

Ordering Worked Examples to Promote Categorization

Brian D. Gane (bgane@gatech.edu)

School of Psychology, Georgia Institute of Technology, 654 Cherry St.
Atlanta, GA 30332

Richard Catrambone (rc7@prism.gatech.edu)

School of Psychology, Georgia Institute of Technology, 654 Cherry St.
Atlanta, GA 30332

Abstract

We report an investigation of novices' ability to categorize probability word problems. Individuals can effectively learn from worked examples (Ward & Sweller, 1990), especially if multiple examples with the same structural features are presented (Gick & Holyoak, 1983). We were interested in whether the order of worked examples could affect a learner's ability to categorize new problems. Participants studied 5 types of probability word problems; each type had 2 worked examples. Two orders of worked examples--blocked and random--were used. For isomorphic test items, the random ordering reduced participants' total number of categorization errors and their categorization error rates. For transfer test items, ordering did not affect categorization errors. We argue these results demonstrate that the random ordering facilitated construction of problem schemas, which subsequently improved participant's categorization ability. However, the random ordering appeared to have no effect on learners' ability to restructure problem schemas as required by transfer items.

Keywords: Instructional design; order effects; problem solving; worked examples

Introduction

Successful problem solving requires accurately identifying the structure of the problem and constructing a solution strategy based on this structure. Consider one type of problem solving: word problems. To solve a word problem an individual must construct a mental representation of the problem and then search long-term memory to find a similar problem representation. If this recognition process is successful, the individual then has access to an existing problem representation, a schema, that can be used to guide his or her solution to the current problem (Chi, Glaser, & Rees, 1982). Problem-solving thus involves several cognitive processes: constructing a mental representation of the problem, recognizing a similar representation (i.e., a schema) in long-term memory, activating that problem schema, mapping a representation of the current problem to the existing schema, and then using the schema to solve the problem. In this paper we use the term *categorization* to refer to a subset of these processes: constructing a problem representation, recognizing a similar representation in memory, and activating that schema. How learners build problem schemas and complete the categorization process

can be affected by instructional design. This paper investigates one instructional design feature: how the ordering of examples affects construction of schemas and categorization of problems.

Instruction via examples

Worked examples, in which the solution steps for a sample problem are written out for the learner, are effective instructional aids. For novices in particular, worked examples can be more instructive than solving problems because they allow learners to devote mental resources to building and refining their problem schemas (Ward & Sweller, 1990). However, there is strong empirical evidence that to successfully apply a problem schema, more than one example with the same schema must be studied. One hypothesized reason that multiple examples of the same schema are necessary is that multiple examples are needed in order for learners to isolate the *structural features* of the examples (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983). Word problems invariably have both *surface features* and structural features. Surface features are aspects of the cover story that are incidental; they are irrelevant and not related to the problem schema in long-term memory. In contrast, structural features form the basis for the problem schema; structural features define and distinguish one particular problem schema from another. Consider the following probability word problem.

Jonie is putting her groceries on the conveyer belt in the supermarket. She has a potato, an onion, a tomato, a squash, and broccoli in her basket. Assuming that she randomly picks produce from the basket, what is the probability that she will put the tomato on the conveyer belt first and the squash on the belt second?

This problem has several surface features that are incidental: groceries (as a set, and the individual grocery items), picking items out of a container, and the values in the problem (5 items total, 2 items being selected). On the other hand, the problem has two crucial structural features: one must consider the order in which items are selected (i.e., permutation) and an item is not returned after it has been selected (i.e., without replacement).

Research on problem solving has consistently shown that novices have difficulty separating surface features from structural features (Catrambone, 1998; Chang,

Koedinger, & Lovett, 2003; Ross, 1989). One danger, therefore, is that novices might construct an erroneous schema by including surface features, rather than structural features. A related danger is that novices might construct a problem schema with incorrect structural features. Consider a domain, such as probability, that has multiple problem types, each with its own set of structural features (see Table 1). Novices learning about this domain would need to create multiple schemas, each with a unique set of structural features. Given evidence that novices should study multiple examples for a particular problem type (Gick & Holyoak, 1983), learners would need to study multiple examples for each of the different problem types. They then have to apply this knowledge by categorizing new problems on the basis of existing schemas that they created while studying (Quilici & Mayer, 1996). An important question for instructional designers arises: can one order examples in a way that maximizes a novice's ability to form multiple problem schemas and categorize novel problems according to these schemas?

Block structural features

One approach to ordering worked examples is to block multiple examples with the same structural features, so that learners study consecutive examples from the same problem category (see Table 1). This blocked ordering might encourage learners to focus on structural features. By changing the surface features but retaining structural features across a block of related examples, learners have to disregard the surface features; in the process they might isolate the structural features and incorporate these into their problem schemas (Quilici & Mayer, 1996).

Randomize structural features

A different approach to ordering worked examples comes from research on distribution of practice (Jacoby, 1978) and contextual interference (Shea & Morgan, 1979). Increasing the spacing between studying related items increases the probability of recalling that item (Jacoby, 1978). Thus, when studying worked examples, increasing the spacing between related items might force the learner to reconstruct the earlier studied example (i.e., build a representation of the current example and then recognize a related representation in long-term memory). The effort of reconstructing the representation could lead to better recall and more robust schemas. There are other theoretical explanations about why a distribution of structurally identical worked examples should be used. Some researchers (e.g., Shea & Morgan, 1979) argue that examples of different types should be interspersed because it allows learners to compare and contrast different features of each example. When the types of examples are not blocked, and appear random to the learner, then the learner does not know a priori to which category the example belongs. Thus the learner must compare previous items with the current item and compare and contrast until he or she isolates the features

that are structurally relevant (van Merriënboer, Schuurman, de Crook, & Paas, 2002).

This paper investigates whether a blocked or random order of worked examples facilitates a learner's ability to build an accurate problem schema and to subsequently categorize new problems. This research question is part of a larger study investigating learning within the domain of probability word problems (Gane & Catrambone, 2006). To investigate how the order of worked examples affects categorization, we created two sequences of worked examples: blocked and random. Some of the worked examples varied their structural features (i.e., five types of probability word problems); all the worked examples varied their surface features (i.e., 10 cover stories). If consecutive examples of the same problem allow learners to isolate surface features and build an accurate schema, a blocked schedule should be better at helping a learner to build an accurate schema. If, however, distributing examples across learning forces learners to reconstruct problem schemas or contrast different structural features, a random schedule should be better.

Method

Participants

Undergraduates that had not previously completed a college-level course on probability or statistics participated. Sixty-five participants completed the study and were compensated with course credit.

Materials

Learning materials. Part one of the probability learning materials was an introduction to the probability domain, which explained general concepts (e.g., the concept of a random experiment or selection with replacement). Part two of the learning materials contained a set of 10 worked examples of probability word problems. The worked examples had each step worked and explicated with text descriptions of the step and the rationale for the step.

There were two worked examples for each type of formula used to solve a problem; each worked example had a unique cover story. Five types of word problems were used (see Table 1). Note that the first type of problem is different from the other four; the first problem illustrates that when a conjunction of probability events is of interest, the probability of the individual events are multiplied together to compute the probability of the overall event. The other four problems illustrate that complex events can also be solved by applying a specific formula. In these problems the appropriate formula is determined by two features: (1) whether the order in which items are selected is of importance (i.e., permutation or combination) and (2) whether items are replaced after each trial (i.e., with replacement or without replacement). These last four types of problems use a strategy we label the *category approach*. In the category approach strategy, learners are taught to classify the

Table 1: Worked example problem types and associated formulas.

#	Type	Formula	
		A	p()
1	Multiply probability	--	$p() = 1 / a_1 * 1 / a_2$
2	Permutation without replacement	$A = n! / (n - k)!$	$p() = 1 / A$
3	Permutation with replacement	$A = n^k$	$p() = 1 / A$
4	Combination without replacement	$A = n! / [(n - k)! * k!]$	$p() = 1 / A$
5	Combination with replacement	$A = (n + k - 1)! / [(n - 1)! * k!]$	$p() = 1 / [A - n + (n * n/k)]$

Table 2: Order of worked examples for the blocked and random group.

Order	Blocked	Random
1	Multiply probability	Multiply probability
2	Multiply probability	Permutation without replacement
3	Permutation without replacement	Combination with replacement
4	Permutation without replacement	Multiply probability
5	Permutation with replacement	Permutation with replacement
6	Permutation with replacement	Combination without replacement
7	Combination without replacement	Combination with replacement
8	Combination without replacement	Permutation without replacement
9	Combination with replacement	Combination without replacement
10	Combination with replacement	Permutation with replacement

problem based on the two features and then select the appropriate formula for that category. Note that any complex probability that can be solved with the category approach can also be solved by multiplying individual simple probabilities (i.e., formula 1 in Table 1). Although this was not explicitly explained to participants, some recognized it and thus solved problems with an *individual-event approach* (Gerjets, Scheiter, & Catrambone, 2004), rather than the category approach that

Test items. After studying the learning materials, participants were given an 11-item test. Five items were isomorphs of the worked examples. These isomorphic items were each a particular category of probability word problem and therefore had one correct formula (see Table 3). Additionally, there were six transfer items (3 near transfer and 3 far transfer items). These transfer items each contained essentially two word problems. That is, there are two different complex probabilities that need to be computed individually and then combined using multiplication. Each of the complex probabilities have one correct formula associated with them; near transfer items have the same two categories (and thus the same two formulas) whereas far transfer items have two different categories (and thus two different formulas; see Table 3).

we expected them to use. The implications of this are discussed in the results section.

The order of the worked examples varied by condition. In the blocked group, both examples of a given problem type were grouped together (see Table 2). In the random group, the worked examples were ordered such that from the point of view of the participant, the type of problem appeared to be randomly ordered; no worked example was preceded or followed by a worked example of the same problem type (see Table 2).

Procedure

Participants first completed the learning phase, in which they worked through materials (introduction and worked examples) at their own pace. Participants were told that they would be given a formula sheet during the test phase and thus did not have to memorize the formulas but that they would need to learn how to apply the formulas. The test phase followed the learning phase. Participants were given an 11-item test to complete within 35 minutes. They were free to work through the items in any order, and were encouraged to skip items if they could not solve them, in order to attempt each item.

Table 3: Sample isomorphic, near transfer, and far transfer items.

Item type	Cover story	Solution	
Isomorph	Ten knights participate in the "9th King's Tournament". The king provides the tournament with 12 horses. The knights have to pick their horses blindfolded. The heaviest knight gets to pick first, then the second heaviest, and so on. <i>What is the probability of the heaviest knight getting the biggest horse, the second heaviest knight getting the second biggest horse, and the third heaviest knight getting the third biggest horse?</i>	$A = n! / (n - k)!$ $A = 12! / (12 - 3)!$ $p() = 1 / A$ $p() = 1 / 330$	
Near transfer	At a soccer game there are two dressing rooms for the two teams. The 11 players from Oxford wear T-shirts with odd numbers from 1 to 21 and the 11 players from Manchester have even numbers from 2 to 22. Because the aisle from the dressing rooms is very narrow only one player at a time can enter the field. The players of the two teams leave their rooms alternately with a player from Oxford going first. <i>What is the probability of first 5 players entering the field having the numbers five, two, thirteen, eight, and one (i.e., the first has the number five, the second has the number two, and so on)?</i>	$A_1 = n! / (n - k)!$ $A_1 = 11! / (11 - 3)!$ $A_1 = 990$ $p() = 1 / A_1 * 1 / A_2$ $p() = 1 / 108900$	$A_2 = n! / (n - k)!$ $A_2 = 11! / (11 - 2)!$ $A_2 = 110$
Far transfer	In a car race 12 different European countries participate with one driver per country. There are 5 prizes for the participants: The winner receives \$10,000, the second place finisher gets \$5,000 and the third place finisher receives \$1,000. The fourth and fifth place finisher each receives \$500. <i>What is the probability that the Italian driver wins \$10,000, the German \$5,000, the Swedish \$1,000, and that the French and Danish drivers each win \$500?</i>	$A_1 = n! / (n - k)!$ $A_1 = 12! / (12 - 3)!$ $A_1 = 1320$ $p() = 1 / A_1 * 1 / A_2$ $p() = 1 / 47520$	$A_2 = n! / [(n - k)! * k!]$ $A_2 = 9! / [(9 - 2)! * 2!]$ $A_2 = 36$

Results

We conducted a preliminary analysis using data from 40 of the 65 participants. We were interested in assessing participants' ability to categorize the test items. We used one specific problem-solving behavior, selecting the correct formula, as an indicator of correct categorization (Quilici & Mayer, 1996).

Scoring

We first scored each item for the strategy used to solve it (e.g., a category approach or an individual-event approach). This was necessary because participants sometimes switched between a category approach and an individual-event approach from item to item. In addition, some participants used both the category and individual-event approach to solve the same item; when participants

used both approaches we excluded their answer because it was unclear which strategy was their primary problem-solving method. Each attempt that was solved with the category approach was then scored to determine whether the participant selected the correct formula (e.g., applied the formula $n!/(n-k)!$ to a permutation without replacement item).

We computed a second measure, total categorization errors, that were defined as an instance in which the participant used the category approach, but selected an incorrect (or unclassifiable) formula for that item. Thus, for each participant, categorization errors equal the number of items solved with a category approach minus the number of correct categorizations (i.e., errors = category approach - correct categorization).

We also used a third measure, *error rate*, which was the proportion of errors to categorization opportunities (i.e., rate = errors / category approach). To illustrate, consider a hypothetical participant, Participant 1. Participant 1 used the category approach on three out of the five isomorphic items, and correctly categorized two of the isomorphic items. Therefore, his error score is 1 (3-2) and his error rate is .33 (1/3).

Analysis strategy

To analyze categorization errors and error rates we excluded participants that did not use the category approach for any of the items. Again, consider hypothetical participants. If Participant 2 did not use the category approach for any item then he would have 0 errors and an error rate of 0. However, if Participant 3 used the category approach for every item and categorized every item correctly, then she would also have 0 errors and an error rate of 0. Therefore, we excluded any participant who performed like Participant 2; this makes values of 0 easily interpretable.

Although three measures were computed (category approach attempts, categorization errors, and categorization error rates), we believe categorization error rates are the most theoretically relevant. Error rates allow one to compare participants' categorization data, despite differences in the number of categorization attempts made. We report the number of categorization attempts to provide context for the error rates. However, we report inferential analyses only on categorization errors and categorization error rates. These two measures were analyzed using a one-way ANOVA with order as a between-subject factor.

Isomorphic and transfer items were analyzed separately. Isomorphic items have one formula whereas transfer items have two formulas that are multiplied together to get the final answer. Therefore, each transfer item has two parts, and each part has a specific formula that must be selected. Participants could have between zero and five categorization errors across the five isomorphic items (one error per item) and between zero and 12 categorization errors across the six transfer items (two errors per item).

On isomorphic items the blocked group made significantly more categorization errors (see Table 4), $F(1, 27) = 6.73, MSE = 1.4, p = .02$. The blocked group also had a higher categorization error rate than the random group (see Table 4), $F(1, 27) = 4.76, MSE = 0.10, p = .04$ ¹. On transfer items the blocked and random groups did not differ in the number of categorization errors nor in their categorization error rates (see Table 4), both F 's < 1.

Discussion

The random condition had fewer errors than the blocked condition on items that were isomorphic to the worked examples. The random order was more successful than the blocked order in facilitating schema construction and subsequent categorization. This suggests that participants in the random condition were more successful at creating schemas that could be retrieved from long-term memory and applied to the problem representation, despite the fact that the test items had different surface features than the studied examples.

However, the order of the worked examples did not appear to affect errors on transfer items; no reliable order effect was found. The transfer items required participants to create a problem representation that was different from any stored schema. That is, participants had to combine two complex probability formulas (e.g., formulas 2 - 4 in Table 1) by multiplying the probability events together (e.g., formula 1 in Table 1). These transfer items therefore had different surface *and* structural features compared to the studied worked examples. Based on these results, it does not appear that alternating the structural features of worked examples (i.e., random ordering) is sufficient to allow learners to create schemas that are robust enough to support reorganization to deal with novel problem representations.

Table 4: Means (and standard deviations) of each dependent measure for isomorphic and transfer items.

Measure	Order			
	Blocked		Random	
	Isomorph			
Category approach	3.4	(1.0)	2.9	(1.4)
Errors	2.4	(1.2)	1.3	(1.2)
Error rate	.72	(.26)	.46	(.39)
	Transfer			
Category approach	7.2	(3.6)	8.4	(3.2)
Errors	4.9	(2.6)	5.2	(2.7)
Error rate	.74	(.23)	.66	(.26)

¹ When study time is controlled for, the difference in errors remain significant, but the difference in error rate does not.

Although these results suggest that structurally similar examples should be separated, we believe that our findings are not necessarily inconsistent with Quilici and Mayer (1996). Quilici and Mayer (Experiment 2) had two experimental groups that each had structurally similar problem types blocked together. However, one of these groups, surface-emphasizing, had surface features intentionally confounded with structural features; when the problem type changed, the cover story changed as well. The other group, structure-emphasizing, did not have surface features confounded with structural features; surface features changed independent of structural features. The structure-emphasizing group was better than the surface-emphasizing group at categorizing new problems. Therefore, they argued that when surface features are grouped, structural features should not be confounded. In this study we used a new cover story for each example in order to avoid confounding structural and surface features. Therefore, our design allows us to answer a question that Quilici and Mayer could not²: when structural and surface features are not confounded, should examples with the same structure be studied together (i.e., blocked condition), or with intervening examples of different problem types (i.e., random condition)? Our results suggest that randomizing the order of worked examples, such that consecutive worked examples do not have the same structural features, is successful in reducing categorization errors.

The benefit of a random, as opposed to blocked, ordering might occur because learners compare and contrast the structural features of the previous example with the current example, and thus are better able to build a schema with the appropriate structural features (van Merriënboer et al., 2002). On the other hand, the random order might be beneficial because the spacing between examples of the same problem type is increased. As more items intervene between study trials, learners increasingly have to reconstruct their problem schema from long-term memory; this reconstruction process might benefit retention (Jacoby, 1978). Although this study is unable to distinguish between these two possibilities, future research could attempt disentangle these explanations.

In conclusion, these results suggest that instructional designers should carefully consider the order in which they present examples to learners, especially when designing lessons with more than one problem type. Although grouping structurally similar examples might seem intuitive, spacing the presentation of similar examples can improve learners' categorization ability. Further research needs to specify which cognitive processes are facilitated by randomizing the order of examples. Identifying these cognitive processes can provide predictions about specific changes to the ordering of sequences of examples and the effect that these ordered

sequences will have on schema formation and categorization.

References

- Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, *127*, 355-376.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1147-1156.
- Chang, N., Koedinger, K. R., & Lovett, M. C. (2003). Learning spurious correlations instead of deeper relations. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 228-233). Boston, MA: Cognitive Science Society.
- Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.
- Gane, B. D., & Catrambone, R. (2006, April). *Can modular examples and contextual interference improve transfer?* Paper presented at the American Educational Research Association annual meeting, San Francisco, CA.
- Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science*, *32*, 33-58.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 39-65.
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649-667.
- Quilici, J. L. & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*, 144-161.
- Ross, B. H. (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*, 456-468.
- Shea, J. B. & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 179-187.
- van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning & Instruction*, *12*, 11-37.
- Ward, M. & Sweller, J. (1990). Structuring effective worked examples. *Cognition & Instruction*, *7*, 1-3.

²Quilici and Mayer (Experiment 3) attempted to answer this question, but they used a different methodology than our study. Their manipulation of the order of structural features did not cause significant differences in learners' categorization ability.