Journal of Experimental Psychology:
Learning, Memory, and Cognition
1989, Vol. 15, No. 6, 1147-1156

Copyright 1989 by the American Psychological Association, Inc.
0278-7393/89/$00.75

Overcoming Contextual Limitations on Problem-Solving Transfer

Richard Catrambone
Georgia Institute of Technology

Keith J. Holyoak
University of California, Los Angeles

Five experiments investigated transfer from multiple analogs to a superficially dissimilar target problem. When subjects explicitly compared the analogs and then immediately attempted to solve the target problem in the context of a single experiment, transfer was obtained with significant frequency even without a hint that the analogs and target were related. Prehint transfer was sharply reduced or eliminated when the source analogs and the target were presented in different contexts, even when the transfer test was immediate. However, prehint transfer was enhanced, even after a context shift and a week-long delay between reading the source analogs and solving the problem, when the following conditions were met: The target problem was reworded slightly to emphasize a structural feature that it shared with the analogs; three rather than two source analogs were provided; and detailed, schema-oriented questions were used to help subjects focus on the problem-relevant aspects of the stories. Although spontaneous transfer between small numbers of dissimilar analogs is difficult to obtain, it can be achieved by manipulations that foster abstraction of a problem schema from the training examples.

The conditions under which people are able to recognize and exploit analogies in the course of problem solving has been the focus of extensive investigation in recent years. Several studies have shown that novice problem solvers in domains such as statistics, computer programming, and word-processing often use analogies between example problems as guides (e.g., Catrambone & Carroll, 1987; Catrambone & Holyoak, 1986; Pirolli & Anderson, 1985; Reed, 1987; Ross, 1984). Spontaneous analogical transfer is much less frequent, however, when the source and target analogs are drawn from different domains and are superficially dissimilar (Gick & Holyoak, 1980; Hayes & Simon, 1977; Reed, Ernst, & Banerji, 1974; Ross, 1987). Gick and Holyoak (1980) found that interdomain transfer often depended on provision of a hint to use the prior analog. Problem solvers often failed to notice the relevance of an analogy that they were readily able to apply once it was called to their attention. Spontaneous transfer can be obtained, however, even after a delay of several days between presentation of the source and the target, if the source and target have at least one salient surface similarity (Holyoak & Koh, 1987).

A number of studies have shown that interdomain transfer can be facilitated by manipulations designed to encourage the formation of generalized rules or schemas. The most direct manipulation of this sort involves explicit instruction in abstract rules, coupled with examples, in domains such as statistics and algebra. Such abstract training can produce substantial rates of spontaneous transfer (e.g., Bassok & Holyoak, 1989; Fong, Krantz, & Nisbett, 1986). Other research has shown that manipulations that emphasize the problem-solving methods or actions in a domain, rather than a means–ends strategy of reducing the differences between an initial state and a goal state, are effective in helping students acquire mental structures that allow them to classify and solve new problems (Lewis & Anderson, 1985; Sweller, Mawer, & Ward, 1983).

There is also evidence that provision of multiple examples without explicit instruction in generalized rules can facilitate transfer (Gick & Holyoak, 1983). The use of multiple examples could lead to the creation of a generalized rule or schema that can be applied to a target problem more successfully than could the examples from which it was formed. It is also possible that the schema could improve the access and application of the examples. The research reported here was designed to investigate boundary conditions on facilitation of transfer attributable to prior inductive generalization from examples.

This study is a direct extension of the work of Gick and Holyoak (1983). Those investigators had subjects first read two stories illustrating problems that were solved by a "convergence" procedure (e.g., a story about a general who captured a fortress by sending small groups of men down several converging roads, and another about a fire chief who put out a blaze by having many buckets of water thrown at the fire at once by a circle of fire fighters). The stories were presented in the guise of an experiment on story comprehension; subjects wrote summaries of each story and also wrote descriptions of the similarities between the two. The latter task was intended to encourage induction of a schematic representation capturing the relatively abstract similarities between the two source analogs. Immediately afterward, the subjects attempted to

---

Experiment 1 was presented at the 93rd annual meeting of the American Psychological Association, Los Angeles, August 1985. Experiment 3 was presented at the 59th meeting of the Midwestern Psychological Association, Chicago, May 1987. Experiments 4 and 5 were presented at the 29th meeting of the Psychonomic Society, Chicago, November 1988.

This research was supported by Army Research Contract MDA903-86-K-0297.

We thank Michael Barron and Gordon Kato for their able assistance in running subjects in Experiments 1 and 2. Mary Gick, Laura Novick, Dorrit Billman, and several reviewers provided helpful comments on an earlier draft of the article.

Correspondence concerning this article should be addressed to Richard Catrambone, School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332.

solve Duncker's (1945) "radiation" problem (in which a doctor must find a way to use rays to destroy a patient's stomach tumor without injuring the surrounding healthy tissue), first without any hint that the prior stories were relevant to the task, and then again after receiving a hint to use the stories. Transfer (i.e., the production of the convergence solution to the radiation problem) before a hint was hypothesized to indicate subjects' ability to spontaneously *notice* and *apply* the analogy, whereas transfer after a hint indicated subjects' ability to apply the analogy once its relevance had been suggested. Thus, after-hint transfer should be less difficult than before-hint transfer because one of the steps—spontaneously noticing the relevance of an analogy—is no longer required.

Gick and Holyoak (1983) found that provision of two analogs rather than one produced significantly greater transfer, both prior to the hint and in total (including hint-aided solutions). Moreover, for subjects who received two source analogs, the likelihood of transfer was predictable from the content of the written descriptions of the story similarities. These descriptions were classified into three levels of "schema quality," reflecting the extent to which they included key aspects of the convergence solutions as similarities. The likelihood of both immediate and hint-aided transfer increased with schema quality. Gick and Holyoak interpreted these findings as evidence that prior induction of a schematic representation of a problem category facilitates transfer (cf. Chi, Feltovich, & Glaser, 1981).

In the experiments of Gick and Holyoak (1983), as in most other studies of transfer, the target problem was presented to subjects immediately after they had studied the source analogs, without either an intervening delay or a change in the context (beyond the obvious shift from a part of the experiment involving story processing to a part involving problem solving). For both theoretical and pedagogic reasons, of course, it is important to establish the conditions under which transfer can be obtained even after more substantial contextual shifts. From a theoretical perspective, such studies are likely to allow firmer inferences about the nature of long-term memory representations that underlie knowledge transfer. Also, from a pedagogic point of view, only manipulations that permit transfer after a delay are likely to be of practical value as instructional aids.

In many ways the paradigm and materials employed by Gick and Holyoak (1983) provide the minimal possible intervention that might possibly facilitate long-term transfer by means of prior inductive generalization. The source analogs lack any surface resemblance to the target problem, only two examples are provided (the minimal number that permits comparison across examples), and the subjects are novices with respect to the class of convergence problems (cf. Novick, 1988). If this simple manipulation produced transfer after substantial delays, it would suggest that exposure to multiple source analogs, even in the absence of direct instruction, is sufficient to yield robust transfer. If not, the results would indicate that stronger manipulations will be necessary to achieve a positive outcome.

Spencer and Weisberg (1986) conducted an initial study that investigated transfer as a function of delay (target problem presented immediately after source analogs or 6 min later) and contextual similarity (both source analogs and target problem administered by the same experimenter, or source analogs presented by an experimenter and target problem presented by an instructor as a class demonstration). These investigators replicated Gick and Holyoak's findings that schema quality, as measured by an assessment of similarity descriptions, predicted transfer both with and without the aid of a hint when the target problem was given in the same context. In addition, schema quality had a positive effect on *hint-aided* transfer in the different-context, delayed-test condition (Spencer & Weisberg, 1986, Experiment 1). However, transfer in the absence of a hint was extremely low if the context was changed, and it did not vary as a function of schema quality.

The results of Spencer and Weisberg (1986) suggest that presentation of multiple source analogs is not sufficient to ensure transfer across different contexts. Their results are not conclusive, however, because most of their subjects produced poor-quality schemas, and some of their negative findings might be attributable to floor effects. The experiments reported here continue the investigation of analogical transfer after delays and contextual changes.

Experiment 1

Experiment 1 was designed to replicate and extend Gick and Holyoak's (1983) findings concerning the role of schema quality as a predictor of immediate transfer, in order to provide a basis for comparison with subsequent studies of transfer after delays. Whereas Gick and Holyoak (1983) always required subjects who read two analogs to compare the two explicitly—a manipulation intended to foster abstraction of a schema based on the common structure of the analogs—this experiment included groups who received two analogs without any such comparison instructions. If the comparison groups yield greater transfer than the no-comparison groups, it would support the conclusion that schema induction, rather than exposure to multiple examples per se, is a crucial determinant of transfer.

Method

Subjects. Seventy-seven University of Michigan students participated in the experiment either for course credit or for pay. The subjects were randomly divided approximately equally across the four conditions described below.

Procedure. The experiment was administered in one session lasting between 35 and 50 min. Subjects were tested in groups of anywhere from 10 to 30 (as was the case in all the experiments unless indicated otherwise). Subjects were divided into four groups, defined by the factorial combination of comparison versus no-comparison instructions and two analogs versus one analog plus one disanalogous story. A cover sheet informed subjects that the experimenters were selecting stories to use in a study of story comprehension. Subjects were given a total of 5 min to read both stories. They were told that when they were done, the next part of the experiment would be distributed to them. All subjects then wrote summaries of the two stories while the stories were still in front of them, and those in the

comparison groups also wrote descriptions of how the two stories were similar. In this task, subjects were simply instructed to "describe as clearly as possible the ways in which the situations in the two stories seem similar." This phase took approximately 20 min. Immediately afterward, the prior materials were removed, and all subjects read the radiation problem (a version we will refer to as the *dosage* version, previously used by Holyoak & Koh, 1987; see Appendix A) and listed as many possible solutions as they could. Finally, subjects were given a hint to consider solutions to the radiation problem suggested by one or both of the prior stories. This hint consisted of the following question: "What solution to the ray problem is suggested by the stories?"

Subjects who received two source analogs read "The General" and "The Fire Chief" (see Appendix II in Gick & Holyoak, 1983); those who received one analog and a control story read one of the above plus "The Wine Merchants" (an unrelated story; see Appendix IV in Gick & Holyoak, 1980). Within the latter condition, each of the two convergence analogs was used equally often across subjects, and the order of the analog and the control story was counterbalanced. Neither of the latter factors affected performance, and the results reported below are collapsed over these variables.

In summary, there were four groups based on the crossing of number of analogs (one or two) and comparison instructions (comparison or no comparison).

Results and Discussion

Subjects' answers to the radiation problem were scored for inclusion of the convergence solution. To be scored as a convergence solution, an answer had to include the following components: (a) rays applied to the tumor from several directions and (b) applied simultaneously. Answers were not required to explicitly mention weak rays because this notion could be implied in the idea of using multiple rays. All other answers were labeled as nonconvergence solutions. Solutions were also broken down according to whether they occurred prior to a hint or in total (including hint-aided solutions). Statistical tests were performed by using the maximum-likelihood chi-square (G^2) (see Bishop, Fienberg, & Holland, 1975).

Table 1 presents the distribution of frequencies of convergence solutions as a function of whether one or two analogs were provided and whether or not subjects were required to compare the similarities of the two stories. The comparison manipulation had no significant effect for subjects who read one analog plus the control story. For subjects who read two analogous stories, however, comparison instructions increased the percentage who gave convergence solutions prior to the hint (47% of those in the two-analog comparison group vs. only 16% in the no-comparison group, $G^2(1) = 4.54$, $p < .05$. This experiment did not include a control group that received no analog at all. However, the subject population was highly similar to that which in similar previous studies has reliably yielded a rate of not more than 10% convergence solutions in the absence of any prior source analog (Gick & Holyoak, 1980; Holyoak & Koh, 1987). With that figure as the baseline, only the two-analog comparison group produced convergence solutions prior to the hint significantly more frequently than would be expected in the absence of an analog, $G^2(1) = 17.3$, $p < .001$.

Table 1
Percentage of Subjects Producing Convergence Solution: Experiment 1

Group	Before hint	Total
Comparison		
Two analogs (n = 19)	47	68
One analog plus control (n = 19)	16	53
No comparison		
Two analogs (n = 19)	16	74
One analog plus control (n = 20)	25	30

Combining results over comparison conditions, subjects who received two analogs produced a greater frequency of convergence solutions in total (i.e., before and after the hint) than did subjects who received one analog plus the control story (71% vs. 41%), $G^2(1) = 7.16$, $p < .01$. This effect is not solely due to the performance of the two-analog compare group; in fact, the two-analog no-comparison subjects did slightly better than the two-analog compare subjects in total (74% vs. 68%). It is also worth noting that the one-analog no-comparison group did not improve much on its before-hint performance after receiving a hint (25% vs. 30%), whereas the comparison group did (16% vs. 53%). This was probably due to the fact that although the compare group had just one analog in memory, the analog received additional processing from the comparison task that may have enabled subjects to use it when solving the target problem. The no-comparison group did not have this benefit.

For subjects in the comparison groups, written descriptions of story similarities were scored for inclusion of elements of the convergence solution (converging forces, small in magnitude, and applied simultaneously) by using the criteria described by Gick and Holyoak (1983). Gick and Holyoak defined three levels of schema quality (good, intermediate, and poor); however, because good schemas were relatively infrequent in the experiments reported here, data will be reported collapsing good and intermediate schemas into a single category, which will be termed *appropriate*. A similarity description was scored as embodying an appropriate schema if at least one of the above components of the convergence solution was mentioned; otherwise, it was scored as poor.

Not surprisingly, subjects who compared one analog with an unrelated control story invariably wrote poor schemas; this result will not be considered further. For subjects who compared two analogs, those who wrote appropriate rather than poor schemas tended to produce more convergence solutions prior to the hint (58% vs. 29%), although given the small number of subjects, this difference was not significant, $G^2(1) = 1.61$, $p = .20$. Total convergence solutions, including those that were hint-aided, did not differ significantly as a function of schema quality (67% vs. 71%), $G^2(1) = 0.05$, $p = .83$.

In general, the results of Experiment 1 replicate and extend those of Gick and Holyoak (1983). Explicit instructions to compare two analogs increased transfer prior to a hint, and for subjects who compared two analogs, solution-appropriate schemas at least showed a trend toward being correlated with more successful transfer. Like the earlier findings of Gick and Holyoak, however, our results were obtained under conditions of minimal delay or contextual change.

Experiment 2

Experiment 2 was designed to assess the robustness of transfer from multiple analogs when a delay of varying length and a change of context intervened between presentation of the source analogs and the target. Spencer and Weisberg (1986), using similar materials, found that a delay or context change essentially eliminated transfer in the absence of a hint.

Method

Subjects. Subjects were 97 introductory psychology students at the University of Michigan who participated in the study to satisfy a course requirement. Due to scheduling difficulties, unequal numbers of subjects were randomly assigned to the four conditions.

Procedure. All subjects were told that the experimenter was collecting data for several different experiments (as was in fact the case) and that they would be receiving several booklets with different tasks for them to do. Subjects worked at their own pace. They were divided into four groups, defined by the factorial combination of comparison versus no-comparison instructions and 1-week delay versus 30-min delay.

Subjects first received a booklet containing two story analogs. The set of four convergence analogs provided in Gick and Holyoak (1983, Appendix II) was used. Each subject received either two military stories, two fire-fighting stories, or one of each. Because differences in story combinations had no influence on transfer, all reported results are collapsed over this variable.

After reading and summarizing the stories, half the subjects (the comparison group) wrote descriptions of the similarities between the two stories. The other half (the no-comparison group) summarized the stories a second time; however, they could not look back at the stories while writing the second set of summaries. This additional summarization task was required of the no-comparison subjects in an attempt to make sheer amount of processing more equal between the two groups. We wished to control for the possibility that comparison instructions might improve transfer performance not by fostering schema formation but simply by virtue of forcing greater processing of the individual stories. After doing these tasks, subjects spent about 30 min filling out a questionnaire for a social psychology experiment. This task served as a change of context. Half the subjects, those in the 30-min delay condition, then were given the radiation problem to solve, at first without a hint to use the stories and then with such a hint. The other subjects, in the 1-week delay condition, ended their first session after filling out the social psychology questionnaire. One week later these subjects returned individually to a different room, where they began by spending about 30 min reading stories (for the social psychology experiment) and making judgments on a computer terminal. They then were given the radiation problem to solve.

Results and Discussion

Table 2 presents the frequencies with which subjects in each of the four conditions produced convergence solutions. None of the conditions produced frequencies of convergence solutions prior to the hint that exceeded the baseline expectation of 10%. When hint-aided solutions are also considered, the solution frequency was higher for the 30-min delay condition than for the 1-week delay condition (65% vs. 42%), $G^2(1) = 4.64$, $p < .05$.

Unlike the no-delay condition tested in Experiment 1, the comparison manipulation in Experiment 2 had no significant effect at either the 30-min [before hint: $G^2(1) = 0.053$, $p = .82$; total: $G^2(1) = 1.29$, $p = .26$] or the 1-week delay [before hint: $G^2(1) = 0.002$, $p = .96$; total: $G^2(1) = 0.27$, $p = .61$]. The absence of a significant advantage for the comparison condition, which should theoretically have fostered schema formation, implies that the "greater effort" hypothesis suggested as an account of the advantage obtained in Experiment 1 cannot be ruled out. However, the lack of spontaneous transfer after a delay resulted in a floor effect, so that no conclusions about the effect of comparison instructions can be drawn from the results of Experiment 2, except that such processing is not a sufficient condition for spontaneous transfer.

Table 2
Percentage of Subjects Producing Convergence Solution: Experiment 2

Condition	Before hint	Total
30-min delay (n = 66)		
Comparison (n = 35)	11	71
No comparison (n = 31)	10	58
1-week delay (n = 31)		
Comparison (n = 15)	7	47
No comparison (n = 16)	6	38

An analysis of the impact of schema quality for the comparison conditions is presented in Table 3. Even subjects who wrote appropriate schemas failed to produce convergence solutions prior to the hint more frequently than the baseline expectation of 10%. However, subjects who wrote appropriate schemas produced a higher frequency of convergence solutions when hint-aided solutions are included (78% vs. 48%), $G^2(1) = 4.73$, $p < .05$. This difference is mainly due to the 1-week-delay condition [although the interaction is not significant, $G^2(1) = 0.003$, $p = .95$]; subjects who had written poor schemas had great difficulty making use of the source analogs after a week's delay even when directly asked to do so.

In summary, Experiment 2 strengthens the claim that schema quality predicts the likelihood of a subject's applying an analogy from the source analogs to the target when given a hint, particularly after longer delays. Without a hint, however, the schema did not seem to be sufficiently salient to yield spontaneous transfer. The results thus essentially replicate those of Spencer and Weisberg (1986).

Experiment 3

The results of the first two experiments indicate that without providing a hint to use the prior stories (either by an

Table 3
Percentage of Subjects in Comparison Conditions Producing Convergence Solution as a Function of Schema Quality: Experiment 2

Condition	Before hint	Total
30-min delay (n = 33)		
Appropriate schemas (n = 20)	11	80
Poor schemas (n = 13)	15	62
1-week delay (n = 15)		
Appropriate schemas (n = 7)	14	71
Poor schemas (n = 8)	0	25

explicit suggestion or by demand characteristics), little or no analogical transfer is obtained after a delay of even 30 min that is filled with an interpolated task. The interpolated task used in Experiment 2, an unrelated experiment, presumably encouraged subjects to treat each new task as unrelated to previous ones, eliminating any demand characteristics that might suggest using the prior stories to help solve the target problem. At the same time, the substantial transfer obtained once a hint was given indicates that the problem-relevant features (i.e., the features of the convergence solution) of the source analogs were still available in memory after a delay as long as a week, especially for subjects who had written appropriate schemas.

It is possible that even though subjects were storing some of the relevant features of the analogs, these features might not have been easily accessible because the target problem might have failed to activate them. Experiment 3 was performed to investigate the possibility that long-term transfer might be improved by varying the wording of the target problem so as to better cue one of the solution-relevant features of the prior stories.

Method

Subjects. Subjects were 65 introductory psychology students at the University of Michigan who participated in the study to satisfy a course requirement.

Procedure. The design and procedure were essentially identical to those used in Experiment 2. The interpolated task was a memory experiment rather than a questionnaire. The major change in Experiment 3 was that the "dosage" version of the target problem (see Appendix A) was reworded slightly. The seventh sentence was changed from "A sustained large dose of the rays will effectively destroy the tumor" to "If the rays reach the tumor all at once at a sufficiently high intensity, the tumor will be destroyed." In addition, the word *dosage* was changed to *intensity* in the eighth and ninth sentences. This "intensity" version of the radiation problem is closer to that used by Gick and Holyoak (1980, 1983) and seems more likely to activate the notion of simultaneous forces, a key structural feature of the source analogs.

Results and Discussion

Table 4 presents the frequencies with which subjects in each of the four conditions produced convergence solutions. In Experiment 3, unlike Experiment 2, there was no significant effect of delay on the production of convergence solutions when hint-aided solutions are considered. As in Experiment 2, none of the conditions produced frequencies of convergence solutions prior to the hint that significantly exceeded the baseline expectation of 10%, and the comparison manipulation had no significant effect.

Table 4
Percentage of Subjects Producing Convergence Solution: Experiment 3

Condition	Before hint	Total
30-min delay (n = 39)		
Comparison (n = 20)	20	65
No comparison (n = 19)	10	63
1-week delay (n = 26)		
Comparison (n = 13)	23	62
No comparison (n = 13)	15	46

An analysis of the impact of schema quality for the comparison conditions, presented in Table 5, provides some evidence that comparison instructions had an impact on spontaneous transfer in this experiment, at least for subjects who wrote appropriate schemas. Subjects who wrote appropriate schemas produced convergence solutions prior to the hint significantly more frequently than did subjects who wrote inappropriate schemas (40% vs. 6%), $G^2(1) = 6.19$, $p < .02$. Schema quality did not have a significant effect when hint-aided solutions are considered, $G^2(1) = 1.13$, $p = .29$; however, as in Experiment 2, the 1-week-delay subjects with inappropriate schemas tended to perform much more poorly than the other three groups (see Table 5). Nevertheless, the interaction between delay and schema quality was not significant, $G^2(1) = 1.95$, $p = .18$.

Experiment 4

The results of Experiment 3, when compared with those of Experiment 2, suggest that when the target problem is more analogous in its structure to the source analogs, then comparison instructions that yield appropriate schemas can result in some spontaneous transfer after a delay of a week.[1] It seemed possible that further improvement in transfer might be obtained if the comparison instructions provided subjects with more direction toward discovery of the underlying schematic problem structure shared by the source analogs. Many researchers have recognized the importance of helping students to learn new ways of thinking about problems or domains in order to be more successful at solving problems in those domains (e.g., Catrambone & Holyoak, 1987; Chi et al., 1981; Fong et al., 1986). Brown, Kane, and Echols (1986; see also Brown & Kane, 1988) found that children who were led to focus on the goal structure of several training problems transferred their knowledge to an analogous target problem more efficiently than did subjects who did not focus on the goal structure. Brown et al.'s results suggest that directing subjects to compare the source analogs on problem-oriented dimensions—similarity of goals, obstacles, and solution methods—while varying the superficial features will foster schematic representations that are able to support transfer across contextual changes (see also Gick & Holyoak, 1987; Keane, 1985).

One source of difficulty for subjects in Experiments 1–3 may have been that the comparison instructions introduced

[1] The dosage and intensity versions used in this study differ in two ways: (a) The latter includes the word *intensity* instead of *dosage*, and (b) the latter includes reference to the rays reaching the tissue "all at once." An unpublished experiment by Laura Novick, Barbara Spellman, and Keith Holyoak compared transfer with versions of the ray problem that varied these two differences separately. The results indicated that inclusion of the structural feature of simultaneous application of the rays, rather than the word *intensity*, is the basis for greater ease of transfer.

Table 5
Percentage of Subjects in Comparison Conditions Producing Convergence Solution as a Function of Schema Quality: Experiment 3

Condition	Before hint	Total
30-min delay (n = 20)		
Appropriate schemas (n = 8)	38	62
Poor schemas (n = 12)	8	67
1-week delay (n = 13)		
Appropriate schemas (n = 7)	43	86
Poor schemas	0	33

by Gick and Holyoak (1983) and adopted in the first three studies here provide no guidance as to the type of similarities between the story analogs that ought to be emphasized. Although a sizable proportion of subjects in Experiments 1–3 wrote "appropriate" schemas, their written comparisons also included a large number of features that were irrelevant to the convergence solution used in the story analogs. Subjects tended to focus on general problem features and various story themes, such as the presence of a hero who saved the day, which is quite reasonable, given that they were reading stories and thus had a preference to process them by using standard story-understanding strategies (van Dijk & Kintsch, 1983).

Experiment 4 was designed to test whether more directed comparison instructions could induce subjects to focus on the convergence features of the story analogs and whether such focusing would help them apply the analogy to the target problem. Two sets of comparison instructions were used, one more directive than the other. In addition, both the dosage and the intensity versions of the ray problem were used.

Method

Subjects. Subjects were 90 introductory psychology students participating in the experiment either for cash or to satisfy a course requirement. Subjects were randomly divided approximately equally across conditions and worked at their own pace throughout the experiment.

Procedure. The general framework was the same as in the prior experiments. The design involved four basic groups, defined by a factorial combination of two levels of instructions and two levels of target problem wording. All subjects were told they were participating in several experiments. They read two analogs ("The General" and "The Fire Chief") and wrote summaries of them. Then they wrote comparisons of the analogs. Two sets of comparison instructions were used (see Appendix B). Groups 1 and 2 received comparison questions that isolated the issues of goal similarity, obstacle similarity, and method similarity in the two stories. Groups 3 and 4 received yet more detailed comparison questions that further isolated certain features of the problem states and methods. After answering the questions, subjects were provided with "ideal" answers and then received the comparison questions again. They were then asked to reproduce the ideal answers from memory. The ideal answers to the comparison questions for all groups contained the three convergence elements. After writing the comparisons, subjects in Groups 3 and 4 solved a problem ("The Aquarium") that was analogous to the "The General" and "The Fire Chief," after being explicitly told to use the prior analogs to help generate an answer. This introduction of an additional example and direct problem-solving experience is similar to a manipulation that Brown et al. (1986) found to improve transfer and was intended to further assist subjects in thinking of the analogs in terms of the relevant convergence features rather than simply as stories. After attempting to solve the aquarium problem on their own, subjects in Groups 3 and 4 read a convergence solution to it (see Appendix C). The majority of these subjects (66%) derived the convergence solution to the aquarium problem themselves.

Finally, subjects solved the radiation problem either immediately or after working on an interpolated task (an experiment in decision making) for approximately 30 min. This delay manipulation had no effect; the instructions appeared to eliminate demand characteristics of the situation even for subjects who did not have a delay imposed. That is, the no-delay subjects did not appear to have any advantage over the delay subjects. Accordingly, all reported results will be based on data combined across level of delay. Subjects first solved the radiation problem without a hint to consider the prior analogs and were then given a hint to consider them. Subjects in Groups 1 and 3 received the dosage version of the ray problem, whereas those in Groups 2 and 4 received the intensity version. A control group (n = 16) that did not read the story analogs also solved the radiation problem. Half of the control subjects received the dosage version and half received the intensity version.

To summarize, subjects in Groups 1 and 2 received less directive comparison questions than did subjects in Groups 3 and 4. Subjects in Groups 1 and 3 received the dosage version of the target problem, whereas subjects in Groups 2 and 4 received the intensity version.

Results and Discussion

Table 6 presents the frequencies with which subjects in the five groups produced the convergence solution both before the hint and in total. There were no significant differences in proportion of convergence solutions produced before the hint among Groups 1, 2, and 3 and the control group, $G^2(3) = 4.30$, $p = .23$. Because no control subjects produced the convergence solution, the control solution frequencies for the dosage and intensity versions obviously did not differ.

Combining results across the version of the target problem that noncontrol subjects received, subjects who received the more directive set of comparison instructions plus the aquarium problem produced the convergence solution before the hint more often than did subjects who received the less directive set of comparison instructions (36% vs. 10%), $G^2(1) = 7.10$, $p < .01$. The former subjects maintained an advantage in total solutions (70% vs. 42%), $G^2(1) = 5.52$, $p < .02$. Combining results across the version of the comparison instructions that noncontrol subjects received, subjects receiving the intensity version of the ray problem tended to produce the convergence solution before the hint more often than did subjects with the dosage version (32% vs. 14%), $G^2(1) = 3.36$, $p < .07$. This difference disappeared when hint-aided solutions are also considered (58% vs. 51%), $G^2(1) = 0.34$, $p = .56$.

Table 6
Percentage of Subjects Producing Convergence Solution for Ray Problem: Experiment 4

Instruction	Before hint	After hint
Less-directive comparison		
Group 1 (dosage version; n = 19)	10	47
Group 2 (intensity version; n = 19)	10	37
More-directive comparison		
Group 3 (dosage version; n = 17)	18	56[a]
Group 4 (intensity version; n = 19)	53	82[b]
Control group (n = 16)	0	—

[a] Does not include 1 subject who did not attempt the radiation problem after the hint.
[b] Does not include 2 subjects who did not attempt the radiation problem after the hint.

Finally, there was a trend toward a significant before-hint interaction between type of comparison instructions and version of target problem, $G^2(1) = 3.35$, $p = .07$ [the after-hint interaction is not significant, $G^2(1) = 2.38$, $p = .13$]. The before-hint interaction suggests that the combination of the more directive comparison questions, coupled with use of the additional aquarium problem and the intensity version of the ray problem, adds up to a fairly powerful manipulation for facilitating spontaneous transfer, even after a 30-min delay.

Subjects' answers to the comparison questions (prior to being given the ideal answers) were examined for the presence of the three convergence features (small forces, surrounding/converging on the target, and applied simultaneously). A slightly different scoring system was used in this experiment compared with the prior ones because for Groups 3 and 4 two of the comparison questions directly asked why it was important that the soldiers and the water arrived at their destination at the same time. Consequently, subjects' responses to the comparison questions were scored as appropriate if they mentioned the notion of small forces or surrounding a target and scored as poor for any other answer. Subjects in Groups 3 and 4 (86%) wrote appropriate schemas far more often than did those in Groups 1 and 2 (26%), $G^2(1) = 28.91$, $p < .0001$, suggesting that the more specific nature of the comparison questions that Groups 3 and 4 answered enabled them to focus better on the convergence features. Combining results across groups, subjects who wrote appropriate schemas produced the convergence solution prior to the hint more often than did those who wrote poor schemas (32% vs. 12%), $G^2(1) = 4.17$, $p < .05$. Those subjects with appropriate schemas also produced the convergence solution in total more often (68% vs. 39%), $G^2(1) = 6.09$, $p < .02$. More detailed analyses of the effects of schema quality within each of the four groups were not attempted because of small cell sizes (e.g., only two poor schemas were written by subjects in Group 3).

In summary, the results of Experiment 4 are broadly consistent with the findings that Brown et al. (1986) obtained by using simpler problems and much younger subjects. More extensive comparison instructions help subjects to focus on the relevant features of the training examples and to apply these features to a new problem, especially when the problem is worded to cue some relevant features of the examples.

Experiment 5

The results of Experiment 4 indicate that more-directive comparison instructions, coupled with a version of the target problem that provided a better cue for an important structural feature, yielded substantial transfer after a delay of 30 min. Experiment 5 was performed to determine whether such conditions would suffice to produce spontaneous transfer after a much longer delay, 1 week. In addition, Experiment 5 was intended to partially determine which aspects of the more directive condition employed in Experiment 4 were most crucial in facilitating transfer. The more-directive comparison condition differed from the less-directive condition in two potentially important ways: (a) the use of more detailed comparison questions and (b) the addition of a third example that provided direct problem-solving experience during training.[2] Experiment 5 was designed to separate the effect of *studying* versus *solving* a third example.

In Experiment 5, all subjects received the intensity version of the ray problem, and all received the more directive version of the comparison instructions from Experiment 4. One group of subjects received two story analogs plus the aquarium problem, whereas the other group received three story analogs, with the aquarium problem rewritten as a story with a solution. Finally, all subjects received the ray problem after a week's delay.

Method

Subjects. Subjects were 23 students in introductory psychology who participated in the experiment for course credit.

Procedure. There were two groups in this experiment. In Session 1, Group 1 ($n = 11$) received three stories to read, summarize, and compare. The stories were "The General," "The Fire Chief," and "The Aquarium" (which was changed into a story with a solution provided). The comparison instructions were adapted from those used by Groups 3 and 4 in the prior experiment to include "The Aquarium." Group 2 ($n = 12$) read, summarized, and compared "The General" and "The Fire Chief" and then received "The Aquarium" (in its problem version) to solve with a remainder to use the prior stories. The convergence solution to the aquarium problem was generated by 91% of the subjects in Group 2; all subjects in this group then read a convergence solution to it.

In Session 2, which occurred a week later, subjects spent 15 min working on a separate experiment in decision making and then received the target problem. As before, subjects were told they were participating in several experiments.

Results and Discussion

Transfer was extremely robust for both groups, and they did not differ significantly in the frequency with which they produced the convergence solution before the hint (Group 1 = 64%, Group 2 = 83%), $G^2(1) = 1.17$, $p = .28$, or in total (91% vs. 92%), $G^2(1) = 0.004$, $p = .95$. There was a trend for greater spontaneous transfer in Group 2, which had solved the aquarium problem during training, than in Group 1, which had received it as a story. Subjects who produced appropriate schemas also produced the convergence solution prior to the hint more often than did subjects who produced poor schemas (81% vs. 0%), $G^2(1) = 5.95$, $p < .02$. However, this comparison is hampered by the fact that only 2 subjects produced poor schemas. After the hint, all subjects except 2 (who wrote appropriate schemas) produced the convergence solution.

Perhaps the most interesting result from Experiment 5 is that subjects produced the convergence solution far more often before the hint than did 1-week-delay subjects in prior experiments. Overall, 74% of subjects in Experiment 5 pro-

[2] Another possibility is that the aquarium story is somehow a more effective analog than the convergence stories taken from Gick and Holyoak (1983). However, an unpublished experiment by Laura Novick and Keith Holyoak, in which the aquarium story was used as a single source analog, revealed that spontaneous transfer to the radiation problem was as infrequent for the aquarium story as for a story from Gick and Holyoak (1983).

duced the convergence solution before the hint, which compares very favorably with the 1-week-delay subjects in Experiments 2 and 3 (7% and 23%, respectively). It seems clear that extensive comparison instructions based on three source analogs are sufficient to allow most subjects to learn the important features of the class of convergence problems and to recognize the applicability of the solution method to a new example presented after a substantial delay and context change. It does not appear to be critical whether the third analog is solved or simply read (in story form). This last result agrees with other evidence that learners can study solved examples (as opposed to working them out) and still perform quite well on transfer problems (Catrambone & Holyoak, 1988; Chi, Bassok, Lewis, Reimann, & Glaser, 1989).

General Discussion

The results of these experiments indicate that presentation of fairly extensive comparison questions along with three analogs seems sufficient to enable reliable transfer to a superficially dissimilar target analog in the absence of an externally provided hint to apply the prior source analogs. When the target problem is presented immediately after only two story analogs and very general comparison instructions in the context of a single experiment, then transfer prior to a hint is also obtained with significant frequency (Gick & Holyoak, 1983; Spencer & Weisberg, 1986; Experiment 1 of our study). Such transfer is usually sharply reduced or eliminated when the source analogs and the target are presented in different contexts, even when the transfer test is immediate (Spencer & Weisberg, 1986; our Experiments 2–3). In fact, performance does not get much worse even after a delay of a week. This suggests that the main obstacle to transfer in these situations is a shift in context, not the passage of time (although it could be argued that it is impossible to have a passage of time without a shift in context).

The results of Experiments 4 and 5, however, indicate that if the comparison instructions are extensive and the target problem is phrased in a way that cues an important feature from the source analogs, then transfer is greatly improved, with the majority of subjects demonstrating spontaneous transfer after a delay as long as a week. Our results do not allow us to distinguish whether more directed questions, a third example, or some combination of these factors is necessary for long-term transfer. However, the results of Experiment 5 indicate that direct problem-solving experience, although it may be helpful, is not crucial. It seems most likely that all of these factors—more examples, more-directive comparison instructions, and problem-solving experience—can contribute to the acquisition of generalized schematic knowledge of a problem category, which in turn allows more flexible transfer (Brown & Kane, 1988; Brown et al., 1986). We have demonstrated that this transfer can occur even after contextual changes and delays.

Our results suggest that comparison exercises must be carefully constructed in order to encourage learners to focus on the aspects of training examples that a teacher deems most important. Neither undirected similarity comparisons (Experiments 2–3) nor highly abstract questions about structural similarities (Experiment 4) proved effective in allowing long-term transfer. In general, novices in a domain tend to focus on those features of examples with which they are most familiar and to miss the underlying concepts that the examples are supposed to demonstrate (Catrambone & Holyoak, 1987; Chi et al., 1981). In addition, the results of Experiments 3–5 indicate that the phrasing of the transfer problem is important, suggesting that learners may need to learn how to encode novel problems in terms of features that are shared by prior examples. Two factors thus seem to be implicated in mediating transfer. One involves the quality of the representation of the commonalities among multiple source analogs; the other involves the presence of cues in the target problem that activate relevant features of the source analogs. In both cases the effect is to increase the likelihood that relevant features from prior training examples will be recalled and applied to the target problem.

These results extend a broad picture of the conditions that govern analogical transfer that has emerged in recent years. When the source and target analog share many salient surface properties, spontaneous transfer is quite likely to occur even in the absence of a hint—even, in fact, when the analogy is misleading (Carroll, Mack, Lewis, Grischkowsky, & Robertson, 1985; Ross, 1984). Transfer in the absence of a hint is much more difficult to obtain between isolated analogs drawn from different domains (Gick & Holyoak, 1980) unless the situations have some salient surface similarity (Holyoak & Koh, 1987) or unless critical features are emphasized in the comparison questions and target problems (Brown et al., 1986). Interdomain transfer is facilitated when multiple source analogs are first compared, highlighting the underlying relational structure (Gick & Holyoak, 1983). Multiple examples often allow transfer without a hint when the context is relatively constant, whereas a single source example typically would not suffice. However, processing multiple source analogs is not sufficient to ensure transfer under more demanding conditions in which the context is changed, as in the present experiments and those of Spencer and Weisberg (1986). Rather, students must be directed to focus on the relevant aspects of examples (Lewis & Anderson, 1985; Sweller et al., 1983). A fruitful avenue of research may involve searching for ways of helping learners to focus on relevant features of training examples in a variety of domains and to learn to reliably identify these features in transfer problems (e.g., Catrambone & Holyoak, 1988).

References

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 153–166.

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. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20*, 493–523.

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.

Carroll, J. M., Mack, R. L., Lewis, C. H., Grischkowsky, N. L., & Robertson, S. R. (1985). Exploring exploring a word processor.

Human-Computer Interaction, 1, 283–307.

Catrambone, R., & Holyoak, K. J. (1988). Learning subgoals and methods for solving problems. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 153–159). Hillsdale, NJ: Erlbaum.

Catrambone, R., & Holyoak, K. J. (1986, May). *Problem similarity and transfer.* Paper presented at the 58th meeting of the Midwestern Psychological Association, Chicago.

Catrambone, R., & Holyoak, K. J. (1987). Transfer in problem solving as a function of the procedural variety of training examples. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 36–49). Hillsdale, NJ: Erlbaum.

Catrambone, R., & Holyoak, K. J. (1988). Learning subgoals and methods for solving problems. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 153–159). Hillsdale, NJ: Erlbaum.

Chi, M. T., Bassok, M., Lewis, M. W., 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., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

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

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*, 253–292.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355.

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

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 9–46). Orlando, FL: Academic Press.

Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (Vol. 2, pp. 21–44). Hillsdale, NJ: Erlbaum.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*, 332–340.

Keane, M. (1985). On drawing analogies when solving problems: A theory and test of solution generation in an analogical problem-solving task. *British Journal of Psychology, 76*, 449–458.

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

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 510–520.

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

Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 124–139.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology, 6*, 436–450.

Ross, B. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371–416.

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*, 629–639.

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

Sweller, J., Mawer, R. F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112*, 639–661.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Appendix A

Dosage Version of Radiation Problem Used in Experiments 1–2 and Groups 1 and 3 in Experiment 4

Suppose you are a doctor faced with the following problem. A malignant tumor has developed in the stomach of one of your patients. If the tumor is not treated soon, the cancer will spread throughout the patient's body, resulting in death. Because of some medical complication it is impossible to perform an operation to remove the tumor or restrict its blood supply. There is therefore no simple way to treat the patient's condition. However, you have available a kind of ray that can be used to destroy the tumor. A sustained large dose of the rays will effectively destroy the tumor. Unfortunately, at this dosage the rays will also destroy the healthy tissue that they pass through on the way to the tumor. At a lesser dosage the rays would not harm the healthy tissue, but they would not destroy the tumor either.

What type of procedure might be used to destroy the tumor with the rays, and at the same time avoid destroying the healthy tissue? Suggest as many possible solutions as you can. Write down all the possibilities you can think of, even ones that may not really be practical. Don't worry about not having enough medical knowledge. Use any information you can think of to help solve the problem.

Appendix B

Comparison Questions and Ideal Answers

Groups 1 and 2 in Experiment 4

1. Describe how the main goal of the general and fire chief is similar.

The main goal of the general and the fire chief is similar in that in both cases they want to focus a large force (soldiers/water) on some target (the fortress/the fire) that cannot be used in full strength from a single direction.

2. How are the obstacles that made it difficult to capture the fortress and put out the fire similar?
   The obstacles are similar in the following ways:
   a. Only a small force can be aimed at the target from any given direction (only a small number of soldiers can go down a given path/only a single bucket of water can be thrown from any position).
   b. A small force by itself is not sufficient to do the job of a large force.
3. List all the important similarities you can think of in the *methods* used to capture the fortress and put out the fire.
   The methods are similar in the following ways:
   a. A large force (total number of soldiers/total amount of water available) was split up into many small forces.
   b. The small forces surrounded the target.
   c. The small forces simultaneously focused or converged on the target from all directions, thus adding up to a large force that focused on the target.

Groups 3 and 4 in Experiment 4 and All Subjects in Experiment 5

1. The fortress is difficult to capture because a large army of soldiers can not attack it from one direction.
   The fire is difficult to put out because a large amount of water cannot be thrown at it from one direction.
   Write a third sentence that is just like the two above, except that "The fortress" and "The fire" are replaced by a word or term that is more general, "army of soldiers" and "amount of water" are replaced by a more general term, and "attack" and "thrown at" are replaced by a more general term.
   A target is difficult to overcome because a large force cannot be aimed at it from one direction.
2. There were enough soldiers to attack the fortress, but they could not do it from one direction. What happened instead?
   The soldiers were broken into small groups, the groups surrounded the fortress, and then moved in on it all at the same time.
   There was enough water to put out the fire, but it could not be shot from one direction. What happened instead?
   The water was put into small buckets, the buckets surrounded the fire, and were then thrown onto it all at the same time.
   Based on the above answers, what is similar about the method used to capture the fortress and put out the fire?
   In both cases a large force was broken into many small forces. These forces were then aimed at a central target and made to arrive at the target all at the same time, thus adding up to a larger force again.
3. Why was it important that all the soldiers arrive at the fortress at the same time?
   By arriving at the same time the groups of soldiers added up to a very large and powerful group.
   Why was it important that all the water hit the fire at the same time?
   By hitting the fire at the same time the many buckets of water added up to a very large and powerful amount of water.

Appendix C

Aquarium Problem Given to Groups 3 and 4 in Experiment 4 and Group 2 in Experiment 5

Problem Statement

A major aquarium in a city on the East Coast decided to create a large aquarium display containing a replica of the sunken ocean liner the Titanic amid the sea environment of its resting place, which is deep in the Atlantic Ocean off the coast of Newfoundland. A professional aquarium designer was assigned to the project. She placed a small replica of the vessel in the center of a large tank, with a realistic sea bed. Then she added to the tank sea plants and fish of the sort that live in the Atlantic at the depth of the sunken Titanic. The display was virtually finished when the designer was confronted with a major problem she had failed to anticipate. In order to maintain the deep-water environment required by the fish and plants, the tank had to be kept quite dark, as the deep-water organisms were not adapted to light. However, if the tank was kept completely dark, people would not be able to see the small replica of the Titanic in the center of the tank, which, after all, was the main point of the exhibit. Putting lights inside the model of the wreck looked too artificial. The designer considered shining a powerful spotlight on the model of the vessel. However, if the spotlight was located inside the tank, it would raise the temperature of the water too high; and if it was located outside the tank, the bright beam seriously disrupted the feeding habits of some of the fish. So it looked like the display was going to have an embarrassing shortcoming.

What could be done to light the display? Write your solution below. Remember to use the same principles that were used in the stories you read.

Solution

Just before the display was to open, the designer hit upon a new idea. She had several low-powered spotlights placed at different locations around the outside of the tank, all focused on the replica of the ship. Each of the lights was quite dim, so the light-averse fish were not disturbed as they swam around the Titanic. But since all of the lights were focused on the ship, their beams added up so as to illuminate it enough that its realistic details could be seen by viewers. When the display opened, everyone thought it was both realistic and esthetically striking.

Received June 20, 1988
Revision received March 23, 1989
Accepted March 28, 1989 ■