiger (Near Transfer) problem: 91% and 94% correct for the Same and Different groups, respectively. On the cookie problem (Far Transfer) the DIFFERENT group did somewhat better than the SAME group: 42% versus 23%, $G^2(1) = 2.9$, $p < .09$. The DIFFERENT group also did better on the birthday problem (Far Transfer), 33% versus 23%, but this difference did not approach conventional significance levels, $G^2(1) = .97$, $p > .3$. Overall, 32% of the subjects solved the cookie problem and 28% solved the birthday problem.

Subjects errors were analysed separately for the cookie and birthday problems. The first type of error for the cookie problem (called ULAMBDA in Table 1) is a failure to recognize the goal of the problem, to solve for λ. That is, the subject does not realize that the average number of raisins per cookie is λ. The second error type (PX0) is a failure to recognize that $P(X=0) = .01$ is provided in the problem. The third category (FREQ1) is whether a subject attempted to make up a frequency table as a way of solving the problem (i.e., they generated hypothetical data). If a subject made up a frequency table, this would indicate that he or she was most likely trying to make the target problem appear like the examples in order to use the familiar procedure. This approach is an error since there is no way to create a useful frequency table with the information given.

There are also three error categories for the birthday problem. The first category (SLAMBDA) is a failure to recognize that λ is the average number of people that are born on any given day. This value is simply the number of people (500) divided by the number of days in the year. (A priori it seemed unlikely that a subject would understand that λ would be the average number of people born on a given day but fail to realize that this value would be 500/365. This assumption was supported by the protocols.) The second category (PX3) is a failure to realize that the problem's goal was to solve for $P(X=3)$ rather than to create a frequency table or to find only the expected value of X. The third category (FREQ2) is identical to the third category for the cookie problem; it counts how often subjects tried to make up a frequency table as an aid to solving the problem. Again, this approach will not help to solve the problem.

Sixty-eight percent (48 out of 71) of the subjects failed to solve the cookie problem and 72% (51 out of 71) failed to solve the birthday problem. Table 1 indicates the error types and their frequencies for the two far transfer problems. It also presents the frequencies collapsed across the group dimension since analyses indicated there were no differences between the groups (for subjects who got a problem wrong) with respect to the frequency of different error types.

Table 1

Percentage of Subjects Who Made Particular Types of Errors (Experiment 1)

Transfer Problem	Error Type	Group		
		SAME	DIFFERENT	Total
Cookie Problem				
		n=27	n=21	n=48
	ULAMBDA	85 (23)	67 (14)	77(37)
	PX1	96 (26)	86 (18)	92(44)
	FREQ1	41 (11)	62 (13)	50(24)
Birthday Problem				
		n=27	n=24	n=51
	SLAMBDA	96 (26)	100 (24)	98(50)
	PX3	67 (18)	62 (15)	65(33)
	FREQ2	15 (4)	21 (5)	18(9)

Note. Frequencies are given in parentheses. Percentages are based on the number of subjects who made a particular error divided by the number of subjects in each group who got the problem wrong (given at the top of each column for each of the transfer problems), not the total of number of subjects in the group.

Discussion

It was intuitively plausible to expect both groups of subjects to solve the Detroit Tiger problem equally well since it used the same series of operations as the examples. However, both groups were expected to do equally poorly on the far transfer problems because we suspected that the manipulation of SAME versus superficially DIFFERENT training examples to be unrelated to whether or not subjects learned the underlying subgoals and methods in the training examples. These expectations were largely confirmed.

It seems clear that subjects who had difficulty with the far transfer problems had difficulty because they had primarily learned a series of operations for solving problems of the training type and had not learned the underlying subgoals or formed generalizations of the methods. Sixty-eight percent of the subjects could not solve the cookie problem and for 92% of those subjects the reason seemed to be that they did not realize that they were given a piece of useful information, namely that $P(X=0) = .01$, and thus they could not figure out how to solve for λ. In addition, the fact that 77% of these unsuccessful subjects did not even realize they were solving for λ indicates that they did not recognize solving for λ as a subgoal, but rather were looking to apply the operations from the examples. This claim is further supported by the fact that half of the subjects tried to make up an observed frequency table from which to calculate λ. However, most of these subjects still went on to calculate an expected frequency table. This suggests that they did not make up the observed frequency table to calculate λ per se, but rather the table was created to help them apply the stereotyped operations so they could reach the only goal they seemed to know: to create an expected frequency table.

Experiment 2

Experiment 1 indicated that manipulations of superficial problem characteristics were not sufficient to induce subjects to isolate subgoals and methods. In Experiment 2 we manipulated the subgoals and methods used in the training problems.

Subjects were given four problems to study. The ONE-PROCEDURE group was just like the DIFFERENT group in Experiment 1: the problems used the same procedure but were different superficially. The TWO-PROCEDURE group received two problems using the same procedure as the Detroit Tiger problem and two problems using the same procedure as the cookie problem. It would not be surprising if the TWO-PROCEDURE subjects could solve the cookie problem successfully. However, the more interesting issue is whether they learned anything more than two sets of operations for solving two types of problems. That is, did they simply learn that frequency table problems require one approach and non-frequency table problems require a different approach (i.e., they learned a superficial selection rule and did not learn subgoals or methods), or did they learn that problems can have different goals, subgoal orders, and methods for obtaining those subgoals?

Subjects then attempted to solve two instances of a new problem type (the birthday problem and one isomorphic to it, the "football" problem—not illustrated here) in addition to problems whose solution procedures were already familiar to them (i.e., the Detroit Tiger problem and/or the cookie problem). Subjects' answers and errors were examined for indications that they were simply trying to apply one of two series of operations or whether they had recognized that particular subgoals existed (finding λ, then finding $P(X=x)$) and that new methods would be needed.

Method

Subjects. Fifty students from a probability class were recruited and paid $7 for their participation.

Materials and Procedure. The procedure was identical to the one in Experiment 1. The only difference was the materials. There were three groups of subjects in this experiment. The TIGER group studied four training problems that used the same solution procedure as the Detroit Tiger target problem. The COOKIE group studied four training problems which used the same solution procedure as the cookie target problem. The TWO-PROCEDURE group studied two problems which used the Detroit Tiger problem procedure and two problems which used the cookie problem procedure. All subjects then received four target problems to solve: the Detroit Tiger problem, the cookie problem, the birthday problem, and the football problem.

Results and Discussion

For some of the analyses reported below, the comparisons are between the three groups: TIGER, COOKIE, and TWO-PROCEDURE. For other analyses the TIGER and COOKIE groups are collapsed into a ONE-PROCEDURE group and thus the comparison will be between ONE-PROCEDURE and TWO-PROCEDURE subjects. In addition, the terms "near" and "far" transfer can not be used as they were in Experiment 1 since, for the COOKIE group, the cookie problem is now a near transfer problem and the Detroit Tiger problem is a far transfer problem. Thus, the target problems will be referred to by their names. Table 2 summarizes the type of transfer problem the target problems represent for each group.

Table 2

Degree of Transfer Required in Target Problems as a Function of Subject Group

Group	Near Transfer:	Far Transfer:
TIGER	Detroit Tiger	cookie, birthday, football
COOKIE	cookie	Detroit Tiger, birthday, football
TWO-PROCEDURE	Detroit Tiger, cookie	birthday, football

While all of the TIGER and TWO-PROCEDURE subjects solved the Detroit Tiger problem correctly, only 13% of the COOKIE subjects did. This difference is, of course, significant, $G^2(2) = 45.5$, $p < .0001$. Similarly, while most of the COOKIE and TWO-PROCEDURE subjects solved the cookie problem correctly (87% and 86%, respectively), a much lower percentage (31%) of the TIGER subjects did, $G^2(2) = 13.9$, $p < .001$. There is no difference in solution rates among the three groups for the birthday or football problems which are far transfer problems for all subjects. Overall, 52% of the subjects solved the birthday problem and 50% solved the football problem. It should be noted that the 52% solution rate for the birthday problem is significantly greater than the 28% solution rate for that problem for subjects in Experiment 1, $z = 2.7$, $p < .007$.

Of the nine TIGER subjects who failed to solve the cookie problem, 78% failed to realize that the goal was to solve for λ, 100% did not realize that $P(X=0) = .01$ was provided in the problem, and 33% tried to make up a frequency table as an aid to solve the problem. These frequencies are similar to the ones obtained in Experiment 1.

Of the 13 COOKIE subjects who failed to solve the Detroit Tiger problem, 12 of them tried to calculate λ by taking an observed frequency for some X and plugging that into the Poisson equation and solving for λ. For those subjects who chose $X=0$, they would get an equation such as $P(X=0) = 85/162 = [e^{-\lambda}\lambda^0]/0!$. This reduces to $.52 = e^{-\lambda}$, which yields $\lambda = .65$. Given that the λ generated by the frequency table method is .66, this "cookie" approach works quite well, but in other situations it could be quite poor in comparison with the frequency table method (since it would ignore available frequency data). In addition, for the 12 subjects who took this "cookie" approach, eight of them stopped after solving for λ and did not generate the predicted frequency table. This suggests that they were performing a series of operations rather than solving for the goal of the problem. Four of the other subjects used the observed frequencies of each X in turn to solve for λ. It becomes quite messy to solve for λ when an X other than 0 is used and these subjects would set up the equations and then stop. The remaining subject who got the problem wrong calculated λ using the frequency table approach, but did not go on to generate predicted frequencies for the various values of X.

The types of errors made by subjects who were unsuccessful in solving the birthday or the football problems are presented in Table 3. The errors are presented as a function of whether subjects received examples illustrating one procedure or two (i.e., the TIGER and COOKIE groups are collapsed into the ONE-PROCEDURE group).

Table 3

Percentage of Subjects Who Made Particular Types of Errors (Experiment 2)

Transfer Problem	Error Type	Group: One Procedure	Group: Two Procedure	Total
Birthday		n=14	n=10	n=24
	SLAMBDA	93 (13)	100 (10)	96 (23)
	PX3	50 (7)	10 (1)	33 (8)
	FREQ	7 (1)	20 (2)	12 (3)
Football		n=14	n=11	n=25
	SLAMBDA	93 (13)	100 (11)	96 (24)
	PX1	43 (6)	9 (1)	28 (7)
	FREQ	14 (2)	0 (0)	8 (2)

Note. Frequencies are given in parentheses. Percentages are based on the number of subjects who made a particular error divided by the number of subjects in each group who got the problem wrong (given at the top of each column for each of the transfer problems), not the total of number of subjects in the group.

Both ONE-PROCEDURE and TWO-PROCEDURE subjects solved the birthday and football problems about 50% of the time. These are far transfer problems for both groups. We had expected the TWO-PROCEDURE subjects to do better since we hypothesized they would have been likely to isolate subgoals such as λ and $P(X=x)$ and generalize the methods for finding them. Nevertheless, one difference did emerge in both problems. Of the ONE-PROCEDURE subjects who failed to solve the birthday problem, only 50% realized they were to solve for $P(X=3)$ while 90% of the TWO-PROCEDURE subjects realized this. This difference is significant, $G^2(1) = 4.64$, $p < .04$. Similarly, 50% of the ONE-PROCEDURE subjects realized they were to solve for $P(X=1)$ in the football problem while 91% of the TWO-PROCEDURE subjects realized this. Again, the difference is significant, $G^2(1) = 3.82$, $p = .05$. This result suggests that TWO-PROCEDURE subjects may have at least isolated subgoals, but were unable to apply the correct method to the birthday and football problems. Most subjects did calculate λ in the birthday and football problems, but they tended to use nonsensical values such as 365/500 or 3/500 for the birthday problem. TWO-PROCEDURE subjects did not seem to learn anything about examining λ for its reasonableness, yet they did adapt to the new goal constraint (i.e., finding only a particular $P(X=x)$) while ONE-PROCEDURE subjects did not.

General Discussion

The difficulties that subjects in both experiments had with the far transfer problems suggest that procedural variety plus explicit pointing out of subgoals and methods may be required to teach students how to solve problems which have different subgoal orders and modified methods compared to training problems.

Procedural variety may mean that students should be exposed to problems that provide different givens, have different appearances, and/or which require solving for different unknowns.

These variations would presumably induce students to isolate different methods for achieving certain subgoals and to realize that there can be different goals and subgoals for solving problems in the same domain (Owen & Sweller, 1985). This induction could also be facilitated by presenting examples which give the data in different forms (such as giving λ directly rather than having it calculated from a table). The need for having students see examples which solve for different unknowns is suggested by the large number of subjects in Experiment 1 who failed to realize that they were solving for something new, namely λ, in the cookie problem. The importance of presenting similar information in different forms (e.g., tables versus text, ready-to-use values versus "low-level" values which require additional calculations before they can be used in equations) seems reasonable in light of the fact that λ was a quite simple thing to calculate in the birthday problem, yet students failed to see it or to calculate it correctly. In fact, students in Experiment 1 often used the more laborious method of making up a frequency table in order to (incorrectly) calculate λ. This problem is similar to the error Reed et al.'s (1985, Experiment 4) subjects made when they tried to use the more complex solution methods from the training examples on the simpler target problems. Both our results and Reed et al.'s indicate that students were learning series of operations rather than, or more easily than, subgoals and methods for solving problems.

We have tried to suggest that an important component of the "power" of examples is the variation that is provided in a sequence of examples. Winston's (1973) arch perceiver could only learn concepts when the examples it was presented with were given in a particular order. Negative instances of a concept were just as important (and sometimes more important) than positive instances. Failure-driven memory is an important component of Schank's (1982) model of learning. So too here, negative examples (in the form of training problems that have different subgoals and methods) are important. If a student sees several problems that are dealt with in different ways, he or she may be more likely to isolate the subgoals and methods rather than viewing the problems as a series of operations which ultimately produce some output. He or she may also form generalizations of methods. However, the student may need guidance to help him or her focus on the subgoals and methods, at least initially (Lewis & Anderson, 1985). We are currently conducting a transfer experiment using materials which provide subjects with explanatory information highlighting the subgoals and methods that are present in each training example.

It may be possible to develop a methodology for constructing examples for textbooks in quantitative domains. This methodology would involve first identifying the subgoals and methods that students need to learn (Kieras, in press; Kieras & Bovair, 1986). Then example problems and explanatory materials which highlight these subgoals and methods can be constructed. The careful procedural variation might allow students to see beyond the superficial features of examples.

Acknowledgments

This research was supported by Army Research Contract MDA903-86-K-0297. We would like to thank J.E. Keith Smith and Mary Gick for their comments on this paper.

References

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

Gentner, D. & Gentner, D.R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A.L. Stevens (Eds.), Mental Models. Hillsdale, N.J.: Erlbaum.

Gick, M.L. & Holyoak, K.J. (1980). Analogical problem solving. Cognitive Psychology, 12, 306–55.

Gick, M.L. & Holyoak, K.J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15, 1-38.

Greeno, J.G., Riley, M.S., & Gelman, R. (1984). Conceptual competence and children's counting. Cognitive Psychology, 16, 94-143.

Hayes, J.R., Waterman, D.A., & Robinson, C.S. (1977). Identifying the relevant aspects of a problem text. Cognitive Science, 1, 297-313.

Kieras, D.E. (in press). The role of cognitive simulation models in the development of advanced training and testing systems. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), Diagnostic monitoring of skill and knowledge acquisition. Hillsdale, NJ: Erlbaum.

Kieras, D.E. & Bovair, S. (1986). The acquisition of procedures from text: A production-system analysis of transfer of training. Journal of Memory and Language, 25, 507-524.

Larkin, J., McDermott, J., Simon, D.P., & Simon, H. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

Lewis, M.W., & Anderson, J.R. (1985). Discrimination of operator schemata in problem solving: Learning from examples. Cognitive Psychology, 17, 26-65.

Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? Journal of Educational Psychology, 77, 272-284.

Reed, S.K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 106-125.

Schank, R.C. (1982). Dynamic memory: A theory of reminding and learning in computers and people. New York: Cambridge University Press.

Schoenfeld, A.H., & Herrmann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 484-494.

Spencer, R.M., & Weisberg, R.W. (1986). Context-dependent effects on analogical transfer. Memory & Cognition, 14, 442-449.

Tenney, Y.J., & Gentner, D. (1984). What makes analogies accessible: Experiments on the water-flow analogy for electricity. In Proceedings of the international conference on research concerning students' knowledge of electricity, Germany.

VanLehn, K. (1985). Arithmetic procedures are induced from examples, (Tech. Report No. ISL-12). Xerox Palo Alto Research Center.

Winston, P. (1973). Learning to identify toy block structures. In R.L. Solso (Ed.), Contemporary issues in cognitive psychology: The Loyola symposium. Washington, D.C.: V.W. Winston & Sons. Hillsdale, N.J.: Erlbaum.