Characterizing the intuitive representation in problem solving: Evidence from evaluating mathematical strategies

JAMES A. DIXON and COLLEEN F. MOORE
University of Wisconsin, Madison, Wisconsin

Two experiments were conducted to investigate the nature of the intuitive problem representation used in evaluating mathematical strategies. The first experiment tested between two representations: a representation composed of principles and an integrated representation. Subjects judged the correctness of unseen math strategies based only on the answers they produced for a set of temperature mixture problems. The distance of the given answers from the correct answers and whether the answers violated one of the principles of temperature mixture were manipulated. The results supported the principle representation hypothesis. In the second experiment we manipulated subjects' understanding of an acid mixture task with a brief paragraph of instruction on one of the principles. Subjects then completed an estimation task intended to measure their understanding of the problem domain. The evaluation task from the first experiment was then presented, but with acid mixture instead of temperature mixture. The results showed that intuitive understanding of the domain mediates the effect of instruction on evaluating problems. Additionally, the results supported the hypothesis that subjects perform a mapping process between their intuitive understanding and math strategies.

Problem solving often requires applying some formal strategy or algorithm to a problem. How these formal strategies are selected and evaluated is a central issue in theories of problem solving. Many theories of problem solving have suggested that the person's qualitative or intuitive representation of a problem is an important factor in arriving at an appropriate formal strategy. For example, Larkin (1983) proposed that when people are solving physics problems with mathematics, they construct a representation of the problem and use this representation to select appropriate math strategies. Similar frameworks have been proposed for the domain of children's counting (Briars & Siegler, 1984; Greeno, Riley, & Gelman, 1984), word problems (Briars & Larkin, 1984; Kintsch & Greeno, 1985), and mixture problems (Dixon & Moore, 1996; Reed & Evans, 1987). According to these models, what the person understands about the problem domain affects the type of strategies they select to solve the problem.

Most models of problem solving also agree that the intuitive representation of the problem domain specifies how the variables in the task relate to one another. A number of computer simulations have demonstrated that information of this type is sufficient to select a formal strategy (see, e.g., de Kleer, 1975, 1977; Kintsch & Greeno, 1985; Larkin, 1983).

There is agreement on the importance of the representation of the problem domain as well as agreement that the representation must specify the relationships between variables in the domain, but it is not clear how the representation is used in the selection and evaluation of formal strategies. In past work, we have proposed that people engage in a mapping process between their intuitive representation and a representation of the formal strategies in order to select a strategy (Dixon & Moore, 1996). Consistent with the work cited above, we proposed that people have a distinct representation of the problem domain. We also proposed that people have a representation of the formal strategies that exist in their repertoire. According to our hypothesis, some mapping process takes place between what the person understands about the task and his/her representation of the formal strategies. The mapping process results in the selection of a formal strategy.

For example, consider a college student attempting to solve a temperature mixture task. The student is shown an initial container of water with the quantity and temperature specified. The student is then shown a second container the contents of which will be added to the initial container. The quantity and temperature of the added water is also specified. The task is to compute the temperature of the water when the contents are combined. Suppose the student's understanding of temperature mixture includes the knowledge that the combined water temperature will be between the temperatures of the contents of the initial and added containers. We hypothesize that the student would use this knowledge to help select a math strategy. One way people might do this is to map what they understand about the problem domain and what they understand about the functional properties of

This research was supported in part by a grant from the Graduate School, University of Wisconsin, Madison. The authors thank Katrina Phelps, Catherine Silverman, Don Schwartz, and Sidney Dechovitz for help with data analyses, and also thank the participating schools. Correspondence should be addressed to J. A. Dixon, Department of Psychology, Trinity University, 715 Stadium Dr., San Antonio, TX 78212-7200 (e-mail: jdixon@trinity.edu).

math strategies. In this example, the student might arrive at unweighted averaging of the two temperatures on the basis that averaging will produce answers that are between the temperatures of the two containers. How such mapping takes place is an important question.

One major obstacle to investigating this mapping process is that the nature of the intuitive representation that people use to arrive at a math strategy is unknown. There are a number of ways in which the relationships between variables might be represented in order to achieve a mapping to a math strategy. In past research, we proposed

that problems involving the mixture of different temperatures are represented in terms of principles (Dixon & Moore, 1996). The principles define the relationships between variables in the problem domain. Other researchers have also proposed principles as the intuitive representation for other problem domains (e.g., Reed & Evans, 1987). However, the hypothesis that people use principles to guide problem solving with formal strategies, as opposed to some other representation, has not been tested. Without knowing how the intuitive knowledge of the domain is represented, it is very difficult to

Figure 1. Representation of temperature mixture principles as fuzzy sets specifying the degree of membership (ordinate) of each possible temperature value (abscissa).

define the mapping process between the intuitive representation and the representation of formal strategies.

The purpose of the first experiment, therefore, was to test two hypotheses about the nature of the intuitive representation. The first hypothesis is that subjects represent the domain of temperature mixture in terms of a particular set of principles that operate in that domain. According to this hypothesis, the principles are explicitly represented and used in selecting and evaluating math strategies. We propose an instantiation of this hypothesis in which each principle specifies a fuzzy set of potential answers (Zadeh, 1965). The upper left panel of Figure 1 shows an example of a principle represented in this way for the temperature mixture task. The principle, which we call the "range principle," specifies that the final temperature must be between the initial and added temperatures. The potential answers are on the horizontal axis. Degree of membership in each set is specified by the height of the curve. The members of the set are all between the two temperatures, T_i (initial temperature) and T_a (added temperature). Goodness of membership increases as the answers approach the center of the region bracketed by the two combined temperatures. The differences in degree of membership reflect the idea that answers near the center of the region are more characteristic of the principle than are answers at the extreme ends of the region. Fuzzy set representations of the other principles involved in representing temperature mixture, shown in the other panels of Figure 1, are discussed later. (See Ragade & Gupta, 1977, or Zadeh, 1965, for discussions of fuzzy set theory. See Massaro, 1994, Oden, 1984, 1988, and Oden & Massaro, 1978, for examples of fuzzy set theory applied to modeling cognitive and perceptual processes.)

The second hypothesis is that subjects represent the relationships in the domain in an integrated manner. Figure 2 shows one prominent example of this type of representation. Here the temperature mixture problem is represented in the form proposed by information integration theory (Anderson, 1974, 1991). Although the integration function receives its input from an input function and delivers the results to an output function, knowledge about the domain is contained solely in the integration function itself. We refer to this as an integrated representation because knowledge about the domain is accessible only by passing the values for the variables through the integration function. Further, although the integration function correctly integrates the temperature and quantity information, it yields answers

Figure 2. Schematic representation of Anderson's (1974, 1991) information integration theory.

only to particular problems. That is, relationships between variables are not represented explicitly. Therefore, the only way to evaluate a potential answer is to compare it with the results obtained by passing values through the integration function.

Another example of an integrated representation is a "mental model," in which the problem situation is represented by some working model of the relationships between variables in the problem. For example, de Kleer's (1975, 1977) NEWTON system uses an integrated model of the entities in the problem situation to guide selection of formal strategies. A set of heuristic rules operates on the integrated representation to select the formal strategies.

Either type of representation, principle or integrated, can explain the current evidence on the relationship between understanding of the domain and use of mathematical strategies. For example, some of the best evidence about this relationship comes from the domain of physics problem solving. In one study, subjects who mentioned appropriate principles while categorizing problems performed more accurately on later problems than did subjects who did not mention principles (Hardiman, Dufresne, & Mestre, 1989). However, it is unclear from this evidence whether the principles are explicitly represented or are extracted from an integrated representation as part of the process of verbalizing knowledge.

The distinction between a representation composed of independent principles and one that forms an integrated whole is important in studying how the intuitive representation influences the selection and evaluation of mathematical strategies. If knowledge of the domain is represented in terms of principles, subjects may select or evaluate a math strategy by examining the match between the math strategy and all the principles or some subset of them. Because the principles are represented independently, they may be used independently in the selection or evaluation process.

If intuitive knowledge of the domain is represented in an integrated manner (e.g., cognitive algebra or a working model), the different relationships between the variables in the task are not represented independently and cannot be used separately. Obviously, then, the nature of the intuitive representation will have major ramifications for models of problem solving.

In both experiments reported here, we had subjects evaluate math strategies on the basis of answers supposedly generated by the math strategies. Subjects were shown two problems complete with answers. They were told that both problems had been solved by the same math strategy and were asked to evaluate whether that strategy was correct or not.

There were two motivations for having subjects evaluate problems. First, an important part of problem solving is evaluating the effectiveness of different formal strategies, so the type of representation used in this process was of interest. Second, we had hypothesized in past work (Dixon & Moore, 1996) that subjects select a math strategy to solve a particular problem by comparing the functional properties of the math strategy to those of the

problem domain. We intended the evaluation task to simulate that part of the strategy selection process that involves evaluating how well the math strategy matches intuitive understanding. The information about what the math strategy does is given by the relationships between the answers and the problems. Subjects may be able to use this information as they would their stored knowledge of how each math strategy functions; that is, they may compare it to their intuitive understanding of the task to assess its appropriateness.

EXPERIMENT 1

In order to test the integrated and principle representations hypotheses, we had college students evaluate answers to temperature mixture problems. In our problems, the initial container of water, which always contained 2 cups at 40°, was presented first. A second container, which would be added to the initial container, was then presented. The contents of this added container varied across trials. The question was, "What will the temperature of all the water together be?"

A number of principles govern the temperature mixture domain. The principles are explained in Figure 1. According to the hypothesis of principle representation, subjects explicitly represent these principles. Two lines of evidence support the idea that subjects represent these principles. First, developmental differences in understanding mixture tasks are consistent with the hypothesis that children are acquiring the principles that govern the tasks (Ahl, Moore, & Dixon, 1992; Dixon & Moore, 1996; Moore, Dixon, & Haines, 1991). Second, when performing mixture tasks, subjects sometimes mention principles, both spontaneously and when asked to explain their judgments (Dixon & Moore, 1996; Reed & Evans, 1987; Strauss & Stavy, 1982).

One instantiation of a principle representation involves expressing each principle as a fuzzy set of the potential answers. According to this instantiation of the principle hypothesis, upon encountering a problem, a subject represents his/her understanding of the problem by generating the fuzzy set of potential responses specified by each principle.¹ For example, a subject who understands the range principle would generate a fuzzy set like that shown in the upper left panel of Figure 1.

Each principle in Figure 1 is shown with its fuzzy set representation. Each principle specifies a fuzzy set of the potential answers that appear on the horizontal axis. Degree of membership in the set is given by the height of the curve and ranges from 0 (indicating no membership) to 1 (indicating perfect membership). The fuzzy set representations of the principles require some explanation. First, the above-below principle specifies that if the added temperature is warmer (colder) than the initial temperature, the final temperature must also be warmer (colder) than the initial temperature. Consistent with the verbal description of the principle, only answers that are warmer than the initial temperature have membership in the fuzzy set. The height of the curve first increases as

answers move farther from the initial temperature. This captures the idea that answers that are only slightly warmer than the initial temperature are not as good members of the set as answers that are considerably warmer. Similarly, the height of the curves decreases after reaching a peak, reflecting the idea that temperatures that are very hot are not good members of the set defined as "warmer." For example, 100° water would be more appropriately described as "much hotter" than 40° water rather than "warmer." A fuzzy set for "much hotter" would be displaced to the right. This highlights an important aspect of the intuitive representations. The intuitive representations contain information about relations, but this information is *semantic* rather than strictly mathematical or logical. The relation "warmer than" is similar but not identical to the relation "greater than."

Second, consider the equal-temperatures-equal principle (EQT). According to this principle, if the initial and added temperatures are the same, the final temperature must be the same. The membership function for this principle is a very steeply peaked curve that centers on the value of the initial and added temperatures. The shape of the curve reflects the idea that very few values are members of this set to any degree. With younger subjects, we have evidence that the distribution may be less steeply peaked. By college age, however, the semantic notion of "same as" is probably very close to the logical or mathematical notion of equality. Newstead and colleagues (Newstead & Collis, 1987; Newstead & Griggs, 1984) reported similar results in that "all" and "always" were not understood to be equivalent to their logical meanings.

Finally, consider the monotonicity and crossover principles. These two principles are different from the others in that they state the relationship between answers to two problems. The monotonicity principle states that given two problems with equal quantities of added water, the problem with the warmer added temperature will have an answer that is warmer than the answer to the other problem.

Similarly, the crossover principle states that given two problems with equal added temperatures, the problem with the greater added quantity will have an answer closer to the added temperature. Because both these principles require comparing two problems, we have expressed the relationship as a fuzzy set of potential answers to the second problem conditional on the subject's answer to the first problem. The subject's answer to the first problem is labeled on the horizontal axis as R1. The curve defines the membership function for potential answers to the second problem.

According to the integrated representation hypothesis, the principles in Figure 1 are not represented explicitly. Although a subject's actions (e.g., pattern of judgments) may appear to follow principles, the principles have no psychological status (Anderson, 1987). Rather, the appearance that a subject's actions follow principles is simply a consequence of the structure of the integrated representation. By consistently using the integrated representa-

tion, the subject gives the false impression of explicitly understanding principles. When presented with a temperature mixture problem, the subject converts the objective values to subjective scale values. The subjective scale values are then passed to the integration function. The integration function reduces these values to a single result and then passes the result to the output function. The subject does not explicitly consider the relationships that operate in the domain. The integration function produces answers consistent with the principles but does not represent them in a way that is accessible to the subject.

In our first experiment, college students were presented with two temperature mixture problems simultaneously. They were told that both problems had been solved by the same mathematical strategy but were not told what the math strategy was. Subjects were asked to judge whether or not the math strategy was correct on the basis of the answers given. Immediately after choosing "correct" or "incorrect," subjects were asked to rate their confidence in their response and finally to rate how good or bad the math strategy was. The pairs of problems varied orthogonally along two dimensions: (1) the amount the problem pair was wrong (i.e., average distance of the pair of answers from the correct answers) and (2) the amount the problem pair of answers violated one of the principles in Figure 1. Problem pairs violated one of five principles, two of five principles, or did not violate any principles. Note that problem answers may be wrong without violating any of the principles. Trials with correct answers were also presented (see the Appendix for examples of problem pairs).

Past research has shown that college students' intuitive understanding of the temperature mixture task is very sophisticated and that with few exceptions, they appear to understand all the principles described in Figure 1 (Ahl et al., 1992; Dixon & Moore, 1996; Moore et al., 1991). Therefore, according to the principle representation hypothesis, pairs of problem answers that violate principles should be easier for college students to reject as incorrect than those that do not violate principles. For example, suppose a problem states that 2 cups of 40° water is combined with 2 cups of 50° water and the answer given by the math strategy is 34°. This problem violates the above-below principle. If subjects explicitly represent principles, violating the principles should make evaluating the problem easier. This is most easily shown by examining the fuzzy set representation of the above-below principle in Figure 1. Answers with temperatures cooler than T_i , 40°, have zero membership in the fuzzy set and therefore should be very easy to reject.

Additionally, according to the principle representation hypothesis, if a principle is violated, differences in the amount of principle violation should not have an effect. For example, the problem described in the paragraph above violates the above-below principle by 6° because the above-below principle specifies that the answer must be warmer than 40°. If the answer given had been 30° in-

stead of 34°, the amount of principle violation would have been greater (10°). Again, reference to the fuzzy set representation of the above-below principle shows this prediction clearly. Answers that are less than T_i all have zero membership and should, therefore, be equally easy to reject.

The principle representation hypothesis also predicts that the amount by which the problem pair is incorrect (i.e., the average absolute distance from the correct answers) will not have an effect on performance when a principle is violated. When a principle is violated, there should be no effect of amount wrong because answers that violate a principle have zero membership in the fuzzy set regardless of their distance from the correct answers. This prediction stands in contrast to effects observed in research on verifying answers to math problems. A common finding in research on math problems is that the amount wrong, or "split," has relatively large effects on response times (RTs) (e.g., Ashcraft & Battaglia, 1978). This effect is usually attributed either to comparison of the proposed answer to an estimate (Ashcraft & Stazyk, 1981; Restle, 1970) or to an answer retrieved from memory (Ashcraft & Battaglia, 1978; Miller, Perlmutter, & Keating, 1984).

When a principle is not violated, each answer has a nonzero degree of membership in the fuzzy set of potential answers. Therefore, when a principle is not violated, amount wrong will have an effect to the extent that a subject's fuzzy set representations differentiate correct answers from answers that do not violate a principle but that are incorrect. Answers that are farther from the correct answers will have a lower degree of membership in the fuzzy set and will therefore be easier to reject than answers that are closer to the correct answers.²

These predictions of the principle representation hypothesis are similar to the categorization effect observed in work on symbolic distance. For example, Maki (1981) had subjects verify the truth of statements about the relative east-west location of cities. The cities were either in the same state or different states. For both natural and artificial categories (i.e., states), comparisons that did not cross category boundaries showed reliable distance effects. However, when comparisons involved crossing category boundaries, RTs were faster and there was no effect of distance.

The integrated representation hypothesis yields different predictions from those yielded by the principle representation hypothesis. First, with the integrated representation, the distance from the correct answers (i.e., amount wrong) should have a strong effect regardless of whether or not a principle is violated. For college students, the values generated by an integrated representation should be very close to the correct answers (Moore et al., 1991). Therefore, subjects should reject proposed answers more easily the farther they are from the correct answers. Second, because the principles are not explicitly represented and the process involves only a comparison of the answer the subject generates to the proposed

answer, violating a principle should not facilitate evaluating the answers to problems when it is manipulated independently of amount wrong. By the same logic, the amount of principle violation should not have an effect.³

Method

Subjects. Thirty college students participated for extra credit in an introductory psychology course.

Procedure. Subjects were presented with two temperature mixture problems simultaneously. They were told that both problems had been solved by the same mathematical strategy. Subjects were asked to judge whether or not the math strategy was correct or incorrect as quickly as possible, and their RTs were recorded by the experimenter with a hand-held stopwatch. They then rated how confident they were of their judgment. A 9-point confidence scale was used (1 = *absolutely sure*; 9 = *very unsure*). Finally, subjects rated how good or bad the math strategy was. A 9-point scale was used (1 = *very good*; 9 = *very bad*). We chose to examine all four of the following dependent variables: proportion of errors, RT, and ratings of confidence and goodness, because a number of recent studies have shown that different dependent measures can yield very different results (e.g., Haines, Dixon, & Moore, 1996; Mellers, Chang, Birnbaum, & Ordonez, 1992).

Materials. Each temperature mixture problem depicted drawings of three containers of water that were 1.5 in. wide and 2 in. high. The initial container was on the subject's left and the added container was on the subject's right. Two arrows indicated that both the initial and added containers were being poured into a third container located in the center. The quantity of water in each of the containers was represented in the drawing. The center container had the combined quantity of the initial and added containers. A thermometer was drawn next to both the initial and added containers. The thermometer was 2.75 in. in height and was marked in 10° increments from 0° to 80°. The height of a black line on the thermometer indicated the temperature of the water in the initial and added containers. The quantity and temperature of the water in the initial and added containers were also given numerically directly below each container (e.g., 1 cup, 50° temperature). A hypothetical numerical answer was presented to the right of the drawing of each problem. Two temperature mixture problems were presented simultaneously on a single 8 × 11 in. card. One problem was on the top half of the card and the other problem was on the bottom half.

Design. A total of 89 trials were presented. Forty of the trials violated one of the five principles. For these trials, both the amount the problem pair was wrong and the amount of principle violation were varied. Nine trials did not violate any principles, but the amount wrong was varied (no-violation trials). Sixteen trials presented correct answers. Twenty-four trials violated two principles simultaneously. For these trials, the amount wrong was held constant, but the degree to which the trial violated each principle was varied.

The degree to which the trial violated the principle was defined as the absolute amount by which the presented answers were outside the region specified by the principle. For example, the range principle states that the final temperature must be between the initial and added temperatures. A hypothetical answer that does not fall between the initial and added temperatures violates this principle. The amount by which the hypothetical answer falls outside the limits of the initial and added temperatures is the degree to which it violates the range principle. The amount wrong was defined as the average absolute difference between the correct answers for the pair of problems and the presented answers.⁴

For all the problems, the initial container had 2 cups of 40° water. The quantity and temperature of the contents of the added container varied. The contents of the added container came from a 3 (added quantity) × 5 (added temperature) factorial design. The added

quantities were 1, 2, or 3 cups. The added temperatures were 20°, 30°, 40°, 50°, or 60°.

The principles can be classified into two types: within-problem principles, which can be used to evaluate a single problem (range, above-below, and EQT), and compare-problem principles, which require comparing two problems (crossover, monotonicity). For trials that violated within-problem principles, the violation occurred in only one problem in the pair. Whether the problem that violated the within-problem principle was in the top or bottom position for a particular problem was counterbalanced.

The compare-problem principles, monotonicity and crossover, require comparing two problems in order to evaluate the trial. For example, the following problem pair violates monotonicity: (1) 2 cups of 40° water combined with 1 cup of 50° water yields 46° water, and (2) 2 cups of 40° water combined with 1 cup of 60° water yields 42° water. Monotonicity is violated in that the ordering of the hypothetical answers conflicts with the ordering of the added temperatures. Violations of the crossover principle are similar. For example, the following problem pair violates crossover: (1) 2 cups of 40° water combined with 1 cup of 20° water yields 35° water, and (2) 2 cups of 40° water combined with 3 cups of 20° water yields 37° water. Crossover is violated in that the greater quantity of the 20° water should result in a lower rather than a higher combined temperature.

Monotonicity and crossover can be used only to evaluate trials that contain those particular types of problem pairs that express the relationship defined by monotonicity or crossover. To apply the monotonicity principle, both problems must have the same quantity of added water, but the added temperatures must differ (monotonicity trials). To apply crossover to evaluate the trial, the temperature of the added water must be the same for both problems in the pair, but the added quantities must differ (crossover trials).

In order to control for the possible effects of the different types of problem pairs, we manipulated the type of problem pair for within-problem principle violation trials, no-violation trials, and objectively correct trials: (1) problem pairs that had different added temperatures and different added quantities (no-relation problem pairs), (2) problem pairs that had the same added quantities but different added temperatures (monotonicity problem pairs), and (3) problem pairs that had the same added temperatures but different added quantities (crossover problem pairs). For trials that violated EQT, only the first two types of problem pairs were used: (1) no-relation problem pairs and (2) monotonicity problem pairs. For trials that were objectively correct, an additional type of trial was used: problem pairs that contained one problem with 40° added temperature. For the 16 objectively correct trials, there were four examples of each of the four types of problem pairs. Further details of each subdesign are given in Table 1.

Results

For each subdesign, we analyzed the four dependent measures in a multivariate analysis of variance (MANOVA): proportion of errors, RT, ratings of confidence in judgment, and ratings of the goodness of the math strategy. First, we consider the manipulation of type of problem: whether the problem pairs were monotonicity, crossover, or no-relation problem pairs. The manipulation of type of problem pair was to control for a possible confound in comparing the within-problem principles and compare-problem principles. Neither the principle representation nor the integrated representation hypothesis predicts an effect of type of problem pair. A separate MANOVA including all factors (see the Method section for details on individual subdesigns) was con-

Table 1
Levels of Amount of Principle Violation and Amount Wrong for
Subdesigns Violating One and Two Principles

One or No Principle Violated	Amount of Principle Violation	Amount Wrong
Above-below	4.0, 7.0	6.6, 10.5
Range	4.0, 7.0	6.6, 10.5
EQT	4.0, 7.0	3.6, 6.8
Monotonicity	1.4, 4.0	3.6, 6.8
Crossover	4.0, 7.0	4.3, 6.2
No violation	0	3.6, 6.6, 10.5
Two Principles Violated	Amount of Principle (1) Violation	Amount of Principle (2) Violation
AB(1)–EQT(2)	.8, 2.6	5.6, 7.6
Range(1)–EQT(2)	2.2, 5.6	1.0, 4.0
Mono(1)–EQT(2)	7.0, 8.4	10.8, 13.2
Mono(1)–AB(2)	7.0, 9.6	2.6, 5.2
Mono(1)–Range(2)	7.0, 9.6	2.6, 5.2
AB(1)–Crossover(2)	.8, 5.2	6.8, 8.0

Note—The numbers in each cell refer to the levels of the two independent variables: amount of principle violation and amount wrong. The lower panel shows the amount of principle violation for problem pairs that violated two principles. The amount of principle violation for the principle listed first is in the center column. The amount of principle violation for principle listed second is in the right column. EQT, equal-temperatures-equal; AB, above-below; mono, monotonicity.

ducted for each subdesign. The omnibus multivariate main effect test for problem type was not significant for any of the subdesigns except range [$F(8,22) = 1.99$, $F(8,22) = 3.10$, $F(4,26) = 1.61$, and $F(8,22) = 1.13$ for above-below, range, EQT, and no-violation, respectively].⁵ The omnibus effect for the range subdesign appears to be due to monotonicity trials taking longer to evaluate than other types. We include the type of problem pair as a factor in the remaining analyses where applicable.

Effect of amount of principle violation. The effect of amount of principle violation was examined in a separate MANOVA including all factors for each subdesign. For subdesigns that violated only one principle, there were no significant effects of amount of principle violation [$F(4,26) = 1.30, 1.14, .90, 1.84$, and $.27$ for above-below, range, EQT, monotonicity, and crossover, respectively]. Similarly, for the six subdesigns that violated two principles, 5 of 12 omnibus tests had F s of less than 1 and only one was significant. The different amounts of principle violation that we used in this study did not seem to affect subjects' abilities to evaluate the problem pairs. The finding that the amount of principle violation did not have an effect is consistent with both the principle and integrated representation hypotheses, but has different implications depending on whether violating a principle or the amount wrong affects performance.

Effect of violating a principle. The upper panels of Figure 3 show the median RT (right side) and proportion of errors (left side) for problem pairs that violated two principles, violated a single principle, and did not violate any principles. The lower panels show the ratings of confidence (left side) and goodness (right side) for the same problem pairs. All the problem pairs presented in Figure 3 averaged between 6.2° and 7° from the correct answers and may therefore be considered equally wrong.

The data have been averaged over levels of trial type and amount of principle violation.

Response time, proportion of errors, and ratings of confidence and goodness were highly correlated. There is no evidence in Figure 3 for a speed-accuracy tradeoff. First, consider the problem pairs that violated two principles, shown in the six columns on the left of each panel. It appears that violating EQT facilitated accurate evaluation of the problem pairs. Trials that violated EQT—the three columns on the far left of Figure 3 were evaluated more quickly and accurately, with greater confidence, and rated as poorer math strategies than were trials that did not violate EQT [MANOVA omnibus $F(4,26) = 13.58$].

Problem pairs that violated a single principle are shown in the next five columns to the right. Problem pairs that violated within-problem principles (above-below, range, and EQT) were similar on all dependent measures. Problem pairs that violated a compare-problem principle (monotonicity and crossover) were more difficult to evaluate than pairs that violated a within-problem principle [MANOVA omnibus $F(4,26) = 5.56$].

Problem pairs that did not violate any of the principles are shown in the far right column (no-violation trials). These problem pairs did not violate any of the principles but their amount wrong was the same as that in the other problem pairs represented in the figure. All four dependent measures show that these problems were difficult to evaluate accurately. Compared to problem pairs that violated two principles, the no-violation trials were more difficult to evaluate [two principles violated vs. no violation; MANOVA omnibus $F(4,26) = 15.53$].

Similarly, problem pairs that violated one within-problem principle were easier to evaluate accurately than were the no-violation problems [MANOVA omnibus $F(4,26) = 10.44$]. Problem pairs that violated only a

Figure 3. Mean (median for response time) values for the four dependent variables as a function of the temperature mixture principle violated (Experiment 1). Vertical bars represent ± 1 SE. EQT, equal-temperatures-equal; mono, monotonicity.

compare-problem principle—monotonicity or crossover—were similar to no-violation trials in terms of how difficult they were to evaluate. The small differences in the dependent measures seen in the figure are not significant [MANOVA omnibus $F(4,26) = 1.82$].

In summary, problem pair answers that violated two principles and those that violated a single within-problem principle were judged more accurately, more quickly, and with greater confidence, and they were rated as having been generated by worse strategies than problem pair answers that did not violate any principles. The similarity between trials that violated a single compare-problem principle and trials that did not violate any principles suggests that subjects may not have been able to use the information obtainable by comparing the problems or that subjects may not have used those particular principles. Overall, the pattern supports the principle representation hypothesis, especially for within-problem principles. It is difficult to explain these results with the integrated representation hypothesis.

Effect of amount wrong. The principle representation hypothesis predicts that the amount wrong will not affect performance when a principle is violated. When a principle is not violated, amount wrong should have an

effect to the extent that the subject's fuzzy set representations accurately specify the correct answers. The integrated representation hypothesis predicts that the amount wrong will affect performance under all conditions.

The upper panels of Figure 4 show the proportion of errors (left side) and RT (right side) as a function of amount wrong. The lower panels of Figure 4 show ratings of confidence (left side) and goodness (right side) as a function of amount wrong. Each subdesign has been averaged over levels of principle violation and is represented by a separate curve. First, consider the curves for trials that violated above-below, range, and EQT. There was no significant effect of amount wrong for these trials [MANOVA omnibus $F(4,26) = 1.96$, $F(4,26) = 1.73$, and $F(4,26) = 1.30$ for above-below, range, and EQT, respectively]. Similarly, trials that violated crossover or monotonicity also did not show a significant effect of amount wrong [$F(4,26) = 1.34$ and $F(4,26) = 2.74$, respectively]. However, trials that did not violate any of the principles (no-violation trials) did show a significant effect of amount wrong [$F(8,22) = 4.33$].

The results are consistent with the principle representation hypothesis: For trials that violate a principle, amount wrong has little or no effect. Interestingly, no-

violation trials showed an effect of amount wrong especially at the middle and high levels of amount wrong (6.6° and 10.5°). The integrated representation hypothesis cannot explain these results.

Discussion

The results are consistent with the principle representation hypothesis and contrary to the integrated representation hypothesis. There was a systematic effect of violating principles on subjects' evaluations of problems. Problem pairs that violated within-principle principles were evaluated more quickly and accurately than were other problem pairs. Subjects also were more confident of their judgments and rated these math strategies as poorer.

A caveat to the conclusion that the data are consistent with the principle representation hypothesis is that problems that violate a compare-problem principle did not show the same effect as problems that violated a within-principle principle. Subjects may have found it too difficult to coordinate comparison of the ordering of the two problems and the ordering of the two answers, and therefore could not apply the appropriate principle to the problem. Another possibility is that subjects simply did not represent the compare-problem principles. Although either of these explanations is plausible, note that the integrated representation hypothesis cannot plausibly explain the results. The integrated representation hypothe-

sis is disconfirmed by the results of the within-principle principles.

The results of the amount-wrong manipulation are also consistent with the principle representation hypothesis. As predicted, amount wrong had little or no effect for trials that violated a principle. Also as predicted, amount wrong had an effect for trials that did not violate any of the principles. The finding that the effect for no-violation trials was limited to the middle and highest levels of amount wrong is also consistent with the principle representation hypothesis. According to the principle representation hypothesis, the effect of amount wrong for no-violation trials depends on how precisely the fuzzy sets specify the correct answers. Subjects had to be sensitive to different degrees of membership in the fuzzy sets in order to show an effect of amount wrong for these trials. On this interpretation, it appears that subjects were sensitive to different degrees of membership but only at fairly high levels of amount wrong.

Our results are analogous to the categorization effect Maki (1981) observed with symbolic distance. When a categorical boundary is crossed (i.e., a principle is violated), performance is facilitated and the distance (i.e., amount wrong) within the boundary has no further effect. When a categorical boundary is not crossed (no principle is violated), there is an effect of distance. Overall, the first experiment provided support for the principle representation hypothesis.

Figure 4. Mean (median for response time) values for the four dependent variables as a function of manipulated levels of amount wrong (Experiment 1). There is a separate solid curve for each principle violated. The dashed curves represent trials in which no principle was violated.

EXPERIMENT 2

The second experiment had two main goals. First, we wanted to provide additional tests of the principle representation hypothesis using a slightly different domain. A task analogous to temperature mixture that college-age subjects do not understand well is mixture of acid concentrations (Reed, 1988). In the acid mixture task, the concentration of acid is analogous to the temperature variable in temperature mixture. Quantity of liquid operates the same way in both tasks. The same set of principles that govern temperature mixture also govern mixture of acid concentrations. In the second experiment, we attempted to manipulate subjects' understanding of the principles of acid mixture by giving them a brief explanation of one of the principles. Subsequently, we evaluated their intuitive understanding of the principles using an estimation task. We then gave them the same task as that used in the first experiment—evaluating pairs of problems that were both said to have been solved by the same math strategy. In addition to college students, we also recruited eighth graders for the second experiment. We chose eighth graders because it seemed very likely that they would understand acid mixture much more poorly than would the college students. By using both groups, we could examine the effects of instructions on subjects with different baseline understandings of the domain. The second experiment, therefore, allowed us to extend the evidence on the principle representation hypothesis by examining the relationship between instructions on principles, measures of understanding principles, and the evaluation of problem pairs that violate principles.

The principle representation hypothesis predicts that instruction on principles will affect subjects' principle representation, which will in turn affect their ability to evaluate problem pairs. Therefore, understanding of principles should mediate the effect of the instructions on performing the evaluation task. Experiment 2 tested the mediation prediction. The second purpose of Experiment 2 was to provide an additional test of the process by which the intuitive representation influences the use of math strategies. In past work (Dixon & Moore, 1996), we found that understanding a principle appeared to be a necessary but not sufficient condition for selecting a math strategy consistent with that principle. That is, subjects may not use all the principles that they understand. Our previous results suggested that subjects select particular math strategies because they are consistent with some subset of the principles contained in their intuitive representation. We call this subset of principles the "goal model." According to this hypothesis, subjects select a math strategy on the basis of the match between the principles they include in their goal model and the properties of the math strategy. For example, suppose a subject who understands the monotonicity and range principles is trying to select a math strategy to solve the temperature mixture problem. The subject searches through his/her repertoire of math strategies for a match between what

the math strategy does and the principles in his/her goal model. The math strategy the subject selects will depend on which of the principles he/she includes in the goal model. If only monotonicity is included, the subject may select addition of the temperatures as a math strategy because addition has the property of producing answers that are monotonically increasing. If the subject also includes range in the goal model, he/she may select averaging, because averaging has the property of producing answers that are between the two combined temperatures as well as monotonically increasing.

In explaining this pattern of results, the goal model hypothesis makes two related assumptions. The first assumption is that the subjects have an understanding of the functional properties of the basic math operations. For example, subjects must understand that adding two variables produces answers larger than the values of the two variables. Averaging produces answers that are between the values of the two variables. The second assumption is that some subjects include only a subset of the principles they understand about the domain in their goal model. Both developmental and adult research on problem solving suggest that this is a reasonable assumption. Research on the development of problem solving has shown that an important developmental change is that older children know which aspects of the problem are important (e.g., Siegler, 1976). Similarly, an important difference between expert and novice performance is that experts know which aspects of a problem are important whereas novices do not (Chi, Glaser, & Rees, 1982; Larkin, McDermott, Simon, & Simon, 1980; Soloway, Adelson, & Ehrlich, 1988).

Our second experiment eliminated the need for the first assumption and provided additional tests of the second assumption. The need for the first assumption—that subjects understand the functional properties of math operations—was eliminated in that the functional properties of the math operations are given by the relationship between the answers and the problems. If a subject is evaluating a math strategy on the basis of the answers given to two problems, the information about what the hypothetical strategy does is specified by the relationships between the problems and the answers. For example, suppose a subject is given the problem "2 cups of 40% acid combined with 1 cup of 60% acid," and the answer produced by the math strategy is 38% acid. The information about what the strategy is doing is available from what is presented. The subject does not need prior knowledge of how the math strategy works. The second experiment also tested the second assumption—that some subjects include only a subset of the principles they understand in their goal model. These subjects should fail to reject answers that violate principles they understand but that are not included in their goal model. Additionally, the goal model hypothesis predicts that subjects must understand the principle in order to include it in their goal model. We hypothesized, therefore, that subjects who do not understand a principle should not reject problem pairs that violate it.

Method

Subjects. Sixty college students participated for extra credit. Twenty-three eighth graders from a public school and a parochial school participated as volunteers. Parental consent was obtained for the eighth graders.

Materials. Subjects completed two tasks—an estimation task about mixing acid concentrations and evaluation of pairs of problems that had been solved by the same math strategy. For the estimation task, subjects were asked to predict the acid concentration in a container given the initial acid concentration and quantity of the liquid in the container and the acid concentration and quantity of the added liquid. Two schematic containers were used as stimuli, one for the initial liquid and one for the added liquid. The schematic containers were felt board (12×16 in.). The quantity of the liquid was represented by felt strips of three different sizes. Each container was paired with a schematic meter that indicated the acidity of the liquid in the container. The acidity meter was 16 in. in height with a movable marker. The placement of the marker on the meter indicated the acidity of the liquid in the container. No numbers or graduations were present on either the containers or the acidity meters.

The materials for the evaluation task were very similar to those used in the first experiment, with the exception that the trials were presented on a computer screen rather than on cards. The drawings represented acid mixture rather than temperature mixture.

Design. College subjects were randomly assigned to one of six instruction conditions. Five of the instruction conditions involved a short explanation of one of the principles of acid mixture: above-below, range, equal-concentrations-equal (EQC), monotonicity, or crossover. Subjects assigned to the sixth condition were not instructed on any principles. Eighth-grade students were randomly assigned to either the range or crossover instruction conditions. We chose to instruct eighth graders on range and crossover because the number of available students was limited and these principles are acquired last in the development of understanding temperature mixture, suggesting that they would not be known to eighth graders in this domain.

The estimation task was a 3 (added quantity) $\times 5$ (added acid concentration) factorial design. The added quantities were analogous to 1, 2, and 3 cups, but were not labeled numerically. The added acid concentrations were analogous to 20%, 30%, 40%, 50%, and 60% acid, but were not labeled numerically. The added acid concentrations were verbally labeled “very slightly acidic,” “slightly acidic,” “fairly acidic,” “extremely acidic,” and “very extremely acidic.” The added quantities were verbally labeled as “small amount,” “medium amount,” and “large amount.” The initial container always held amounts analogous to 2 cups of liquid with 40% acidity and was described as a “medium” amount of “fairly acidic” liquid. Trials for the estimation task were presented in one of five random orders. The design for the evaluation task was the same as that in the first study. The eighth graders did not evaluate problem pairs that violated two principles because these trials were evaluated similarly to within-principle trials in Experiment 1, and our time with each eighth-grade student was limited. The college students completed the full design.

Procedure. Subjects participated individually. Subjects were told that the experiment consisted of two tasks and that both tasks were about mixing liquids with different concentrations of acid. Before starting the two tasks, subjects received instruction about a principle of mixing acid concentrations as determined by their random assignment. The instructions for each principle are below.

Above-below. When one liquid is always the same amount and has the same concentration, and another liquid with some concentration is combined with it, the resulting mixture will always be in the direction of the added liquid. For example, if one container always has a medium amount of fairly acidic liquid and some amount of extremely acidic liquid is combined with it, the resulting mixture will have an acid concentration that is more acidic than the original fairly acidic liquid. So when one liquid is always the same, com-

binning another liquid with it makes the resulting acid concentration move in the direction of the added liquid’s acid concentration.

Range. When two acid concentrations are combined, the resulting concentration is always between the two original concentrations. For example, if the liquid in one container is very acidic and the other is slightly acidic, when the liquids are combined the resulting mixture will have an acid concentration somewhere between very acidic and slightly acidic. So when acid concentrations are combined, the resulting concentration is always between the concentrations that were combined.

Equal-concentrations-equal. If two liquids have the same acid concentration, when they are combined the resulting mixture will have that same acid concentration. For example, if a container of fairly acidic liquid is combined with another container of fairly acidic liquid, the resulting mixture will also be fairly acidic. So when one container of liquid is combined with another container of liquid having an identical acid concentration, the resulting concentration will be the same.

Monotonicity. If the liquid in one container always has the same acid concentration, the greater the acid concentration of a particular amount of liquid combined with it, the greater the acid concentration of the resulting mixture. For example, if one container always has fairly acidic liquid, when extremely acidic liquid is combined with it, the resulting mixture will be more acidic than if the same amount of slightly acidic liquid were combined with it. So when one container is always the same, the greater the concentration of the liquid added, the greater the concentration of the resulting mixture.

Crossover. The more there is of an acid concentration, the greater its effect on the resulting mixture. For example, if a large amount of very acidic liquid were combined with some other liquid, it would have a greater effect than if there were only a small amount of the very acidic liquid. So the more liquid there is of a particular acid concentration, the greater its effect.

After receiving instruction on the appropriate principle, subjects completed the estimation task. The estimation task was intended to assess intuitive understanding of the domain of acid mixture. The experimenter explained that the felt boards represented containers of liquid and that the felt represented the liquid. The subject was told that a meter went with each container and that the meter indicated how acidic the liquid in each container was. The experimenter explained how the acid meter worked and gave examples of extreme acidities. The experimenter explained that the container on the subject’s right would always start out with a medium amount of fairly acidic liquid. The other container (the added container) would have a different amount and acidity each time. Subjects were asked to judge what the acidity of all the liquid would be when the added liquid was combined with the initial liquid. Subjects responded by adjusting the meter of the initial container to show the acidity of all the liquid combined.

After completing the acid mixture estimation task, subjects were read the instructions for the evaluation task. The instructions for the evaluation task were analogous to those in the first study. Subjects judged whether the math strategy that produced the answers to two acid mixture problems was correct or not. The amount of time it took subjects to make the judgment was recorded by the computer. Subjects were then asked to rate their confidence in their judgments. Finally, subjects rated how good the math strategy was that produced the answers. The rating scales were the same as those in Experiment 1.

Scoring of intuitive principles from estimation task. Each subject’s pattern of judgments in the estimation task was scored for consistency with each of the principles. To the degree that a subject understands a principle, his/her pattern of judgments should be consistent with that principle.

Monotonicity measures understanding of the principle that the more acidic the added liquid, the more acidic the final mixture. Monotonicity was scored by comparing the ordering of the judgments

of adjacent added acid concentrations for each quantity. For example, when 1 cup of liquid with 30% acidity is combined with the liquid in the initial container (2 cups of 40%), the resulting acidity should be more acidic than when 1 cup of liquid with 20% acidity is combined with it. One point was given for each correct ordering of answers for each quantity. One point was subtracted for each incorrect ordering, and nothing was done for ties. The maximum score was 12 and the minimum was -12.

Above-below measures understanding of the principle that the combined acidity should always be in the direction of the added acidity from the initial acidity. For example, if the added liquid were less acidic than the initial liquid, the final acidity should be less than the initial acidity. Six of the added liquid trials were below the initial acidity of 40% (20% and 30%), and six were above (50% and 60%). One point was added for each answer on the appropriate side of the initial acidity. One point was subtracted for each answer on the inappropriate side. Nothing was done for ties (i.e., judged final acidity equal to the initial acidity). Forty-percent added acidity trials were not used. The maximum score was 12 and the minimum was -12.

Range measures understanding of the principle that the acidity of the mixture must fall between the initial and added acidities. Range was scored by adding 1 point for every answer between the added and initial acidities. One point was subtracted for each answer not between the added and initial acidities. Nothing was done for ties (i.e., judged final acidity equal to the initial or added acidities). Forty-percent added acidity trials were not used. The maximum score was 12 and the minimum was -12.

Crossover measures the understanding of the interaction between acid concentration and quantity. Crossover was scored by comparing the ordering of the extreme quantities for each added acid concentration. For example, the final acid concentration should be less when 3 cups of liquid with 20% acid are added to the initial acid concentration than when 1 cup of liquid with 20% is added. One point was added for each correct ordering. One point was subtracted for each incorrect ordering. Nothing was done for ties (i.e., final acid concentration judged to be equal for the compared trials). Forty-percent added acid concentration trials were not used. The maximum score was 4 and the minimum was -4.

Equal-concentrations-equal measures understanding of the principle that the acid concentration will not change if the added liquid has the same concentration as the initial concentration. EQC was scored using the 40% added concentration trials. One point was added for each judged final acid concentration within 1% of 40%. The maximum score was 3 and the minimum was 0.

For ease of comparison, all principles scores were linearly transformed to percentage of the maximum possible score.

Results

First, we consider the effect of the instructions on subjects' intuitive understanding of the principles. Then we examine the relationship between the instruction conditions, understanding the principles, and evaluating problem pairs that violate a principle. Finally, we examine at an individual level how understanding a principle is related to evaluating problem pairs that violate that principle.

Effect of instruction on intuitive understanding. In past work we have used the principle scores from the estimation task as an index of subjects' intuitive understanding of the domain (Ahl et al., 1992; Dixon & Moore, 1996; Moore et al., 1991). Recall that numbers are not available, and so math calculations are extremely unlikely. According to Hammond (Hammond, Hamm, Grassia, & Pearson, 1987), this type of task should induce the use of intuitive cognition. Likewise, Reed (1988) proposed that

similar estimation tasks provide a good index of what people understand about a domain.

Figure 5 shows the mean percentage principle scores for each instruction group for the college students. The most striking result in Figure 5 is that instruction on the range and EQC principles seems to have improved understanding of most of the principles compared with the other instruction conditions. The other instruction conditions do not appear to have differed from each other. To test the hypothesis that the no-instruction, crossover, above-below, and monotonicity instruction conditions did not differ, a one-way analysis of variance (ANOVA) was performed separately on each principle score. The effect of instruction group was not significant for any of the principle scores (all $F_s \leq 1.00$). The similarity between the understanding of the no-instruction group and the crossover, above-below, and monotonicity instruction conditions suggests that the instruction we provided for these principles did not have an effect on subjects' understanding of the task.

For the college students, the range and EQC instruction conditions yielded higher mean principle scores than did the other conditions for all principles except crossover. To compare the range and EQC instruction conditions with the other instruction conditions, we performed a one-way ANOVA separately on each principle score. For this analysis we considered the range and EQC conditions as one group and the other conditions as a second group. The effect of group (range/EQC vs. the other conditions) was significant for all principle scores except crossover [monotonicity, $F(1,48) = 16.46$; above-below, $F(1,39) = 10.22$; range, $F(1,47) = 20.82$; EQC,

Figure 5. Mean percentage principle scores for the college students in the estimation task with a separate curve for each of the instruction groups (Experiment 2). EQC, equal-concentrations-equal; mono, monotonicity.

$F(1,56) = 16.94$; and crossover, $F(1,45) = 2.96$, using Welch's correction for unequal variances]. Instruction on the range and EQC principles appears to have enhanced overall understanding of the task, rather than independently enhancing understanding of the particular principle. One explanation of this result is that instruction on range and EQC helped subjects conceptualize the domain as involving intensive as opposed to extensive quantification.

Figure 6 shows the intuitive principle scores of the two eighth-grade instruction conditions, range and crossover, and the college instruction conditions grouped into range/EQC and other groups. Comparison of the two eighth-grade instruction groups revealed that the range instruction group showed better understanding of all principles except the crossover principle [crossover, $F < 1.00$; monotonicity, $F(1,21) = 8.12$; above-below, $F(1,21) = 7.98$; range, $F(1,21) = 10.33$; EQC, $F(1,21) = 14.68$; $ps < .05$]. The crossover principle was poorly understood by both instruction groups. The eighth-grade subjects who were instructed on range showed understanding very similar to that of the college subjects instructed on range or EQC. For all principles except crossover, the eighth-grade range instruction group was similar to the college range/EQC instruction group. This result is striking in that a single paragraph of instruction was sufficient to bring the performance of these eighth graders up to the level of college-age subjects. Reed and Evans (1987) showed a similar effect with a transfer paradigm. Comparison of the crossover (eighth-grade) and other instruction (college) groups suggests that there was an age difference between the eighth-grade and college students in terms of their initial intuitive understanding of the task. The college-age subjects understood more about

Figure 6. Mean percentage principle score in the estimation task with a separate curve for each age and instruction group (Experiment 2). The vertical bars represent ± 1 SE. EQC, equal-concentrations-equal; mono, monotonicity.

Figure 7. Two causal models representing the relations between the intuitive principle score in the estimation task and the proportion of error in the verification task.

the task initially than did the eighth-grade subjects, as would be expected.

Effect of instruction conditions on evaluating problems. The principle representation hypothesis predicts that the two instruction groups with better understanding of the principles should evaluate problems that violate the principles more accurately than the other instruction groups. Subjects with poor understanding of the principles either apply incorrect principles to the problem or have only partial understanding of the principles. In both cases, these subjects as a group will represent their understanding of the problems in such a way that potential answers that violate the principles have nonzero membership (i.e., are considered viable answers). We tested these predictions for each subdesign in which a principle was violated. We compared the proportion of errors for the range and EQC instruction conditions versus the other instruction conditions.⁶ The effect of instruction group was significant for all subdesigns that violated a within-problem principle [$F(1,79) = 7.28, 3.98$, and 5.87 for above-below, range, and EQC, respectively]. There was no evidence of an effect for subdesigns that violated a compare-problem principle ($F_s < 1$).

Mediating effect of understanding principles. A more complete test of the principle representation hypothesis involves examining the relationships among the instruction conditions, the principle scores, and performance in the evaluation task. According to the principle representation hypothesis, the instruction conditions affect understanding of the principles, which in turn affects performance in the evaluation task. That is, understanding of the principles mediates the effect of instruction condition on the evaluation task. A competing hypothesis is that instruction condition affects the principle score and also directly affects performance on the evaluation task, but that the relationship between principle scores and the evaluation task is spurious. These two models are presented in Figure 7.

The two models make different predictions about the correlations between instruction condition, principle score, and the proportion of error in the evaluation task (Asher, 1983). According to the spurious model, the correlation between principle score and proportion of error should go to zero when the effect of instruction condition is partialled out. Conversely, according to the mediation model, the correlation between instruction condition and proportion of error should go to zero when the effect of principle score is partialled out. The mediation hypothesis also predicts that the correlation between principle score and proportion of error will become smaller when the effect of instruction condition is partialled out.

We tested both models for each principle by examining the correlations between the principle score, the proportion of errors for trials that violated that principle in the evaluation task (averaged over type of problem, amount wrong, and amount of principle violation), and instruction condition. The results for the within-problem principles are presented in Table 2.⁷ The same pattern of results was seen for all the within-problem principles. As predicted by the mediation model, instruction condition was not significantly correlated with proportion of errors when the effect of principle score was partialled out. However, principle score still significantly correlated with proportion of errors when the effect of instruction condition is partialled out. These results support the mediation model. Understanding the principle, as measured by the principle score, appears to have mediated the effect of the instruction conditions on evaluating problems.

Another way of testing the mediating role of principle understanding is to determine whether a global measure of accuracy on the estimation task is as predictive as principle understanding. We calculated the absolute distance from the correct answers on the estimation task. Because estimates that are inconsistent with principles must be less accurate than estimates that are consistent with principles, this accuracy measure necessarily subsumes the variance associated with each principle score. Therefore, the principle representation hypothesis predicts that the accuracy score will not explain any variance in evaluating problems beyond that explained by

the principle score. Because the principle score measures the principle representation that is used in evaluating problems, the additional information contained in the accuracy score will not predict how subjects evaluate problems.

To test this hypothesis, the accuracy score was entered after each principle score in separate multiple regressions. The dependent variable was the mean proportion of error for trials that violated that principle. Grade and instruction condition were entered first in each regression. The results showed that the accuracy score did not significantly increase the multiple R^2 . All changes in R^2 were less than or equal to .02 [$F_s(4,78) = 2.12, .26, 1.40, .74$, and 1.71 for above-below, range, EQC, crossover, and monotonicity, respectively]. Consistent with the principle representation hypothesis, our measure of principle understanding predicts how well subjects will evaluate problems that violate that principle. The overall accuracy measure does not add any predictive value despite being based on the entire pattern of estimation judgments.

Relationship between understanding principles and evaluating problems. Recall that according to the goal model hypothesis, subjects evaluate math strategies by confirming that a strategy is consistent with some set of the principles that constitute their intuitive representation. The goal model hypothesis predicts that subjects who do not understand a principle cannot make that principle part of their goal model and cannot use it to evaluate math strategies. Therefore, subjects who do not understand a principle should have difficulty reliably rejecting math strategies that violate that principle. Subjects who do understand a principle may or may not include it in their goal model while evaluating math strategies. For example, some subjects who understand the above-below principle should confirm math strategies by checking to see whether the strategy produces answers that are consistent with above-below. Other subjects who understand the above-below principle may not appreciate that it is an aspect of the task that should be embodied in the math strategy. These subjects would not include above-below in their goal model. Therefore they would not reliably reject math strategies that violate the above-below principle. Understanding a principle is necessary but not sufficient for using the principle in the goal model.

Table 3 shows the number of subjects who reliably rejected violations of each principle as a function of whether or not they understood the principle. *Understanding a principle* was defined as a principle score $\geq 75\%$ of the maximum score for all principles (except EQC, which had the cutoff for understanding set at 66%, because there is no value possible between 66% and 100%). We defined *reliably rejecting violations of a principle* as rejecting a sufficient proportion of trials so that the probability of coming from the binomial distribution where $p = .5$ is less than .05. Because there were not sufficient trials for the monotonicity and crossover principles to reach the .05 level, we defined reliable rejection for trials that violated these principles as rejecting all trials ($p = .0625$).

Table 2
Correlations for Testing the Mediation and Spurious Models

	Error-Instruction	Error-Principle	EI.P	E.P.I
Predictions				
Spurious	*	*	*	n.s.
Mediation	*	*	n.s.	*
Results				
Above-below	.25*	.37*	.15	.28*
EQC	.26*	.38*	.12	.28*
Range	.19	.42*	.01	.37*

Note—The first two columns in the table show the correlations between error and instruction condition and principle score, respectively. EI.P, correlation between error and instruction with principle partialled out; E.P.I, correlation between error and principle score with instruction condition partialled out; EQC, equal-concentrations-equal. *Significant at $p < .05$.

Table 3
Number of Subjects Who Reliably Rejected Violations of Each Principle, as a Function of Principle Understanding

Principle Understanding	Above-Below		Range		Equal-Concentrations-Equal		Monotonicity		Crossover	
	Reject Trials	Do Not Reject Trials	Reject Trials	Do Not Reject Trials	Reject Trials	Do Not Reject Trials	Reject Trials	Do Not Reject Trials	Reject Trials	Do Not Reject Trials
Understand	44	29	38	22	45	15	32	40	13	27
Do not understand	1	9	5	18	7	16	3	8	13	30

As can be seen in Table 3, for the three within-problem principles (range, above-below, and EQC), few subjects who did not understand a principle reliably rejected trials that violated that principle. However, a large proportion of subjects who *did* understand a principle failed to reliably reject trials that violated that principle. The distributions for the within-problem principles were all significantly different from what might be expected under a hypothesis of independence when tested using the Fisher-Irwin exact test (Siegel & Castellan, 1988).

For the compare-problem principles, the results were mixed. A similar pattern of results can be seen for trials that violated the monotonicity principle, although the distribution was not significantly different from independence. For the crossover principle, the results did not follow the same pattern. Understanding the crossover principle clearly does not predict anything about rejecting trials that violate crossover.

With the exception of the crossover principle, the observed relationship between understanding a principle and rejecting violations of the principle is consistent with the goal model hypothesis. Subjects who do not understand a principle are much less likely to reject trials that violate that principle. However, subjects who do understand a principle may or may not reject trials violating that principle.

GENERAL DISCUSSION

The goal of the first experiment was to test two hypotheses concerning the nature of the representation used in evaluating problems: the principle representation and the integrated representation hypotheses. The results were consistent with the principle representation hypothesis. Problem pair answers that violated within-problem principles were evaluated more accurately and more quickly than were either problem pairs that did not violate any principles or problem pairs that violated a compare-problem principle. Problem pairs that violated a within-problem principle were also evaluated with greater confidence and were rated as having been generated by worse strategies. The amount-wrong manipulation had little or no effect for problem pairs that violated a principle. Interestingly, the amount-wrong manipulation had greater and very consistent effects for problem pairs that did not violate any of the principles. Additionally, the manipulation of amount of principle violation did not appear to have an effect on performance.

The principle representation hypothesis is consistent with this pattern of results. Because principles are represented explicitly as fuzzy sets, violating principles facilitates the evaluation of the problem pairs. The distance from the correct answer has little effect when a principle is violated because all answers that violate a principle have zero membership in the fuzzy set. If the problem pair does not violate any of the principles, subjects can base their responses on the nonzero degree of membership in the fuzzy sets. To the extent that the membership functions of the fuzzy sets distinguish correct and incorrect answers, the amount-wrong manipulation should have an effect. The category effect in work on symbolic distance is analogous to this pattern of results (Maki, 1981).

The first experiment showed that subjects represented the domain using principles. The distinction between the principle representation and the integrated representation is important in delineating how intuitive understanding is used in selecting and evaluating formal strategies. The type of representation constrains the set of candidate processes for selecting and evaluating strategies. If the representation contains explicitly known principles, the principles may be used independently of each other. If the representation is integrated, the different types of relationships between variables cannot be separated from one another. Therefore, processes that use the relationships independently are possible only if the representation is in terms of principles.

The principle representation we propose uses fuzzy sets to represent subjects' semantic knowledge about the relationships in the domain. This semantic knowledge specifies relationships that are similar but not identical to logical or mathematical relations. For example, we propose that in understanding temperature, subjects use concepts such as "warmer than" and "colder than." These concepts are different from the mathematical concepts of "greater than" and "less than" in that they have additional implications. "Warmer than" implies not only that the compared temperature is greater, but also that it is not extremely greater. If 120° water were described as "warmer than" 34° water, the listener would most likely note the discord in the statement despite its logical truth. Similar observations, with supporting evidence, have been offered regarding the meanings of probability terms (Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986) and linguistic quantifiers (Zadeh, 1983; Zimmer, 1988).

In the second experiment, we extended the evidence for the principle representation hypothesis. According to the principle representation hypothesis, the principle scores should mediate the effect of instructions on the evaluation task. The pattern of correlations among the three measures supports this prediction. This finding has a number of interesting aspects. First, it appears that a short paragraph of instruction is sufficient to enhance the understanding of the principles. It seems likely that the instructions may help subjects make an analogy to better understood domains, rather than directly constructing understanding. Reed and Evans (1987) reported good transfer between very similar tasks. It is not entirely clear why subjects instructed on range and EQC made so much greater gains than the subjects in other conditions. Possibly these principles clearly identify the problem as involving intensive quantification rather than extensive quantification.

Second, and more directly in support of the principle representation hypothesis, the measures of understanding for each principle mediate the effect of instructions on evaluating problems violating that principle. By scoring the estimation pattern for consistency with each principle, we arrive at measures for understanding each principle. The instructions affect the degree to which subjects understand principles. Understanding of the principles affects how accurately subjects evaluate the answers to problems. Importantly for the principle representation hypothesis, using the overall accuracy score for the pattern of estimation judgments does not explain any variance in the evaluation task beyond that explained by each principle. This result strongly suggests that subjects use the principles we are measuring in the estimation task to evaluate the problem pairs. The crucial predictor for evaluating problems that violate a principle is the degree to which the estimation task judgment pattern adheres to the *principle*. Including further information about the accuracy of the judgment pattern does not add predictive value. A likely explanation of this result is that subjects make judgments (in the estimation task) and evaluate problems using the same principle representations. Fuzzy set representations can be used to model judgment processes (see, e.g., Massaro, 1994).

The second experiment also tested a hypothesis about the mapping process between the principles and selection of math strategies—the goal model hypothesis. The goal model hypothesis states that subjects use their semantic or intuitive knowledge of the domain to select (and evaluate) math strategies. Subjects can include only principles that they understand in their goal model, but will not necessarily include all understood principles. The relationship we observe between understanding a principle and evaluating a problem pair that violates that principle is consistent with this hypothesis. Subjects who understand a principle may or may not include it in their goal model. Therefore, subjects who understand a principle may or may not reliably reject problem pairs vio-

lating that principle. However, subjects who do not understand a principle should not reliably reject problem pairs violating that principle. For all the principles except crossover, we observed that subjects were much less likely to reject trials that violated a principle if they did not understand that principle. This pattern of results suggests that subjects can use their semantic knowledge about the relationships in the domain to evaluate the mathematical or logical relationships specified by the problem pairs. Siegler (Siegler & Crowley, 1994; Siegler & Jenkins, 1989) has presented evidence suggesting that young children construct similar goal structures when seeking more efficient solution strategies.

In past work, we observed the same relationship between understanding a principle and selecting (as opposed to evaluating) a math strategy (Dixon & Moore, 1996). Some students who understood a principle selected math strategies that were consistent with the principle. Other students who understood a principle selected math strategies that violated that principle. However, students did not select math strategies that were consistent with principles they did not understand. The fact that we observe the same relationship between understanding principles and math strategies for both selection and evaluation suggests that subjects may be constructing similar goal model structures in both tasks. The results of the present study are a more compelling test of the goal model hypothesis because the information about violating the principle is given directly in the problem. We do not have to assume that subjects have any prior knowledge about whether a math strategy will violate the principle or not.

Many lines of past research cited in the introduction have either suggested or assumed that a person's intuitive or qualitative understanding of a domain is an important factor in formal problem solving. We have proposed that a person's intuitive representation of the problem domain and the representation of the formal strategies must be compared in order to either select a strategy or evaluate a strategy. It was difficult, however, to define the comparison process without testing how the intuitive knowledge of the domain is represented. The two experiments we have presented here and our previous work show that the intuitive representation functions in terms of principles. However, a major component of the model remains to be investigated—how the formal strategies are represented. It is possible that the representation of formal strategies is in terms of principles as well. The principles for the formal strategies would also specify what the relations between variables are. Consistent with this idea, Krueger (1986; Krueger & Hallford, 1984), and more recently, Lemaire and Fayol (1995), have shown that subjects are aware of an odd-even rule in addition and multiplication. This rule specifies a relation between the result and the operand. If subjects understand this relatively tangential property of math strategies, it seems likely that they might understand more central princi-

ples. A representation of principles of formal strategies would facilitate a mapping process between understood principles and the formal strategies.

REFERENCES

- AHL, V., MOORE, C. F., & DIXON, J. A. (1992). Development of intuitive and numerical proportional reasoning. *Cognitive Development*, **7**, 81-108.
- ANDERSON, N. H. (1974). Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 1-101). New York: Academic Press.
- ANDERSON, N. H. (1987). Function knowledge: Comment on Reed and Evans. *Journal of Experimental Psychology: General*, **116**, 297-299.
- ANDERSON, N. H. (1991). Functional memory in person cognition. In N. H. Anderson (Ed.), *Contributions to information integration theory* (Vol. 1, pp. 1-55). Hillsdale, NJ: Erlbaum.
- ASHCRAFT, M. H., & BATTAGLIA, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 527-538.
- ASHCRAFT, M. H., & STAZYK, E. H. (1981). Mental addition: A test of three verification models. *Memory & Cognition*, **9**, 185-196.
- ASHER, H. B. (1983). *Causal modeling*. Beverly Hills, CA: Sage.
- BRIARS, D. J., & LARKIN, J. H. (1984). An integrated model of skill in solving elementary word problems. *Cognition & Instruction*, **3**, 245-296.
- BRIARS, D. J., & SIEGLER, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology*, **20**, 607-618.
- 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, pp. 7-75). Hillsdale, NJ: Erlbaum.
- DE KLEER, J. (1975). *Qualitative and quantitative knowledge in classical mechanics*. Unpublished master's thesis, MIT.
- DE KLEER, J. (1977). Multiple representations of knowledge in a mechanics problem solver. *International Joint Conference on Artificial Intelligence*, **5**, 299-304.
- DIXON, J. A., & MOORE, C. F. (1996). The developmental role of intuitive principles in choosing mathematical strategies. *Developmental Psychology*, **32**, 241-253.
- GAINES, B. R. (1977). Foundations of fuzzy reasoning. In M. M. Gupta, G. N. Saridis, & B. R. Gaines (Eds.), *Fuzzy automata and decision processes* (pp. 19-75). New York: North-Holland.
- GREENO, J. G., RILEY, M. S., & GELMAN, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology*, **16**, 94-143.
- HAINES, B. A., DIXON, J. A., & MOORE, C. F. (1996). *Dimensions of task structure influence the development of probability understanding*. Manuscript submitted for publication.
- HAMMOND, K. R., HAMM, R. M., GRASSIA, J., & PEARSON, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions*, **SMC-17**, 753-770.
- HARDIMAN, P. T., DUFRESNE, R., & MESTRE, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition*, **17**, 627-638.
- KINTSCH, W., & GREENO, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, **92**, 109-129.
- KRUEGER, L. E. (1986). Why $2 \times 2 = 5$ looks so wrong: On the odd-even rule in product verification. *Memory & Cognition*, **14**, 141-149.
- KRUEGER, L. E., & HALLFORD, E. W. (1984). Why $2 + 2 = 5$ looks so wrong: On the odd-even rule in sum verification. *Memory & Cognition*, **12**, 171-180.
- LARKIN, J. (1983). The role of problem representation in physics. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 75-98). Hillsdale, NJ: Erlbaum.
- LARKIN, J., McDERMOTT, J., SIMON, D. P., & SIMON, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, **208**, 1335-1342.
- LEMAIRE, P., & FAYOL, M. (1995). When plausibility judgments supersede fact retrieval: The example of the odd-even effect on product verification. *Memory & Cognition*, **23**, 34-48.
- MAKI, R. H. (1981). Categorization and distance effects with spatial linear orders. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 15-32.
- MASSARO, D. W. (1994). A pattern recognition account of decision making. *Memory & Cognition*, **22**, 616-627.
- MELLERS, B. H., CHANG, S., BIRNBAUM, M. H., & ORDONEZ, L. D. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 347-361.
- MILLER, K., PERLMUTTER, M., & KEATING, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 46-60.
- MOORE, C. F., DIXON, J. A., & HAINES, B. A. (1991). Components of understanding in proportional reasoning: A fuzzy set representation of developmental progression. *Child Development*, **62**, 441-459.
- NEWSTEAD, S. E., & COLLIS, J. (1987). Context and interpretation of quantifiers of frequency. *Ergonomics*, **30**, 1447-1462.
- NEWSTEAD, S. E., & GRIGGS, R. A. (1984). Fuzzy quantifiers as an explanation of set inclusion performance. *Psychological Research*, **46**, 377-388.
- ODEN, G. C. (1984). Dependence, independence, and the emergence of word features. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 394-405.
- ODEN, G. C. (1988). FuzzyProp: A symbolic superstrate for connectionist models. *Proceedings of the IEEE International Conference on Neural Networks*, **1**, 293-300.
- ODEN, G. C., & MASSARO, D. W. (1978). Integration of featural information in speech and perception. *Psychological Review*, **85**, 172-191.
- RAGADE, R. K., & GUPTA, M. M. (1977). Fuzzy set theory: Introduction. In M. M. Gupta, G. N. Saridis, & B. R. Gaines (Eds.), *Fuzzy automata and decision processes* (pp. 105-131). New York: North-Holland.
- REED, S. K. (1988). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 124-129.
- REED, S. K., & EVANS, A. C. (1987). Learning functional relations: A theoretical and instructional analysis. *Journal of Experimental Psychology: General*, **116**, 106-118.
- RESTLE, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, **83**, 274-278.
- SIEGEL, S., & CASTELLAN, N. J., JR. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- SIEGLER, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, **8**, 481-520.
- SIEGLER, R. S., & CROWLEY, K. (1994). Constraints on learning in non-privileged domains. *Cognitive Psychology*, **27**, 194-226.
- SIEGLER, R. S., & JENKINS, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.
- SOLOWAY, E., ADELSON, B., & EHRLICH, K. (1988). Knowledge and processes in the comprehension of computer programs. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 129-152). Hillsdale, NJ: Erlbaum.
- STRAUSS, S., & STAVY, R. (1982). U-shaped behavioral growth: Implications for theories of development. In W. W. Hartup (Ed.), *Review of child development research* (Vol. 6, pp. 547-599). Chicago: University of Chicago Press.
- WALLSTEN, T. S., BUDESCU, D. V., RAPOPORT, A., ZWICK, R., & FORSYTH, B. (1986). Measuring the vague meaning of probability terms. *Journal of Experimental Psychology: General*, **115**, 348-365.
- ZADEH, L. A. (1965). Fuzzy sets. *Information & Control*, **8**, 338-353.
- ZADEH, L. A. (1983). A computational approach to fuzzy quantifiers in natural languages. *Computers & Mathematics*, **9**, 149-184.
- ZIMMER, A. C. (1988). A common framework for colloquial quantifiers and probability terms. In T. Zetenyi (Ed.), *Fuzzy sets in psychology* (pp. 73-89). New York: North-Holland.

NOTES

1. The fuzzy set generated for a principle given a particular problem can be considered one member of the family of fuzzy sets specified by that principle. The exact form of the fuzzy set depends on the input parameters—in this case, the temperature and quantity of water in each container. Of course, it is possible to express each membership function

in precise mathematical terms despite the label "fuzzy." The reader may note that the proposed fuzzy sets have discrete points at which membership becomes zero. This may seem antithetical to fuzzy sets, but it is a usual feature of them. See Zadeh's (1965) original paper on fuzzy sets for examples.

2. The fuzzy set representation allows for multiple principles or propositions to be applied to a single problem in a very straightforward manner through the application of logical operators (Gaines, 1977).

3. Neither the principle representation nor the integrated representation hypothesis predicts an effect of amount of principle violation. We manipulated it to examine the possibility that potential answers beyond the limits implied by the principles might have some degree of membership in the fuzzy set. For example, the above-below principle implies that potential answers less than T_i should have zero membership. It is possible, however, that our construction of the principles imposes constraints that are not present in the subjects' representations.

4. It should be noted that the maximum deviation of the answers for each problem pair from the correct answers is related to the amount wrong, as would be expected. This is not a problem, since the purpose of manipulating of amount wrong was to systematically vary the deviation from the correct answers. It should also be noted that the maximum deviation of each problem pair from the correct answers was not

related to violating a principle. For example, consider the maximum deviations for problem pairs that violated different principles but that were all between an average of 6.2° and 7° wrong. For problem pairs that violated two principles, the maximum deviation ranged from 7.35° to 12°. For problem pairs that violated one within-problem principle, the maximum deviation ranged from 8.61° to 11.11°. For problem pairs that violated either a single principle or did not violate any principles, the range was 7.88° to 11.2°. Therefore, the maximum deviation from the correct answers does not explain the effect of violating a principle.

5. All significant effects are at the .05 level unless otherwise noted.

6. Because we focus on proportion of error in the next section, and because proportion of error, RT, and ratings of goodness and confidence were highly correlated in both experiments, we limit the present discussion to proportion of error.

7. In all the regression analyses reported here, the effect of grade was partialled out before principle score entered the equation to control for any differences in performance on the evaluation task due to grade. However, the substantive results are unchanged if grade is not partialled out. We performed the same analysis for the between-problem principles (monotonicity and crossover). However, neither instruction condition nor principle score was significantly correlated with the proportion of errors in the evaluation task.

APPENDIX

Principle	Amount of Principle Violation	Amount Wrong								
		Low (3.6–4.3)			Medium (6.0–7.0)			High (10.5)		
		Answer	Quantity	Temp.	Answer	Quantity	Temp.	Answer	Quantity	Temp.
Above-below	4.0				44	1	30	23	2	20
					49.2	1	50	36	2	60
	7.0				47	2	30	36	3	30
Range	4.0				48.8	2	60	33	3	60
					30.7	1	20	16	3	20
	7.0				54	1	50	43	3	60
EQT	4.0				23	2	30	34	2	20
					48.8	2	60	67	2	60
	7.0				36	1	40			
Monotonicity	4.0	44	2	40	56.2	1	60			
		46.8	2	60	33	3	40			
	7.0	33.5	2	20	45.4	3	60			
Crossover	4.0	47	2	40	42.8	3	50			
		36.8	2	20	41.4	3	60			
	7.0	35.4	2	30	39.8	3	20			
No violation	4.0	31.8	1	30	35.8	3	30			
		34	1	20	33.2	2	20			
	7.0	38	2	20	37.2	3	20			
		30.2	1	30	28.2	1	20			
		37.2	2	30	35.2	3	20			
		38.8	2	30	41.8	1	60	58.2	1	60
		37.4	3	30	49.4	3	60	59.4	2	60

Note—For problem pairs that violated a principle, the two levels of violation are shown in the left column to the right of the principle name. EQT, equal-temperatures-equal.