The Effects of Labels in Examples on Problem Solving Transfer

Richard Catrambone
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332
(404) 894-2682
richard.catrambone@psych.gatech.edu

Abstract

It is hypothesized that labels in examples help learners group a set of steps and to try to explain why those steps belong together. The result of these grouping and self-explanation processes might be the formation of a subgoal. It is conjectured that the meaningfulness of the label itself might not be critical in order for the grouping and self-explanation processes to occur. This conjecture is supported in an experiment in which subjects studying examples in probability that had steps labeled transferred to novel problems more successfully than subjects whose examples did not contain labels. Furthermore, subjects who saw less meaningful labels transferred as successfully as subjects studying examples with more meaningful labels. Thus, it appears that the *meaningfulness* of the label does not seem to affect subgoal formation as much as the *presence* of a label. This result supports the interpretation that subgoal learning is affected by labels and that labels produce this benefit by helping learners group the steps into a purposeful unit, perhaps through a self-explanation process.

Introduction

Learners typically prefer to use examples to help them solve novel problems, yet they have great difficulty generalizing from examples in order to solve those problems (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Pirolli & Anderson, 1985; Reed, Dempster, & Ettinger, 1985). One source of this difficulty is a tendency by learners to focus on features of examples that are superficial, that is, features that can vary from an example to a problem without affecting the solution procedure (Ross, 1987, 1989). Learners will incorrectly interpret superficial changes as cues for procedural changes and superficial similarity as cues for procedural similarity. Conversely, learners often fail to detect components of examples' solutions that represent the domain theory and thus, *are* invariant across examples and problems (Catrambone and Holyoak, 1990, Reed, Ackinclose, & Voss, 1990; Reed et al., 1985).

These invariant components might fruitfully be considered subgoals. "Subgoal" is used here to represent the task structure to be learned for solving problems in a particular domain. It is assumed that these subgoals can be taught to learners (e.g., Catrambone, in press; Catrambone & Holyoak, 1990; Dixon, 1987; Eylon & Reif, 1984). On this view, a subgoal groups a set of steps under a meaningful task or purpose (e.g., Anzai & Simon, 1977; Chi & VanLehn, 1991). For instance, in the probability materials used in the experiment, a set of multiplication and addition steps can be grouped under the subgoal "find the total frequency of the event."

Learners appear to be more successful solving novel problems when they learn the goal structure of the problems they will be solving (e.g., Anzai & Simon, 1979; Brown, Kane, & Echols, 1986; Eylon & Reif, 1984). Subgoals appear to be useful because they provide a purpose for a set of steps. Thus, when the learner is working on a novel problem, the learner can use the subgoals to guide him- or herself to the steps from the old solution procedure that need to be modified in order to achieve the subgoals in the new problem. In contrast, a learner who had simply memorized a string of steps for solving a particular problem type, without grouping sets of steps under the subgoals they achieve, will have fewer cues to direct him or her to the steps that need modification for the novel problem.

Learners might be more likely to learn subgoals when steps for achieving those subgoals are labeled. Labels may perform at least two functions related to subgoal learning. One is to group a set of steps. This grouping might lead a learner to construct a purpose for the steps since the label is implicitly informing the learner that the steps form a coherent unit. A second possible function of a label is to provide a mini-explanation of the steps' purpose. The explanation might be crucial in order for a learner to form a domain-meaningful subgoal. The present experiment was designed to test the importance of the meaningfulness of a label in subgoal formation.

The Test Domain: The Poisson Distribution

The Poisson distribution is often used to approximate binomial probabilities for events occurring with some small probability. The Poisson equation is $P(X{=}x) = \frac{\left[\left(e^{-\lambda}\right)\left(\lambda^{x}\right)\right]}{x!}$, where λ is the average (the expected value) of the random variable X.

Table 1 shows a use of the Poisson distribution to predict the probability of a randomly chosen lawyer owning a certain number of briefcases. The method for the goal of finding λ, the average frequency of the event (e.g., owning a briefcase), could be represented as shown in Table 2a.

Table 1: Training Example with Meaningful, Less-Meaningful, and No Label Solutions

A judge noticed that some of the 219 lawyers at City Hall owned more than one briefcase. She counted the number of briefcases each lawyer owned and found that 180 of the lawyers owned exactly 1 briefcase, 17 owned 2 briefcases, 13 owned 3 briefcases, and 9 owned 4 briefcases. Use the Poisson distribution to determine the probability of a randomly chosen lawyer at City Hall owning exactly two briefcases.

a.) *Meaningful Label Solution*:

Total number of briefcases owned = [1(180) + 2(17) + 3(13) + 4(9)] = 289

$E(X) = \frac{289}{219} = 1.32 = \lambda$ = average number of briefcases owned per lawyer

$$P(X{=}x) = \frac{\left[\left(e^{-\lambda}\right)\left(\lambda^{x}\right)\right]}{x!}$$

$$P(X{=}2) = \frac{[(2.718^{-1.32})(1.32^{2})]}{2!} = \frac{(.27)(1.74)}{2} = .235$$

b.) *Less-Meaningful Label Solution*:

Ω = [1(180) + 2(17) + 3(13) + 4(9)] = 289

$E(X) = \frac{289}{219} = 1.32 = \lambda$ = average number of briefcases owned per lawyer

{Remainder of solution identical to Meaningful Label solution}

c.) *No Label Solution*:

$E(X) = \frac{1(180) + 2(17) + 3(13) + 4(9)}{219} = \frac{289}{219} = 1.32 = \lambda$ = average number of briefcases owned per lawyer

{Remainder of solution identical to Meaningful Label Solution}

Table 2: Representations of the Weighted-Average Method for Finding λ

a.) *Goal:* Find λ
Method: 1) multiply each event-category (e.g., owning exactly 0 briefcases, owning exactly 1 briefcase, etc) by its observed frequency
2) sum the results
3) divide the sum by the total number of trials (number of lawyers) = average number of briefcases per lawyer

b.) *Goal*: Find λ
Method: 1) *Goal*: Find total number of briefcases
Method: a) multiply each event-category by its observed frequency
b) sum the results = total number of briefcases
2) divide total number of briefcases by the total number of trials = average number of briefcases per lawyer

Table 3: Sample Test Problems

a.) *Total Frequency Provided Directly*

A number of celebrities were asked how many commercials they made over the last year. The 20 celebrities made a total of 71 commercials. Use the Poisson distribution to determine the probability that a randomly chosen celebrity made exactly 5 commercials.

Solution (not seen by subjects):

$$E(X) = \frac{71}{20} = 3.55 = \lambda = \text{average number of commercials per celebrity}$$

$$P(X=5) = \frac{[(2.718^{-3.55})(3.55^5)]}{5!} = \frac{(.029)(563.8)}{120} = .135$$

b.) *Total Frequency Calculated by Adding Simple Frequencies*

Over the course of the summer, a group of 5 kids used to walk along the beach each day collecting seashells. We know that on Day 1 Joe found 4 shells, on Day 2 Sue found 2 shells, on Day 3 Mary found 5 shells, on Day 4 Roger found 3 shells, and on Day 5 Bill found 6 shells. Use the Poisson distribution to determine the probability of a randomly chosen kid finding 3 shells on a particular day.

Solution (not seen by subjects):

$$E(X) = \frac{4+2+5+3+6}{5} = \frac{20}{5} = 4.0 = \lambda = \text{average number of shells per kid}$$

$$P(X=3) = \frac{[(2.718^{-4.0})(4.0^3)]}{3!} = \frac{(.018)(64)}{6} = .195$$

Learners are good at memorizing, from examples, the steps for the weighted-average method in order to achieve the subgoal of finding the average frequency of an event for problems with different story-lines (Catrambone & Holyoak, 1990). However, they often fail to notice that steps #1 and #2 in the method in Table 2a could also be viewed as a method for achieving the subgoal of finding the total frequency of the event (e.g., total number of briefcases owned). As a result, learners often have trouble finding the average frequency of an event when a problem provides the total frequency of that event directly (see Table 3a) or requires the total frequency to be calculated in a new way (see Table 3b) (Catrambone, under review; Catrambone & Holyoak, 1990).

Catrambone and Holyoak (1990) demonstrated that if a person learns the subgoal "find the total frequency of the event," he or she is better able to find λ in a novel problem that requires a change from the examples in how total frequency is found. For example, in the method discussed earlier for finding the average number of briefcases owned per lawyer, transfer is improved if the learner's method for finding λ is organized as shown in Table 2b.

Experiment

This experiment examined subgoal learning by manipulating the meaningfulness of the label for a set of steps. The No Label group studied examples demonstrating the weighted average method for finding λ (see Table 1, the "No Label" solution). The Meaningful Label and Less-Meaningful Label groups' examples differed in that the steps for finding the total frequency were explicitly labeled rather than merged with the overall set of steps for finding λ (see Table 1, the "Meaningful Label" and "Less-Meaningful Label" solutions).

Subgoal learning was assessed in two ways. The first was to analyze transfer performance--how successfully subjects found λ--on novel problems. The second was to have these subjects describe how to solve problems in the Poisson domain. If "label" subjects' descriptions mention the labels used in the training examples, this would provide one piece of evidence that labels make learners more likely to group a set of steps. If these subjects solve the novel problems more successfully than the No Label subjects, this would provide converging evidence that label subjects learned a subgoal. If the Meaningful Label subjects outperform the Less-Meaningful Label subjects, this would suggest that the meaningfulness of the label also affects the likelihood of forming a domain-meaningful subgoal.

No difference was predicted among the groups in the frequency of mentioning the notion of finding an average (or λ) in their descriptions since all examples mention the term "average" in the solutions. Finally, all groups were predicted to be quite successful solving transfer problems that are isomorphic to the training examples.

Method

Subjects. Subjects were 100 students recruited from an introductory psychology class at the Georgia Institute of Technology who received course credit for their participation. None of the subjects had taken a probability course prior to participating in the experiment.

Materials and Procedure. Subjects were randomly assigned to one of three groups. The Meaningful Label group (n = 34) studied three examples demonstrating the weighted average method for finding λ in which the steps for finding the total frequency were explicitly labeled with a label that presumably had meaning to the subjects and made mathematical sense given the steps that preceded it. The labels seen by the Meaningful Label group were specific to the context of the problem (e.g, "total number of briefcases owned"), and not phrased at a general level (e.g., "total frequency") in order to minimize additional explicit general domain instruction. The Less-Meaningful Label group (n = 34) studied examples in which the steps for finding the total frequency were labeled with Ω which presumably had little meaning for the subjects in the context of the examples. The No Label group (n = 32) studied examples in which the steps for finding the total frequency were not labeled.

After studying the examples, subjects were asked to write explanations of how they would teach someone to solve problems of the type they had just studied. It was conjectured that these explanations might provide additional information about the organization of subjects' problem solving knowledge after having studied the examples. Subjects' explanations of how to solve problems in the domain were scored for two primary features: an explicit mention of trying to find the total frequency (and/or Ω in the case of the Less-Meaningful group), and an explicit mention of trying to find an average. Two raters independently scored the explanations and problem solutions and agreed on scoring 95% of the time. Disagreements were resolved by discussion.

After writing their explanations, subjects solved six test problems. The first two required the use of the weighted average method for finding λ (isomorphic to the example in Table 1). The third and fourth problems provided the total frequency directly (and thus λ could be found by simply dividing the given number by the total number of trials). The latter type included the problem in Table 3a and another problem isomorphic to it. The fifth and sixth problems involved adding simple frequencies in order to find the total frequency (see Table 3b for an example). Subjects were told not to look back at the examples when solving the test problems or writing their explanations.

Subjects' written solutions were scored for whether they found λ correctly. The frequencies with which the groups correctly found λ and mentioned the subgoals were analyzed using the likelihood ratio chi-square test (G^2; Bishop, Fienberg, & Holland, 1975).

Results

Transfer as a Function of Group. As expected, all three groups did quite well at finding λ on test problems that were isomorphs to the training examples (see Table 4, Problems 1-2). There was no significant difference among the groups on either of the isomorphic problems.

As predicted, there was a significant difference among the groups at finding λ successfully on the four test problems that involved the new method for finding λ (see Table 4, problems 3-6), $G^2(2) = 9.72$, $p = .008$; $G^2(2) = 8.47$, $p = .01$; $G^2(2) = 11.14$, $p = .004$; $G^2(2) = 18.80$, $p = .0001$, for problems 3-6, respectively. The most typical mistake that subjects made on problems #3 and #4, in which the total frequency was given directly in the problem, was to write in the solution area that not enough information was given to solve the problem. This mistake was also made frequently in problems #5 and #6 as well as the mistake of taking each value in the problem and multiplying it by a number from an increasing sequence (i.e., 1, 2, 3, etc), perhaps as an attempt to mimic the procedure from the training examples.

The Meaningful Label group had the highest percentage correct for each of the novel problems and the No Label group had the lowest percentage correct. Pairwise comparisons indicate that the Less-Meaningful Label group outperformed the No Label group on each of these problems, $G^2(1) = 3.87$, $p = .049$; $G^2(1) = 3.92$, $p = .048$; $G^2(1) = 3.86$, $p = .05$; $G^2(1) = 6.98$, $p = .008$, for problems 3-6, respectively. The Meaningful Label group never significantly outperformed the Less-Meaningful Label group, $G^2(1) = 1.32$, $p = .25$; $G^2(1) = 0.77$, $p = .38$; $G^2(1) = 1.98$, $p = .16$; $G^2(1) = 3.17$, $p = .08$, for problems 3-6, respectively.

Explanations as a Function of Group. Subjects' explanations were scored for whether they explicitly mentioned the subgoal of finding the total frequency (and/or Ω for the Less-Meaningful Label group) and the subgoal of finding the average or λ. As expected, the groups were equally likely to mention finding the average (see Table 5), $G^2(2) = 2.17$, $p = .34$, presumably because the average was labeled for all conditions.

The groups significantly differed in the frequency of mentioning finding total frequency in their explanations (see Table 5), $G^2(2) = 13.87$, $p = .001$. Interestingly, the combined rate of mentioning either total frequency or Ω by the Less-Meaningful Label group was the same as the rate of mentioning total frequency by the Meaningful Label group. This suggests that the Less-Meaningful Label group appeared to learn a subgoal, articulated as "total frequency" or "Ω" for the steps for finding total frequency as often as the Meaningful Label group. In contrast, only 16% of the No Label subjects seemed to form a subgoal for these steps. It could be conjectured that some No Label subjects provided idiosyncratic terms for relating the steps for finding total frequency and that these terms were ignored in the subgoal-scoring process. In fact though this never appeared to happen; No Label subjects typically provided the steps without a statement about what they accomplished.

Transfer as a Function of Explanations. As predicted, there were no significant performance differences between subjects who mentioned finding λ and those who did not on any of the test problems (see Table 6), $\underline{G}^2(1) = 0$; $\underline{G}^2(1) = 0.95$, $\underline{p} = .33$; $\underline{G}^2(1) = 1.06$, $\underline{p} = .30$; $\underline{G}^2(1) = 0.07$, $\underline{p} = .79$; $\underline{G}^2(1) = 0.11$, $\underline{p} = .74$; $\underline{G}^2(1) = 0.23$, $\underline{p} = .63$, for problems 1-6, respectively.

The comparison among subjects as a function of whether or not they mentioned total frequency in their explanations produced performance differences on all four novel test problems (see Table 6, problems 3-6), $\underline{G}^2(1) = 13.60$, $\underline{p} = .0002$; $\underline{G}^2(1) = 8.34$, $\underline{p} = .004$; $\underline{G}^2(1) = 11.64$, $\underline{p} = .0006$, $\underline{G}^2(1) = 17.69$, $\underline{p} = .0001$, for problems 3-6, respectively. Pairwise analyses (not presented here) show that subjects who mentioned only Ω did not perform significantly differently on these problems than subjects who mentioned neither total frequency or Ω. However, only 11 subjects mentioned Ω (without also mentioning total frequency), thus comparisons with this group are difficult to interpret.

Table 4: Percentage of Subjects Finding λ Correctly on Test Problems as a Function of Group

	Group			
	Meaningful Label (n = 34)	Less-Meaningful Label (n = 34)	No Label (n = 32)	AVG
Problem #1 (weighted avg.)	100	100	100	100
Problem #2 (weighted avg.)	100	97	97	98
Problem #3 (total freq. direct)	82	71	47	67
Problem #4 (total freq. direct)	82	74	50	69
Problem #5 (simple frequency)	82	68	44	65
Problem #6 (simple frequency)	97	85	56	80
AVG	90	82	66	80

Table 5: Percentage of Subjects Mentioning Features in Explanations as a Function of Group

	Group		
	Meaningful Label	Less-Meaningful Label	No Label
Mentions λ	71	82	84
Mentions Total Frequency	56	24	16
Mentions Ω	0	32	0

Table 6: Percentage of Subjects Finding λ Correctly on Test Problems as a Function of Explanations

	Feature Mentioned				
	Total Frequency (n=32)	Ω (n=11)	Neither (n=57)	λ (n=79)	Not λ (n=21)
Problem #1 (weighted avg.)	100	100	100	100	100
Problem #2 (weighted avg.)	100	100	96	98	100
Problem #3 (total freq. direct)	91	54	56	65	76
Problem #4 (total freq. direct)	88	54	61	68	71
Problem #5 (simple frequency)	88	46	56	66	62
Problem #6 (simple frequency)	100	82	68	81	76

Discussion

The transfer performance of the three training groups is consistent with the hypothesis that the presence of a label helps learners group a set of steps into a meaningful unit. This unit might be characterized as a subgoal. Even a relatively meaningless label aids this grouping function, possibly because it encourages learners to retrieve information from long-term memory in order to explain why a set of steps belong together. The transfer performance as a function of explanations demonstrated that subjects who mentioned finding total frequency transferred better to novel problems. This result is also consistent with the prediction that subjects learning the subgoal to find total frequency would perform better on novel problems than subjects who did not learn that subgoal.

It appears that the grouping effect of labels, not their domain-meaningfulness, is most responsible for subgoal learning, at least for domains in which the learner has relevant prior knowledge. That is, it is suggested that a label, even one with little meaning, leads learners to attempt to explain to themselves why the steps go together. This self-explanation leads to subgoal formation (Chi et al., 1989). Thus, the presence of the label can be seen as a way of aiding the self-explanation process.

If a person is learning to solve problems in a domain for which he or she has little relevant prior knowledge, then perhaps labels that are relatively meaningless will not aid the learner in forming useful subgoals. In those cases, meaningful labels will presumably produce better subgoal learning since the extra domain information provided by a meaningful label will help the learner make sense of the steps and to understand their purpose.

The experiment did not directly test the assumption that when learners are led to group steps they then try to explain to themselves why the steps go together. This assumption could be examined by having subjects talk out loud about what they are learning when they study examples that vary in the use of labels. If data from this approach suggest that even a relatively meaningless label leads learners to try to self-explain why a series of steps are grouped, and if these subjects show superior transfer for the subgoal represented by the grouping, then additional support would be gained for the connections between labeling, grouping, self-explanation, and subgoal learning.

Acknowledgements

This research was supported by Office of Naval Research Grant N00014-91-J-1137.

References

Anzai, Y., & Simon, H.A. (1979). The theory of learning by doing. *Psychological Review, 86(2)*, 124-140.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.

Brown, A.L., Kane, M.J., & Echols, C.H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development, 1*, 103-121.

Catrambone, R. (in press). Improving examples to improve transfer to novel problems. *Memory & Cognition*.

Catrambone, R. (under review). *Using labels to aid subgoal learning: Effects on transfer*. Manuscript submitted for publication.

Catrambone, R., & Holyoak, K.J. (1990). Learning subgoals and methods for solving probability problems. *Memory & Cognition, 18(6)*, 593-603.

Chi, M.T.H., Bassok, M., Lewis, R., Reimann, P., & Glaser, R. (1989). Self explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M.T.H., & VanLehn, K. (1991). The content of physics self-explanations. *Journal of the Learning Sciences, 1(1)*, 69-106.

Dixon, P. (1987). The processing of organizational and component step information in written directions. *Journal of Memory and Language, 26(1)*, 24-35.

Eylon, B., & Reif, F. (1984). Effects of knowledge organization on task performance. *Cognition and Instruction, 1(1)*, 5-44.

Pirolli, P.L., & Anderson, J.R. (1985). The role of learning from examples in the acquisition of recursive programming skill. *Canadian Journal of Psychology, 39(2)*, 240-272.

Reed, S.K., Ackinclose, C.C., & Voss, A.A. (1990). Selecting analogous problems: Similarity versus inclusiveness. *Memory & Cognition, 18(1)*, 83-98.

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(1)*, 106-125.

Ross, B. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(4)*, 629-639.

Ross, B. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(3)*, 456-468.