

## On the spontaneous discovery of a mathematical relation during problem solving

James A. Dixon<sup>a,\*</sup>, Ashley S. Bangert<sup>b</sup>

<sup>a</sup>*Department of Psychology, 406 Babbidge Road, Unit 1020, University of Connecticut, Storrs, CT 06269-1020, USA*

<sup>b</sup>*Department of Psychology, 525 East University, Ann Arbor, MI 48109-1109, USA*

Received 21 July 2003; received in revised form 22 October 2003; accepted 1 December 2003

---

### Abstract

People spontaneously discover new representations during problem solving. Discovery of a mathematical representation is of special interest, because it shows that the underlying structure of the problem has been extracted. In the current study, participants solved gear-system problems as part of a game. Although none of the participants initially used a mathematical representation, many discovered a parity-based, mathematical strategy during problem solving. Two accounts of the spontaneous discovery of mathematical strategies were tested. According to the automatic schema abstraction hypothesis, experience with multiple, unique problem exemplars facilitates extraction of the parity relation. According to the comparison-based abstraction hypothesis, explicitly comparing gear pathways that have different number, but the same parity, should result in extraction of parity. An event history analysis showed that accumulation of experiences with different-number, same-parity comparisons predicted discovery of parity; accumulation of unique exemplars did not. Results suggest that comparison-based abstraction processes can lead to the discovery of a mathematical relation.

© 2004 Cognitive Science Society, Inc. All rights reserved.

*Keywords:* Schema abstraction; Discovery; Mathematical relations; Comparison; Structure mapping; Problem solving

---

The ability to generate new representations for solving problems is a powerful and very useful property of human cognition. People can spontaneously discover new ways of representing problems during problem solving; some of these discoveries are quite dramatic, such as having a sudden “insight” (Kaplan & Simon, 1990), others are more subtle, such as

---

\* Corresponding author.

*E-mail address:* james.dixon@uconn.edu (J.A. Dixon).

modifying a strategy (Shrager & Siegler, 1998). Perhaps the most important type of representational change is a shift from representing the content and surface features of a problem to representing its underlying relational structure. The shift to a relational representation is of particular interest because it indicates that the problem has been reduced to its essence; the crucial relationships in the problem have been extracted and irrelevant details discarded. A relational representation may also provide the basis for a new problem category or schema, thereby allowing other problems with the same relational structure to be quickly and easily understood.

Research on analogical problem solving has identified two general ways in which relational structure may be extracted during problem solving, automatic schema abstraction and comparison-based abstraction (Reeves & Weisberg, 1994).<sup>1</sup> Automatic schema abstraction occurs as a result of experience with multiple exemplars. For example, Reeves and Weisberg (1994) discussed an experiment in which participants were asked to encode either two or four problem exemplars and rate the likelihood of those exemplars occurring in real life. Participants who rated four exemplars showed better problem solving transfer to a structurally analogous problem than participants who had been presented with two exemplars (see also, Chen, 1999; Ross, Nelson, Wetstone, & Tanouye, 1986). Automatic schema abstraction is thought to occur as a consequence of memory processes, particularly retrieval. When a new problem is presented, the degree to which its features match the features of stored exemplars determines the activation of those exemplars. The representation that results from the activation of the stored exemplars forms an abstract schema (Goldinger, 1998; Hintzman, 1986). See Gomez (2002) for a non-retrieval-based account.

Comparison-based abstraction occurs when two analogues are explicitly compared. This can occur independently of an attempt to solve a problem (e.g., participants are simply asked to compare two problems) or as part of a problem solving attempt (e.g., participants reference a specific base problem while trying to solve a target problem). The process of aligning the structures of the two problems highlights their common relations such that the relational structure can be stored separately from their original problems (see Kotovsky & Gentner, 1996; Namy & Gentner, 2002; Ross, 1984, 1989; Ross & Kennedy, 1990).

The vast majority of research on abstraction has been in the context of analogical problem solving and transfer (e.g., Bernardo, 2001; Gick & Holyoak, 1983; Medin & Ross, 1989; Reeves & Weisberg, 1994; Ross & Kennedy, 1990; for a recent review see Goldstone & Sakamoto, 2003), and to a lesser extent with categorization (e.g., Gentner & Namy, 1999; Ross & Warren, 2002). However, automatic and comparison-based abstraction are quite general processes and should be available whenever multiple exemplars are presented or problems compared. In the current study, we examined the contributions of these processes to the spontaneous discovery of a mathematical representation during problem solving. Mathematical relations are, of course, abstract by nature and therefore the discovery of a mathematical representation during problem solving is strong evidence that the relational structure of the problem has been extracted. The current study complements previous research by examining the ability of these models to predict the timing of the discovery of an abstract relation as a function of theoretically derived predictors, predictors which also vary across time.

We asked participants to solve a series of gear-system problems in the context of a train race. Each gear system had a driving gear that provided the force, a variable number of inter-

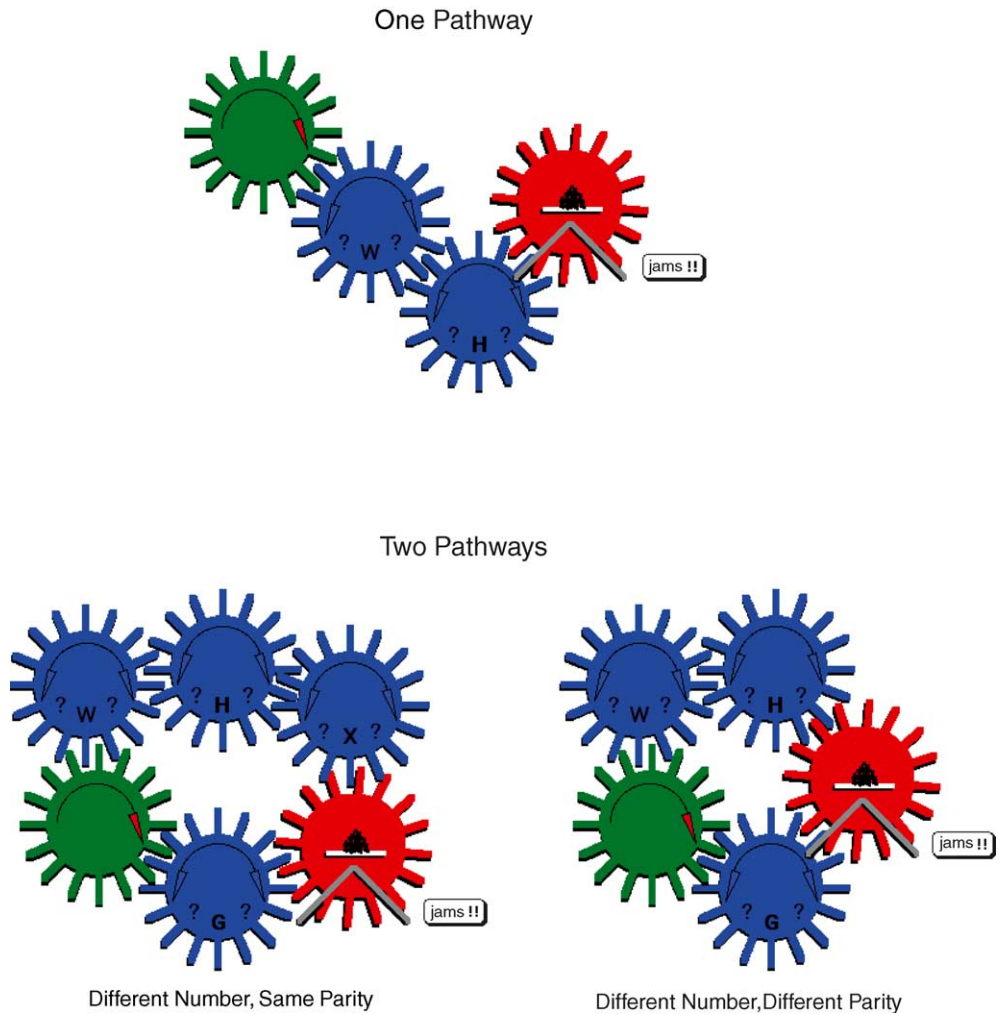


Fig. 1. Examples of three different gear-system problems are shown. Each gear system has a driving gear, marked by a single clockwise-pointing arrow, that provided the force to the system and a target gear that held a small pile of coal on a shelf. The driving and target gears are connected by one (top panel) or two (lower panels) sets of gears or pathways. Participants judged which way the target gear would turn and made their response by selecting either the left or right chute on the target gear or the button labeled “Jams!!.” The gear system on the lower left has a different number of gears in each pathway (i.e., one and three), but the pathways have the same parity (i.e., odd). The gear system on the lower right has a different number of gears in each pathway (i.e., one and two) and these pathways have different parity.

connecting gears, and a target gear (see the top panel of Fig. 1 for an example). Participants were asked to predict whether the target gear would turn clockwise, counterclockwise, or jam. The gear systems were presented as fueling stations; by predicting the movement of the target gear, the participant’s train could obtain fuel.

Two types of gear systems were presented. One-pathway systems had one set of gears connecting the driving and target gears. Two-pathway systems had two sets of connecting gears

(see the top and bottom panels of Fig. 1, respectively, for example). Both types of problems have the same underlying structure, but for two-pathway problems it is necessary to consider both pathways.

Previous research found that the vast majority of adult participants first solved gear-system problems by representing the physical force as it moved through the system (Dixon & Dohn, 2003; Schwartz & Black, 1996). Participants literally traced the movement of the clockwise-turning driving gear, each interconnecting gear, and finally the target gear. Most participants quickly discovered a second, more efficient strategy in which each gear is categorized as clockwise-turning or counterclockwise-turning without tracing the movement of any of the gears (see Dixon & Bangert, 2002, for an account of the processes underlying the discovery of these strategies). Many participants went on to discover a mathematical representation of the problem in which the parity of the number of gears (i.e., whether the number of gears was odd or even) indicates whether the target gear will turn clockwise or counterclockwise.

The question at issue for our current purposes is what processes allow participants to discover the mathematical, parity representation. According to the automatic abstraction view, as participants are exposed to an increasingly diverse set of problems, the probability of discovering an abstract representation should increase. That is, as the set of exemplars stored in memory becomes more diverse, their parallel retrieval will result in an abstract representation of their common underlying feature. In particular, retrieving multiple examples of clockwise-turning (all of which will share “odd number of gears” as a feature) and counterclockwise-turning problems (all of which will share “even number of gears” as feature) should result in the emergence of the parity relation. In essence, the idiosyncratic features of the individual exemplars “balance out,” allowing the common underlying structural features emerge. For example, consider a participant who has solved a large, one-pathway system and small, one-pathway system both of which turned clockwise. When clockwise systems are retrieved from memory, both values of system size are activated essentially canceling each other out.

According to the comparison-based abstraction view, explicitly comparing two pathways that have different numbers of gears, but the same parity (e.g., a pathway with three gears and a pathway with one gear), should result in discovering the parity relation. This prediction follows directly from structure-mapping theory. In structure-mapping theory, the process of comparison creates a maximally consistent structural alignment. Somewhat informally, this means that the entities in each problem are placed in one-to-one correspondence such that corresponding predicates also have corresponding arguments. Further, mappings that result in systems of relations are preferred (Gentner, 1983; Gentner & Markman, 1997; Holyoak & Thagard, 1989; Hummel & Holyoak, 1997; Keane, 1997). Therefore, comparing problems that have different numbers of gears, but the same parity, should highlight the corresponding parity relations (see Gentner & Medina, 1998 or Loewenstein & Gentner, 2001 for additional discussion of structure-mapping and extracting relations).

During problem solving, a participant might directly compare a one-pathway problem to another one-pathway problem, retrieved from memory. If this second problem had a different number of gears, but the same parity, the comparison necessary for extracting parity information could be made. However, it seems more likely that multiple problems of a variety of types would be retrieved, given that the participant has just solved many highly similar problems. Therefore, relevant between-problem comparisons may rarely occur.

Importantly, relevant within-problem comparisons (i.e., comparing two different-number, same-parity pathways) occur quite naturally as part of solving individual two-pathway systems. Some two-pathway systems have sets of gears with different numbers of gears, but the same parity (see Fig. 1, for an example). To solve these problems it is necessary to compare the pathways. Given the evidence that comparison invokes structure mapping (Gentner & Medina, 1998; Gentner & Namy, 1999; Markman & Gentner, 1993, 2000), we expected that solving these same-parity, two-pathway problems should result in structural alignment, highlighting and extraction of the parity relation. For example, suppose a participant encountered the gear system in the lower left panel of Fig. 1 labeled, “Different Number, Same Parity.” He or she would compute the turning force at the target gear for each pathway, perhaps by simulating the movement of the driving gear and the resulting movements of the gears in one pathway and then the other. Because the two pathways agree (i.e., result in the same turning force at the target gear), they are members of the same functional category, pathways that result in clockwise movement, in the current example. Joint membership in a relevant, but poorly understood category should result in structural alignment. That is, participants should compare the two pathways in a search for the property that makes them work the same way. Therefore, each time a participant encounters a different-number/same-parity problem, he or she should engage in structural alignment, activating the parity relation shared by the two pathways as well as other commonalities. Over repeated comparisons, the probability of using the parity relation as part of the representation of the problem should increase. For ease of exposition, we refer to different-number/same-parity problems as “same-parity” problems, henceforth.

Note that the prediction here is that repeatedly solving same-parity problems will result in use of the Counting-Parity strategy on subsequent trials, not on the same-parity problems themselves. Comparison, and ultimately extraction of the parity relation, occurs as a result of solving the problem with other strategies. Therefore, the accumulation of same-parity problems will function as a lagged predictor; the number of previously solved same-parity problems predicts discovery of the Counting-Parity strategy on future trials rather than the current trial.

Solving other two-pathway problems should not lead to the discovery of the parity relation. Specifically, when the two pathways have different number and also different parity (henceforth, different-parity problems), they should be classified as members of separate categories because they function differently (i.e., create opposite forces at the target gear). Namy and Gentner (2002) showed that presenting objects as members of different categories inhibited structural alignment (see also, Gentner & Namy, 1999). One simple explanation for this effect, suggested by Namy and Gentner, is that participants do not seek to align objects that do not share a relevant commonality. In the current context, when two pathways function differently, participants may not align them because there is no evidence that they share a common property. A second potential explanation is that differences between the two objects may be so highly salient or numerous that the alignment process terminates before deep structure is extracted. For example, different-parity problems have a different number of gears in each pathway. If participants search for a difference, the number of gears may be highly and immediately available such that deeper relations (e.g., parity) are not extracted. Therefore, we expected that exposure to same-parity, but not different-parity, problems would facilitate discovery of the parity relation.

## 1. Method

### 1.1. Participants

Fifty-six college students participated as one option to fulfill a course requirement.

### 1.2. Materials and procedure

Participants were asked to solve gear-system problems as part of a train race. The train race and the gear-system problems were presented on a Macintosh PowerPC. At each of 44 fueling stations, a gear system was presented. Each gear system was composed of a driving gear that always turned clockwise, a variable number of interconnecting gears, and a target or fuel gear. The target gear had a shelf that held a small pile of coal. If the target gear turned clockwise, the coal would slide down a chute on the right; if it turned counterclockwise the coal would slide down the chute on the left. It was also possible for two-pathway (but not one-pathway) systems to jam, because the two sets of interconnecting gears could create opposing forces at the target gear. Participants were asked to select the appropriate chute or a button labeled “Jams!!” at each fueling station. Appropriate visual events and sounds accompanied correct and incorrect predictions (i.e., for correct solutions: coal falling into the train and a trumpet sounding; for incorrect solutions: coal falling outside the train and a thudding sound). In addition, participants could watch their progress relative to an opponent train on a map. The opponent train was controlled by the computer (see Dixon & Bangert, 2002 for more complete details).

Participants completed the problems over two sessions; the first session contained 12 trials, the second contained 32 trials. Participants could solve the gear-system problems in any way they wished and could take as long as they wished to generate a solution. The gear systems varied along three dimensions: size (small: four or five gears, large: seven or eight gears), number of pathways (one or two interconnecting pathways), and whether or not an extraneous gear was present. Extraneous gears were not involved in the causal interconnecting pathways.

During the initial, 12-trial session, half the participants saw two different single-pathway gear systems (six trials each). The remaining participants saw two-pathway gear systems; one of these systems jammed (six trials), the remaining systems turned clockwise (three trials) or counterclockwise (three trials). To reduce the complexity of the early trials, none of the gear systems presented in the initial session had extraneous gears.

During the second, 32-trial session, eight single-pathway gear systems were each presented twice (16 trials). Four two-pathway systems were presented twice; these problems jammed (eight trials). Eight two-pathway systems were presented once; half these problems turned clockwise (four trials), half turned counterclockwise (four trials). We used eight, rather than four, different types of non-jamming two-pathway problems to allow both different-number/same-parity problems and same-number/same-parity problems to be included in the set. (No predictions were made for the accumulation of same-number/same-parity problems and they were not considered in the computation of predictor variables.)

The order of the trials within each session was randomized for each participant. The manipulated variation in the problems presented during the first session ensured additional

variability in participants' experience with the gear systems above and beyond that provided by randomizing the order. In the analyses reported below, we capture both these types of manipulated variation via predictor variables. Including additional variables to code explicitly for the between-subjects conditions does not alter the results reported below.

### 1.3. Strategy coding

Participants were asked to think aloud as they solved the gear-system problems. Three major strategies were coded. (1) Figure-Eight: Participants literally traced the force from the driving gear to the target gear (the movement resembles a figure-eight). (2) Left-Right: Participants categorized the gears in an alternating sequence: "left," "right," "left," etc. (Even college participants tend to use "left" and "right" rather than the more appropriate clockwise, counterclockwise designations.) The movement of gears was not traced or mentioned in solving the problem. (3) Counting-Parity: Participants counted the gears and used the parity of the count to determine whether the driving gear turned left or right. Strategies that did not fit any of these categories were coded as "Other." These "Other" strategies were relatively low frequency strategies that are presented collectively here (see Dixon & Bangert, 2002 for details on these strategies individually). The sessions were videotaped to allow for reliability coding. Reliability was very high; three independent raters agreed on 96% of trials from five randomly selected cases.

## 2. Results

### 2.1. Accuracy and response time

The upper panel of Fig. 2 shows the mean proportion correct for the different types of gear systems. A 2 (Size)  $\times$  2 (Number of Pathways)  $\times$  2 (Extraneous Gear) ANOVA showed that participants made significantly more errors on two-pathway than one-pathway problems,  $F(1, 55) = 14.56$ ,  $MSE = 0.01$ . (An alpha level of .05 was used for all analyses unless otherwise indicated.) This effect did not depend significantly on whether an extraneous gear was present,  $F(1, 55) = 2.70$ ,  $MSE = 0.01$ ,  $p > .10$ , or system size,  $F < 1$ . None of the other main effects or interactions were significant,  $F$ 's  $< 1$ .

The lower panel of Fig. 2 shows the mean response time for the different types of gear-system problems. A 2 (Size)  $\times$  2 (Number of Pathways)  $\times$  2 (Extraneous Gear) ANOVA showed that participants took longer to solve larger gear systems,  $F(1, 55) = 40.64$ ,  $MSE = 63.93$ , and two-pathway problems,  $F(1, 55) = 34.26$ ,  $MSE = 61.41$ . Solving problems with more gears takes longer, as does comparing the results of two pathways. The presence of an extraneous gear did not affect response time,  $F < 1$ , nor were any of the interactions significant,  $F$ 's  $< 1$ .

Performance with the Figure-Eight strategy was slower than with the Left-Right strategy,  $M$ 's = 26.47, 14.59,  $F(1, 31) = 17.62$ ,  $MSE = 118.34$ , but not significantly less accurate,  $M$ 's = 0.93, 0.94,  $F < 1$ . The Left-Right strategy was slower than the Counting-Parity strategy,  $M$ 's = 14.59, 9.53,  $F(1, 20) = 13.63$ ,  $MSE = 16.88$ ; but not significantly less accurate,  $M$ 's = 0.94, 0.98,  $F(1, 20) = 4.32$ ,  $MSE = 0.01$ . In sum, all the strategies produced

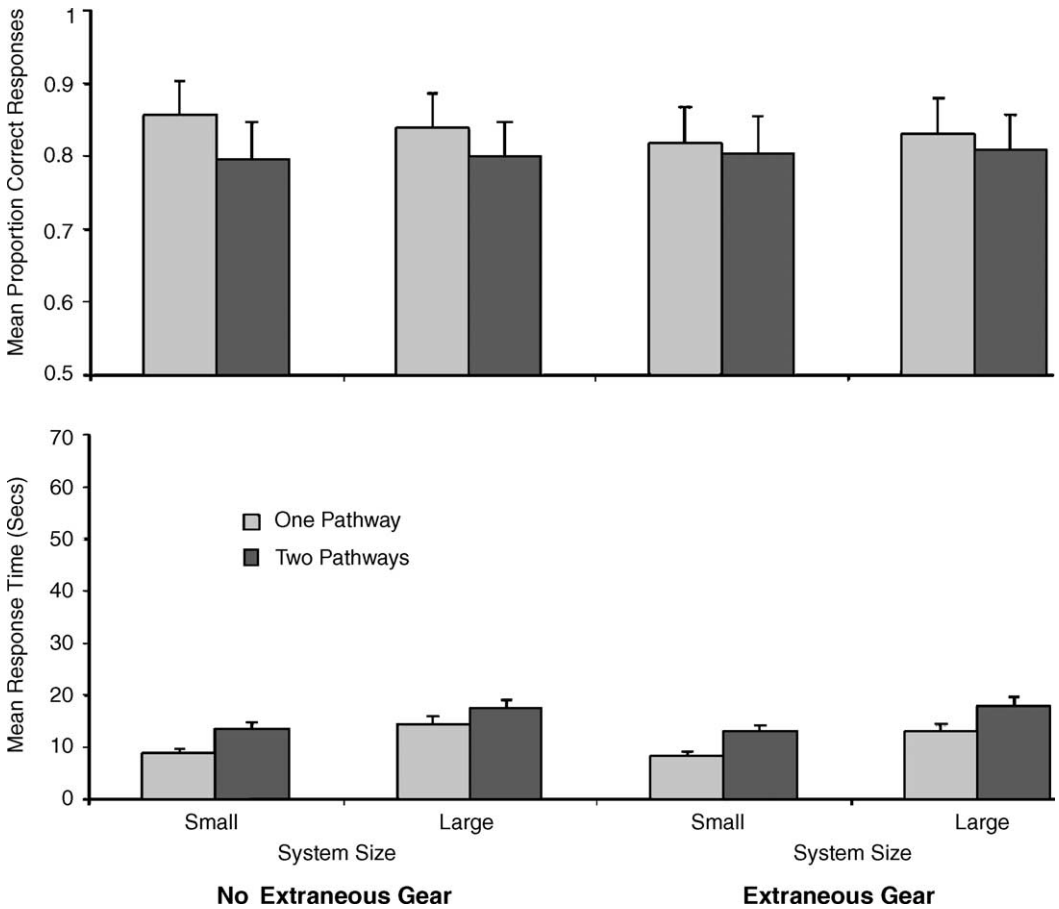


Fig. 2. The mean proportion of correct responses (top panel) and mean response time in seconds (bottom panel) for each type of gear system are presented above. Results for gear systems without extraneous gears are presented on the left; results for systems with extraneous gears are on the right. Error bars indicate one standard error.

very accurate performance, but Left-Right was faster than Figure-Eight, and Counting-Parity was faster than Left-Right.<sup>2</sup>

### 2.2. Strategy use across trials and problem types

Fig. 3 shows the percentage of strategy use across trials. Consistent with past research, the vast majority of participants (70%) first used the Figure-Eight strategy. Many participants discovered the Left-Right strategy over the first session or relatively early in the second session. Approximately 35% of participants used the Left-Right strategy on the last trial of Session 1, about 60% used it on Trial 10 of the second session.

None of the participants used Counting-Parity as their first strategy, indeed very few participants discovered it during the first session. Twenty-five participants discovered Counting-Parity over the course of the study.



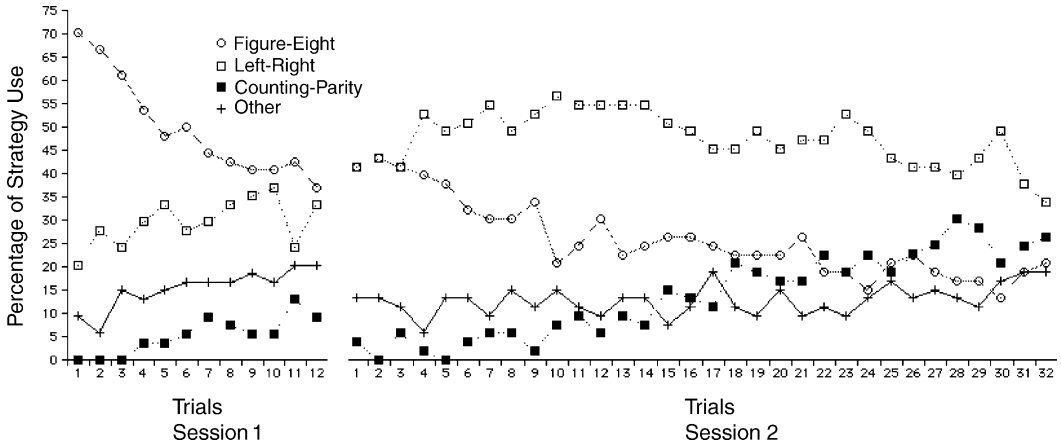


Fig. 3. The percentage of strategy use is shown across trials. The results from the first session (12 trials) are on the left; results from the second session (32 trials) are on the right.

The different strategies were used across the various types of gear systems. Participants used Figure-Eight, Left-Right, and Counting-Parity on 35, 44, and 10% of small-system trials and on 29, 43, and 14% of large-system trials, respectively. (Strategies in the “Other” classification comprised the remaining trials.) The distribution of strategies depended on gear-system size,  $\chi^2(3, N = 2464) = 16.22$ .<sup>3</sup> Participants used Figure-Eight, Left-Right, and Counting-Parity on 31, 45, and 12% of the one-pathway trials and 32, 43, and 12% of two-pathway trials,  $\chi^2(3, N = 2464) = 1.98$ , n.s. For gear systems that did not have an extraneous gear, Figure-Eight, Left-Right, and Counting-Parity were used on 34, 41, and 12% of trials. When an extraneous gear was present the strategies were used on 28, 49, and 11% of trials. The distribution of strategies depended on whether the gear system had an extraneous gear,  $\chi^2(3, N = 2464) = 13.28$ . Because the hypotheses under test do not make predictions about the distribution of strategies across problem types and the data become quite sparse for strategies discovered on later trials, we do not report strategy distributions for combinations of the size, number of pathways, and presence of an extraneous gear.

Finally, we consider the distribution of strategies on the subtypes of two-pathway problems that are important in later analyses. On same-parity problems, participants used Figure-Eight, Left-Right, and Counting-Parity on 34, 49, and 6% of trials. On different-parity problems, Figure-Eight, Left-Right, and Counting-Parity were used on 30, 41, and 16% of trials. The distribution of strategies was related to the type of two-pathway problem,  $\chi^2(3, N = 924) = 18.28$ .

### 2.3. Predicting the discovery of Counting-Parity

#### 2.3.1. Event history analysis

In the current study, we are interested in how (a) exposure to an increasingly diverse sample of exemplars and (b) exposure to two-pathway problems of a particular type predict the discovery (i.e., literally the first use) of Counting-Parity. A class of statistical techniques, called event

history analysis, has been developed to analyze longitudinal event data (see Allison, 1984 or Singer & Willett, 2003 for introductions). Event history analysis allows one to model how the probability of a discrete event, such as the discovery of new strategy, changes as a function of the participant's history prior to that point in time. Two concepts are important for understanding event history analysis: the risk set and the hazard rate. The risk set is comprised of all the participants who have not yet experienced the event of interest (e.g., have not yet discovered Counting-Parity). The risk set changes from trial to trial. Participants who experience the event (discover Counting-Parity) drop out of the set. The hazard rate is the probability of the event occurring on a given trial, called the focal trial, and is estimated by dividing the number of participants who experienced the event on the focal trial by the number currently in the risk set. The hazard rate is, therefore, the population parameter of interest. Predictor variables can be constructed from the participant's history prior to the focal trial (e.g., the number of same-parity problems experienced before the focal trial).

### 2.3.2. Automatic schema abstraction

According to the automatic schema abstraction hypothesis, as the diversity of the category exemplars increases, the probability of extracting the underlying relational structure should increase. In the current study, there were eight unique clockwise-turning exemplars and eight unique counterclockwise-turning exemplars. These problems differed along three dimensions: size (small or large), number of pathways (one or two), and whether an extraneous gear was present. Participants encountered the exemplars in random order over the course of their problem solving (subject to the constraints in initial session described in Section 1). As an individual participant's corpus of exemplars becomes more diverse (i.e., contains more unique exemplars), he or she should be increasingly likely to discover the Counting-Parity strategy.

To test this hypothesis, we used event-history analysis to analyze how the trial-by-trial accumulation of unique exemplars affected the probability of discovering Counting-Parity. The set of exemplars to which a participant has been exposed becomes increasingly diverse across trials. We constructed two predictor variables to reflect the accumulation of unique exemplars. One variable was number of unique clockwise-turning exemplars the participant had viewed; the other was the number of unique counterclockwise exemplars the participant had viewed. These variables simply index the degree to which the participant had been exposed to the variety of clockwise and counterclockwise exemplars, respectively. The question of interest was whether the increasing variety of exemplars prior to the focal trial would predict discovery of the Counting-Parity strategy on that trial. We also included whether the participant had previously used the Left-Right strategy, because past work suggests that this is an important precursor to discovering Counting-Parity (Schwartz & Black, 1996). The event-history models were fit with logistic regression.

When the number of unique clockwise and the number of unique counterclockwise exemplars were used to predict the discovery of the Counting-Parity strategy, the overall model was not significant,  $LR \chi^2(3, N = 1791) = 4.57, p > .05$ .<sup>4</sup> Nor did the individual clockwise or counterclockwise predictors make significant contributions to the model when considered separately, change in  $LR \chi^2$ 's(1) = .79, .52, respectively. The accumulation of diverse category exemplars did not predict discovery of the Counting-Parity relation.<sup>5</sup>

A second approach to this issue is to consider the variability among the features within each category (i.e., clockwise-turning vs. counterclockwise-turning) rather than the diversity of the category exemplars. This approach assumes that the features are coded independently of each other; examining the diversity among the exemplars (as above) assumes that features combine non-additively (Foard & Kemler-Nelson, 1984; Kemler-Nelson, 1993; Ward, 1983) and might, therefore, underestimate the effects of within-category variability. To assess this possibility, we created a binary variable (with the values  $-1$ ,  $1$ ) for each of the three features (i.e., size, number of pathways, and presence of an extraneous gear). Using these three variables, we coded the features of each gear system experienced by an individual prior to the focal trial. Summing a feature variable across the gear systems within a category provides an index of the degree of variability for that feature prior to the focal trial. Scores that are closer to zero indicate greater variability (i.e., the values of the features, coded  $-1$  and  $1$ , are balanced). For example, consider an individual who prior to Trial 19 (an arbitrarily selected trial) saw nine clockwise-turning systems. Assume that of these nine systems, six were large and three were small. This individual would have a score of three for the size feature at Trial 19. If the individual saw a small system that turned clockwise on Trial 19, the score for the size feature would change to two at Trial 20, indicating an increase in variability for this feature.

Summing the unsigned scores across features gives an index of the degree of variability for features within the category. For example, considering again Trial 19 for the individual above, the value of the size variable (3) would be summed with values of the “number of pathways” and “extraneous gear” feature variables. The sum of these variables indexes the degree to which the features vary across the category exemplars that were experienced prior to the focal trial. Therefore, for each individual, we computed the sum of the variability among the features for both the clockwise and counterclockwise categories prior to each focal trial. We used these measures of variability as predictors of discovering the Counting-Parity strategy in an event history analysis. As above, we also included whether the participant had previously discovered the Left-Right strategy in the model. The overall fit of the model was not significant,  $LR \chi^2(3, N = 1791) = 5.42$ , nor did the measures of variability contribute significantly to the model, when considered separately, change in  $LR \chi^2$ :  $s(1) = 1.91$ ,  $1.00$ , for the clockwise and counterclockwise categories, respectively.

### 2.3.3. Comparison-based abstraction

According to the comparison-based abstraction hypothesis, repeatedly solving two-pathway problems in which the pathways have different numbers of gears, but the same parity (e.g., one and three), should lead to a similarity comparison, structural alignment of the gear pathways, and eventual representation of the parity relationship. This effect should be specific to these types of two-pathway problems. In problems that contain two pathways of different parity (e.g., two and three), the pathways create opposing forces at the target gear. Therefore, the perceived difference in forces should lead to categorizing the pathways as different types which should not facilitate alignment.

To test this hypothesis, we calculated the number of same-parity problems and different-parity problems that occurred prior to each focal trial. There were four problems of each type in the corpus. These predictors were entered into an event-history analysis to assess whether participant’s history of exposure to these different problem types predicted the discovery of

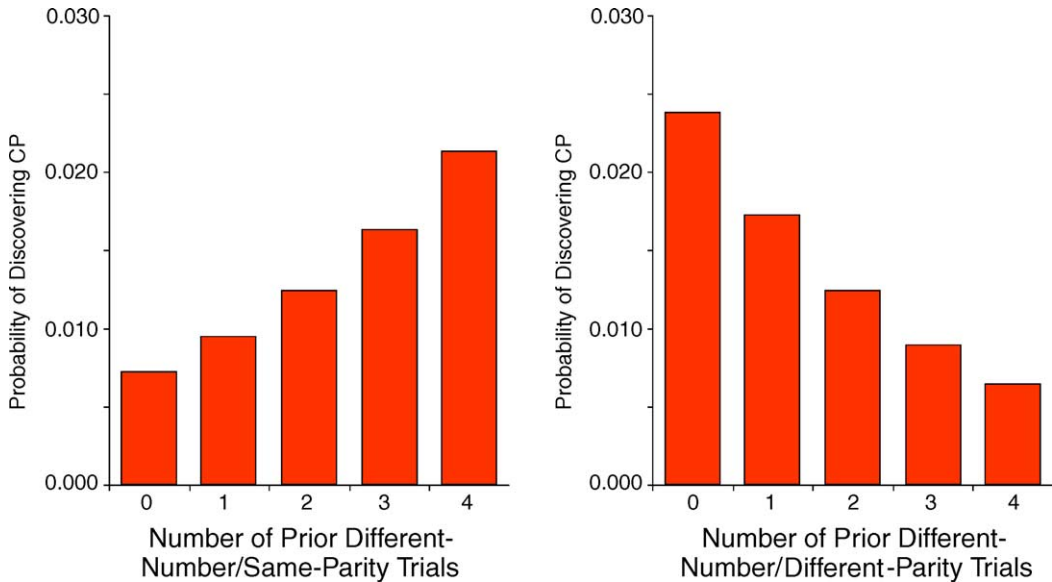


Fig. 4. Model predictions from the event history analysis are shown in the two panels above. The panel on the left shows the probability of discovering the Counting-Parity (CP) strategy as a function of the number of same-parity (but different number), two-pathways systems encountered prior to the current trial. The panel on the right shows the probability of discovering Counting-Parity as a function of the number of different-parity, two-pathway systems encountered prior to the current trial.

Counting-Parity. We included whether the participant had previously discovered the Left-Right strategy in these models as well.

The overall fit of the model was significant,  $LR \chi^2(3, N = 1791) = 9.19$ . All three variables contributed individually to the fit of the model: prior use of Left-Right,  $B = .91$ ,  $SE = 0.47$ , change in  $LR \chi^2(1) = 4.25$ ; number of same-parity problems,  $B = .27$ ,  $SE = 0.12$ , change in  $LR \chi^2(1) = 4.89$ ; number of different-parity problems,  $B = -.33$ ,  $SE = 0.14$ , change in  $LR \chi^2(1) = 6.22$ .<sup>6</sup> Fig. 4 shows the model predictions for the two substantive variables of interest, number of same- and different-parity problems. As can be seen in the figure, as the number of same-parity problems increases, the probability of discovering the Counting-Parity strategy increases. As the number of different-parity problems increases, the probability of discovering Counting-Parity decreases.<sup>7</sup> Consistent with the comparison-based abstraction hypothesis, repeatedly solving problems that encourage structural alignment increases the probability of discovering the parity relationship.

Note that in the analyses above discovery of the Counting-Parity strategy could occur on one- or two-pathway problems; it is an effective strategy for both types of problems. Including the number of pathways in the current problem (the problem presented on the focal trial) does not change the estimated effects of the theoretically important lagged predictors: number of same-parity problems,  $B = .27$ ,  $SE = 0.12$ , change in  $LR \chi^2(1) = 4.86$ ; number of different-parity problems,  $B = -.33$ ,  $SE = 0.14$ , change in  $LR \chi^2(1) = 6.27$ . The effects are the same when the number of pathways for the focal trial problem is included in the model.

### 3. Discussion

Our results show that participants' discovery of the Counting-Parity strategy, a strategy based on the mathematical relation of parity, is predicted by exposure to problems that encourage similarity comparisons. Problems that have two pathways, each with a different number of gears, but the same parity within each pathway, provide the same force at the target gear. These pathways, therefore, function similarly and invite a similarity comparison. As participants encounter these same-parity problems, the probability of discovering the parity relation increases.

Interestingly, encountering different-parity problems actually reduces the probability of discovering the Counting-Parity strategy. [Namy and Gentner \(2002\)](#) showed that structural alignment was facilitated when objects were presented as members of the same category. However, when the same objects were presented as members of different categories structural alignment was inhibited. In the current context, pathways with different parity produce different forces at the target gear and should, therefore, be classified as members of separate categories. Our results suggest that categorizing the pathways as different does not just prevent structural alignment; if this were the case there should be no effect of experiencing these problems. The negative effect of different-parity problems seems to suggest that these problems may highlight aspects of the problem representation that compete with parity. For example, categorizing the pathways as different may highlight salient and irrelevant differences, such as in the number or spatial configuration of the gears. This may result from a structural alignment process that terminates early, because differences between the two pathways are easily available. For example, a difference in parity necessarily entails a difference in number, length, and other perceptually available, quickly alignable features. Repeatedly attending to these irrelevant features may maintain them in the current problem representation and thereby inhibit the emergence of a mathematical representation.

The degree to which participants had experienced a diversity of exemplars from the clockwise- and counterclockwise-turning problem categories did not predict discovery of Counting-Parity. As a participant's category expands to include a wider variety of instances, the joint activation of those instances would be expected to result in the emergence of the relation that is the common denominator for the category. In this case, all the clockwise-turning problems share the same parity (i.e., an odd number of gears); as do the counterclockwise-turning problems (i.e., an even number of gears). We also tested whether the diversity among the features within a category might predict the discovery of parity. These two ways of representing the diversity make different assumptions about how the features are processed (i.e., in combination vs. independently, respectively), but the conclusions from the analyses are the same. The diversity within the categories did not drive the discovery of the parity relation.

This may be due to insufficient diversity in our set of exemplars. Recall that the problems varied across three dimensions: size, structure, and whether an extraneous gear was present. The automatic schema abstraction process may require an even broader set of exemplars to extract relational information. The current results suggest that the automatic schema abstraction process is not necessary for extracting a mathematical relation, but cannot address whether the process might be sufficient under some alternative set of conditions.

The comparison-based abstraction process appears to be sufficient to extract the mathematical relation of parity in this domain. Our results suggest that this process can occur locally; the crucial problems created a similarity comparison between two simultaneously presented pathways, as opposed to comparing the current problem with one retrieved from memory. The opportunity to make local, simultaneous comparisons may dramatically facilitate discovery. The issue is that retrieved representations of stored problems may often be lacking in some crucial details. For example, in the gear domain the exact number of interconnecting gears might not be easily recalled (as in a gist representation, Reyna & Brainerd, 1995). Without that crucial detail, a similarity comparison between a current and previous problem would not reveal parity information, because parity is an attribute of the exact number of gears, not the approximate magnitude of gears (e.g., few, many). To the extent that extraction of the underlying relation requires the comparison of detailed problem information (e.g., the number of gears in each problem), comparing the current problem to one retrieved from memory will be unlikely to result in discovering the relation.

The current results have implications for our understanding of how abstract relations can be spontaneously extracted during problem solving. Note that discovering the mathematical relation was not in any way required in the current study. Participants already had a perfectly adequate solution strategy and were under no time pressure. It seems likely that structure mapping was engaged as part of routine problem solving rather than as part of an active search for a better strategy. Therefore, spontaneous discovery of mathematical and other abstracted relations may be the result of serendipitous opportunities to compare alignable problems.

The event-history approach used in the current study complements more traditional experimental approaches used in previous work on extracting relations (e.g., Bernardo, 2001; Namy & Gentner, 2002; Ross & Kennedy, 1990). Statistical models of cognitive processes, particularly models that analyze real-time behavior (as in the current study), can play an important role in theory testing. Theories of cognition must make predictions about real-time performance of individuals. We should be able to find evidence for theoretically important processes as participants operate moment-to-moment or trial-to-trial.

Although event-history analysis utilizes a regression format, participants were randomly assigned to orders of problem types; the ordering of the problem types completely determined the values of the predictor variables. Therefore, structurally, the current study is an experiment in which participants have been randomly assigned to different sequences of experience; the sequences of experience contain the levels of the relevant independent variables. The effects of this manipulation were evaluated with a statistical methodology that was specifically designed to handle this type of data, discrete events occurring across time. Like other regression methodologies, this analytic approach allows for the construction of many different predictors. This allows the effect of important covariates to be integrated into the model easily and naturally. Of course, because the number of predictors one can construct is limitless (e.g., Mosteller & Tukey, 1977 ladder of powers), it is important for predictors to be tied to substantive theory.

Finally, the results suggest that crafting comparisons that allow for alignment of the underlying relational structure may help students discover a mathematical representation. Explicit, local comparisons, in which the problems are simultaneously presented, may be best because of the potential difficulty in recalling detailed information and recalling the appropriate problem.

Presenting a diversity of examples may also help (see Bruner, 1966; National Council of Teachers of Mathematics, 2000), but our results suggest that even with a sizable corpus of unique exemplars spontaneous discovery of the relation may not occur. A diversity of examples may actually hinder the comparison-based abstraction process if, as in the current domain, some problems do not facilitate alignment.

## Notes

1. Reeves and Weisberg (1994) used the term strategic schema abstraction instead of comparison-based abstraction. The subsequent literature has predominantly used the term comparison.
2. These analyses compare the performance of participants who used both the strategies, therefore, the degrees of freedom vary across comparisons.
3. The  $\chi^2$  tests of significance in this section assume that the observations are independent, an assumption that can be difficult to evaluate for time-series data (see Wickens, 1989). Therefore, we present these analyses for descriptive purposes; the associated significance tests should be viewed with caution.
4. In this trial-level analysis,  $N$  indicates the total number of trials on which participants were “at risk” for the event (i.e., had not yet discovered Counting-Parity).  $LR \chi^2$  is the likelihood ratio chi-square, an index of model fit that is conceptually analogous to the multivariate  $R^2$  and its associated  $F$  in OLS regression. Improvements in model fit, as one adds predictors, are evaluated by assessing the significance of the change in the  $LR \chi^2$ . See Menard (1995) for details.
5. For these analyses, we assume, somewhat unrealistically, that recall of previously seen examples is near perfect. This assumption should work in favor of the automatic schema abstraction hypothesis; the fewer problems lost to forgetting, the more diverse the category. However, to ensure that this assumption was not obscuring the predictive effects of the diversity of category exemplars, we conducted an additional analysis including both a primacy and recency effect. Problems experienced early (in the first four trials) were weighted more heavily: using the weights 9, 6, 4, 2, for Trials 1 through 4, respectively. Problems experienced most recently were also weighted more heavily: using the weights 2, 4, 6, 9, for trials, Focal-4, Focal-3, Focal-2, and Focal-1, respectively. Unique exemplars that are easier to recall, because of their position (very early or very recent), contribute more to the diversity score. The weighted measures of category diversity were entered into an event history analysis, analogous to that conducted with the unweighted measures. The overall fit of the model was not significant,  $LR \chi^2(3, N = 1791) = 5.01$ ; when considered separately the individual predictors did not significantly contribute to the model.
6. For both the automatic and comparison-based abstraction models we conducted additional analyses to address the possibility that accuracy and response times on the relevant trials were important factors. The results showed that including information about the accuracy of participants’ responses and response times for these trials did not significantly improve the fit of the models.

7. Note that the small probability values are expected, because the model predicts the probability of an individual discovering the strategy on a single trial. For example, if discoveries were randomly distributed across trials and all participants discovered the strategy, the average probability of discovery per trial would be .10. If discoveries occurred only during the second session (as the majority of the Counting-Parity discoveries do), the average probability of discovery would be .04. When non-discoverers are present, as in the current study, the probability of discovery becomes even smaller, because non-discoverers remain in the risk set and always contribute to the denominator.

## Acknowledgments

This work was supported in part by NSF Grant No. BCS-9996353.

## References

- Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA: Sage.
- Bernardo, A. B. I. (2001). Principle explanation and strategic schema abstraction. *Memory & Cognition*, *29*, 627–633.
- Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Belknap Press.
- Chen, Z. (1999). Schema induction in children's analogical problem solving. *Journal of Educational Psychology*, *91*, 703–715.
- Dixon, J. A., & Bangert, A. (2002). The prehistory of discovery: Precursors of representational change in solving gear-system problems. *Developmental Psychology*, *38*, 918–933.
- Dixon, J. A., & Dohn, M. C. (2003). Redescription disembeds relations: Evidence from relational transfer and use in problem solving. *Memory & Cognition*, *31*, 1082–1093.
- Foard, C. F., & Klemmer-Nelson, D. G. (1984). Holistic and analytic modes of processing: The multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General*, *113*, 94–111.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*, 45–56.
- Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*, 263–287.
- Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, *14*, 487–513.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–39.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.
- Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*, 414–466.
- Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.
- Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*, 295–355.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.
- Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*, 374–419.



- Keane, M. T. (1997). What makes an analogy difficult? The effects of order and causal structure on analogical mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 946–967.
- Kemler-Nelson, D. G. (1993). Processing integral dimensions: The whole view. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 1105–1113.
- Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, 67, 2797–2822.
- Loewenstein, J., & Gentner, D. (2001). Spatial mapping in preschoolers: Close comparisons facilitate far mappings. *Journal of Cognition and Development*, 2, 189–219.
- Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25, 431–467.
- Markman, A. B., & Gentner, D. (2000). Structure mapping in the comparison process. *American Journal of Psychology*, 113, 501–538.
- Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem solving, and induction. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189–223). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.
- Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison-Wesley.
- Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, 131, 5–15.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, 115, 381–400.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, 7, 1–75.
- Ross, B. H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology*, 16, 371–416.
- Ross, B. H. (1989). Reminders in learning and instruction. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 438–469). Cambridge: Cambridge University Press.
- Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 42–55.
- Ross, B. H., & Warren, J. L. (2002). Learning abstract relations from using categories. *Memory & Cognition*, 5, 657–665.
- Ross, G., Nelson, K., Wetstone, H., & Tanouye, E. (1986). Acquisition and generalization of novel object concepts by young language learners. *Journal of Child Language*, 13, 67–83.
- Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science*, 20, 457–497.
- Shrager, J., & Siegler, R. S. (1998). A model of children's strategy choices and strategy discoveries. *Psychological Science*, 9, 405–410.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.
- Ward, T. B. (1983). Response tempo and separable-integral responding: Evidence for an integral-to-separable processing sequence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 103–112.
- Wickens, T. D. (1989). *Multway contingency tables for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.