Cognitive Development, 7, 81–108 (1992)

Development of Intuitive and Numerical Proportional Reasoning

Valerie Allen Ahl
University of California–Berkeley

Colleen F. Moore
James A. Dixon
University of Wisconsin–Madison

The relationship between intuitive and numerical proportional reasoning was examined using a temperature-mixing task with fifth graders, eighth graders, and college students. In the intuitive task the temperatures and quantities were described verbally, whereas in the numerical task, numbers were used and subjects were instructed to try to use math. Half the subjects were given the intuitive version first, and half were given the numerical version first. To the extent that subjects are capable of using their intuitive knowledge to direct their numerical performance, performing the intuitive version first should make intuitive knowledge more salient and improve performance on the numerical task. Performance in the numerical condition depended on the task-order manipulation, but performance in the intuitive condition was almost the same in the two task orders. Five components were used to generate a profile representing each person's performance on each task version. Subjects were grouped according to the degree of similarity of their component profiles to several hypothesized, qualitatively different prototype patterns. This "fuzzy set" analysis showed that the frequencies of subjects showing different patterns varied across versions of the task and were age related. Performing the intuitive version first decreased the likelihood that numerical temperature would be treated as an extensive rather than an intensive quantity. A theoretical framework is outlined for the relationship between intuitive and numerical task performance.

Proportional reasoning is a pervasive activity that transcends topical barriers in adult life. We use proportions while evaluating such variables as speed, fairness,

Authorship is equally shared. This research was supported in part by a grant to the first author from the Trewartha Research Fund of the University of Wisconsin–Madison, and a grant to the second author from the Graduate School, University of Wisconsin–Madison. We thank Timothy F.H. Allen and Jonas Langer for helpful conversations about the manuscript.

Correspondence and requests for reprints should be sent to Colleen F. Moore, Psychology Department, 1202 W. Johnson Street, UW–Madison, Madison, WI 53706. Requests for reprints should be sent to Valerie Allen Ahl, Psychology Department, University of California, Berkeley, CA 94720. Correspondence may also be sent by electronic mail to: bitnet:CFMOORE@WISCMACC or internet: CFMOORE@MACC.WISC.EDU.

Manuscript received December 27, 1990; revision accepted September 13, 1991

quantity, and probability. One interesting feature of many judgments which involve proportional reasoning is the absence of a precise numerical scale. When a numerical scale is unavailable, everyday knowledge of the task domain is used to make intuitive judgments. It is only in a mathematical context, with a fixed scale and objectively correct result, that the use of proportions to solve problems is formalized. Inhelder and Piaget (1958) made a distinction between qualitative and quantitative proportions, and proposed that a qualitative grasp of proportions precedes the ability to manipulate numerical proportions. However, there has been almost no research exploring the relationship, if any, between numerical proportional reasoning, which involves explicit calculations, and qualitative or intuitive proportional reasoning, which involves estimation and everyday knowledge.

We use the term "intuitive" not in the Piagetian sense, but rather in the same way as Brunswik (1956) and others used it (Hammond, 1982; Hammond, Hamm, Grassi, & Pearson, 1987; Kahneman & Tversky, 1982) to characterize human judgment that is based either on information for which a numerical scale is unavailable, or when a judgment is reached by informal processes without explicit calculation. Thus, intuitive proportional reasoning occurs without actual calculation. Inhelder and Piaget's (1958) term "qualitative proportions" is similar to our use of the term intuitive proportion. We prefer the term intuitive because it is widely used in the judgment and decision-making field. Quantitative or numerical proportions, on the other hand, are formalized and analytical and involve explicit computations. Explicit numerical calculations belong to the category of cognition that Brunswik (1956) referred to as "analytic." Thus, our approach follows the distinction between intuitive and numerical cognition that was made by Brunswik, and further developed by Hammond (1982; Hammond et al., 1987).

Given Inhelder and Piaget's influence on the study of the development of proportional reasoning, it is surprising that a central problem which exists in the literature is a failure to distinguish between tasks that require intuitive versus numerical proportional reasoning skills (Strauss & Bichler, 1988; Surber & Haines, 1987). For example, Siegler and Vago (1978) used the terms "calculating volumes," "compute nonmetrically" and "implicit multiplication" when referring to how subjects made intuitive judgments of a proportional concept—relative fullness—in the absence of numerical information. These terms reflect Siegler and Vago's effort to express the type of judgments Brunswik characterized as intuitive (and which Inhelder and Piaget termed "qualitative" proportional reasoning) and to distinguish that process from proportional reasoning involving explicit numerical calculation.

The failure to distinguish intuitive and numerical proportional reasoning in much of the literature has resulted in inconsistent and confusing experimental results. Numerical proportional reasoning is possible only when numerical values are given, or when it is possible for the subject to infer numerical values and then use them to solve the task. Some studies of proportional reasoning have provided stimuli with numerical values, whereas in other studies, the stimuli were neither accompanied by numbers, nor easy to quantify explicitly (Chapman, 1975; Goldberg, 1966; Martorano, 1977; Noelting, 1980a, 1980b). When comparisons of proportional reasoning have been made across different task domains, the use of numerically valued or easily quantifiable stimuli is often confounded with task domain (Bart & Mertens, 1979; Martorano, 1977; Siegler, 1981).

Our study explored the relationship between intuitive and numerical proportional reasoning for problem isomorphs of a temperature-mixture task. Subjects were asked to predict the temperature of a container of water produced by combining two separate containers at different temperatures. The same problems were presented in two ways. In the numerical condition the use of quantitative or numerical proportions was encouraged by the presence of numerical values for temperatures and quantities and instructions to use math. In the intuitive condition, the temperatures and quantities were presented pictorially and accompanied by verbal descriptions. Numbers were not provided, thereby reducing the likelihood of explicit calculation in the intuitive condition (Hammond et al., 1987).

We manipulated the presentation order of the intuitive and numerical versions of the task in order to explore the ability of subjects to use their intuitive understanding to guide their numerical solutions. In past research it has been reported that when a temperature-mixture task was presented numerically and computation was encouraged, even some eighth graders produced answers that treated temperature as if it were an extensive rather than an intensive quantity[1] (Moore, Dixon, & Haines, 1991; Strauss & Stavy, 1982). Such counterintuitive answers were frequently based on adding the numerical values of the temperatures. If the treatment of temperature as an extensive quantity was due to the subjects' failure to use their intuitive understanding to guide their numerical solutions, then presenting the intuitive version of the task first would be expected to decrease the use of numerical strategies, which treat temperature as extensive. Experiencing the intuitive version of the task first should increase the memory availability of subjects' intuitive knowledge, which in turn, should make available the intensive nature of temperature. An alternative possibility is that children have difficulty using their intuitive understanding to guide their numerical solutions, even when they are aware of the need to do so. Perhaps children do attempt to use their intuitive understanding, but fail in coordinating the two types of knowledge. If this is the case, then the order of presentation of the two versions

[1] Piaget (1941/1965) described intensive quantities as those which are "not susceptible of actual addition" (p. 244), in contrast with extensive quantities such as mass for which the combined result of two quantities is the sum. See Strauss and Stavy (1982) for an interesting analysis of the development of intensive quantities.

of the task would not be expected to influence performance in the numerical version of the task.

It is also possible that only those subjects with very poor intuitive understanding of the task will treat temperature as an extensive quantity in the numerical task. Unfortunately, previous studies have not assessed intuitive task understanding for the same subjects who perform the numerical task. An innovative aspect of our research is that we measured intuitive understanding using recently developed methods (Moore et al., 1991). In this study each subject performed the task in both the intuitive and numerical versions, allowing a direct test of the role of intuitive understanding in determining numerical task performance.

Components of Understanding

Rather than measuring performance globally, we scored the degree to which each person's pattern of responses showed understanding of several different components of the temperature-mixture task (Moore et al., 1991; Reed & Evans, 1987; Surber, 1980, 1984). We chose five components that were previously used by Moore et al. (1991) and are similar to those of Reed and Evans (1987). The components are listed in the Appendix and are best explained with reference to the specific design of this study. For each version of the task, the subject was presented with trials generated by a 2 × 4 × 3 (Standard Temperature × Added Temperature × Added Quantity) analysis of variance (ANOVA). The subject was asked to predict the temperature of the "standard" container after the contents of another container of water had been added to it. Figure 1 shows the pattern of correct answers for the trials. The correct equation for solving the temperature task is a weighted average: $T_F = (Q_S T_S + Q_A T_A)/(Q_S + Q_A)$, where T_F is the final temperature, Q_S and T_S are the quantity and temperature of the standard, and Q_A and T_A are the quantity and temperature of the water added to the standard.

The components were scored by examining the ordinal features of each subject's pattern of responses. For example, in order to score the main effect component, the subject's temperature judgments for the added temperature values of 20° and 80° were compared for each quantity. If the subject's answer for the 80° trial was higher than for the 20° trial, a point was added to the component score. Because there are 3 quantity values for each standard, the maximum raw score for the main effect component was 6. The other components were scored analogously as described in the Appendix. Higher component scores represent response patterns that conform more closely to the correct answers shown in Figure 1.

Several features of the components are noteworthy. First, some of them are selected to measure the same general concept at a coarse- versus fine-grained level. The main effect and monotonicity components measure the understanding that the higher the added temperature, the higher the resulting temperature. The main effect component is coarse grained, whereas the monotonicity component is

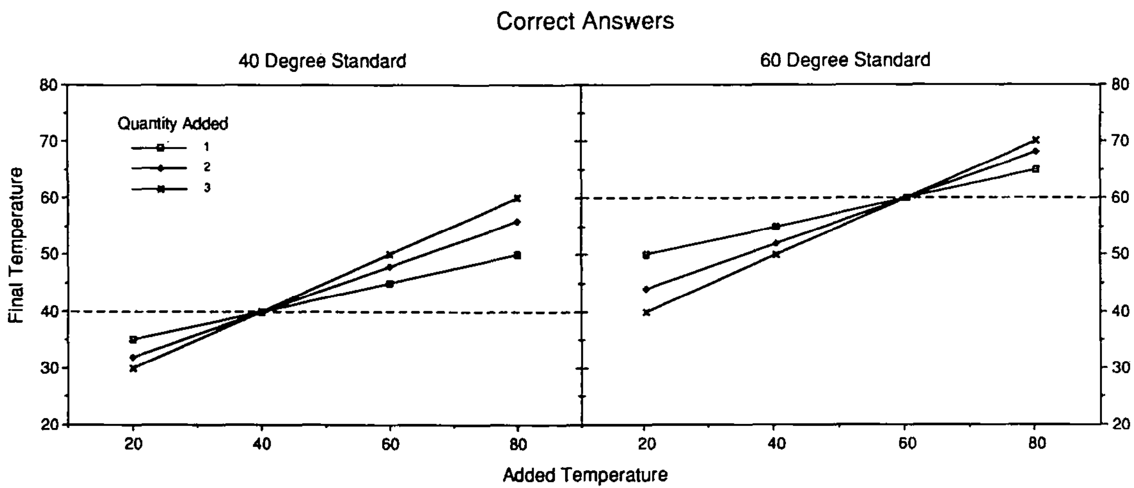

Figure 1. Correct answers for each standard (40° and 60°) plotted as a function of added and final temperature. Each curve represents a different quantity. The dashed lines represent the initial temperature of the standard.

fine grained. The above–below and range components also form a coarse- and fine-grained pair. We selected components in this way because it is possible that initial understanding would be detected by the coarse-grained component, whereas developmental refinements in understanding might be detected only by a more fine-grained variable.[2]

Second, although some of the components are similar to effects that can be extracted from data patterns using ANOVA (e.g., crossover is essentially an ordinal index of the interaction of quantity and temperature, and main effect is an ordinal index of the main effect of added temperature), the components-of-understanding approach allows analysis of individual data patterns that are obscured in group analyses. In the functional measurement approach, the data patterns of individuals are sometimes analyzed using ANOVA, and the results are used to classify individuals as using various rules for making judgments in a task (Anderson & Cuneo, 1978a; Kun, Parsons, & Ruble, 1974; Lopes, 1976). This approach has been criticized in developmental research because of limits on statistical power (Bogartz, 1978), although there is debate about the power issue (Anderson & Cuneo, 1978b; Gigerenzer & Richter, 1990). Another problem in developmental research is that the attentional capacities of the subjects may be exhausted before more than one or two replications of the design can be collected. The components of understanding provide a nonstatistical approach to the analysis of individual data patterns.

Using the components of understanding, the manipulation of task order can be predicted to influence some components more than others. If numerical performance, which treats temperature as an extensive quantity, results from a failure to use intuitive knowledge, then task order should influence the above–below and range component scores from the numerical task. For example, if a subject adds numerical temperature values (therefore treating temperature as extensive), then the above–below and range component scores will be poor because the subject's answers will all be above the standard and out of range. If subjects are less likely to add numerical temperatures after experiencing the intuitive version of the task, then the range and above–below scores should be higher in the intuitive-first than in the numerical-first order on the numerical version of the task. In contrast, the main effect and monotonicity component scores should not be affected by a decrease in the likelihood of treating temperature as an extensive quantity because adding temperatures produces response patterns that will have high main-effect and monotonicity scores.

---

[2] The scores on the coarse- and fine-grained pairs of components are related hierarchically. A person's above–below score must always be greater than or equal to the range score. The relation between main effect and monotonicity is more complex. When the main effect score is perfect, monotonicity must be at least 33% of maximum. One implication of the component structure is that some component score patterns are impossible. For example, it would be impossible to have a pattern that would have maximum scores on all components except main effect.

Fuzzy Set Approach

In most developmental studies, age-group means are used to make inferences about developmental trends. Group means can mask important individual differences, and can even present a distorted picture of the performance of individuals (Surber, 1980; Wohlwill, 1973). Moore et al. (1991) introduced a new approach, based on Zadeh's (1965) fuzzy set theory, for analyzing both developmental sequences and individual differences. In the fuzzy set approach subjects are grouped according to the similarity of their component profiles to prototype component patterns representing different hypothesized developmental landmarks, or "fuzzy developmental levels." The fuzzy developmental levels are essentially categories with fuzzy boundaries (Oden, 1977; Smith & Medin, 1981). The prototypes define the perfect member of each fuzzy developmental level. Degree of membership in a "fuzzy developmental level" is regarded as continuous and is measured by the similarity of a person's profile to the prototype for that developmental level. A subject can be classified as "in" that fuzzy developmental level for which the degree of membership is highest, but the degree of membership allows for individual differences within developmental levels. Degree of membership also allows a subject to be characterized as a member of two or more levels to some extent, as proposed by Flavell (1971).

The fuzzy set approach is similar to Siegler's (1976) rule-assessment method in which each subject's pattern of responses is used to diagnose the subject's strategy for a task. It is interesting that in a preliminary report using the rule-assessment method, the rules were referred to as task-specific developmental "stages" (Siegler & Simon, 1975). The fuzzy set approach also uses response patterns to classify subjects and the classification is task specific. A difference between the approaches is that the fuzzy set approach provides additional information about the *degree* to which the subjects fit the categories to which they were assigned.

In this study, the fuzzy set approach was used as a way of testing the effects of task order at the individual level. We selected the six fuzzy set prototypes that were also used by Moore et al. (1991). The component pattern for each prototype is presented in Table 1 (p. 88). The everything-wrong and everything-right prototypes represent the beginning and end points of development, respectively. The adding prototype represents the pattern of component scores that would be produced by literal addition of the numerical temperatures or any other approach in which temperature is treated as an extensive quantity. The other three prototypes require that temperature be treated as an intensive quantity. They were based on the possibility that range and crossover components might be acquired in a single developmental sequence, or in either order. Both of these components require coordination of two variables. The range component represents the understanding that the final temperature should be between the two combined temperatures. The crossover component represents the understanding that quan-

Table 1. Component Patterns for the Fuzzy Set Prototypes

	Component Score				
Fuzzy Set Prototype	**Main Effect**	**Mono.**	**Above–Below**	**Range**	**Crossover**
1. Everything-right	Perfect	Perfect	Perfect	Perfect	Perfect
2. Everything-but-crossover	Perfect	Perfect	Perfect	Perfect	Chance
3. Everything-but-range	Perfect	Perfect	Perfect	Chance	Perfect
4. Everything-but-range-and-crossover	Perfect	Perfect	Perfect	Chance	Chance
5. Adding	Perfect	Perfect	Chance	Zero	Chance
6. Everything-wrong	Chance	Chance	Chance	Chance	Chance

tity influences the magnitude of the temperature change, and can be seen in the crossover interactions in Figure 1.

Reed and Evans (1987), using an acid mixture task with college students, found that the range and crossover components contributed independently as predictors of response accuracy.[3] Thus, our prototypes allow for the possibility that the range and crossover components develop independently as opposed to in a fixed developmental sequence. If range and crossover are acquired in an invariant sequence, then subjects should be found in either the everything-but-crossover or everything-but-range fuzzy set, but not both. If there are two independent developmental paths for range and crossover, then individuals should appear in both fuzzy sets. Piaget (1960) presented examples of concepts that develop independently, such that alternative developmental sequences are found across individual children within the same knowledge domain. Our study tests for the presence of such alternate developmental paths.

Moore et al. (1991) showed that the distribution of memberships in the fuzzy developmental levels differed between the intuitive and numerical versions of the temperature-mixture task. The interesting question in this experiment is how task order influences the distribution of memberships in the six fuzzy developmental levels in the numerical condition. To the extent that subjects are able to use their intuitive knowledge to guide their numerical task performance, task order should influence the likelihood of membership in the adding fuzzy developmental level, because this level represents treatment of temperature as an extensive quantity. Alternatively, if subjects are unable to use their intuitive knowledge to guide their

[3]Reed and Evans (1987) studied an acid mixture task and used the term "monotonicity" to describe the relationship between stimulus quantity (as a proportion of total quantity) and resulting acid concentration. Reed and Evans's monotonicity component is closest to the component we call crossover. Their range component is analogous to the range component in our study.

numerical performance, then task order should not influence membership in the fuzzy developmental levels.

In summary, this study examines the relationship between intuitive and numerical proportional reasoning using components of understanding and the fuzzy set approach introduced by Moore et al. (1991). By manipulating the order of the intuitive and numerical tasks, the experiment tests the hypothesis that intuitive knowledge can be used to regulate numerical strategies, even when intuitive task understanding is incomplete.

METHOD

Subjects

A total of 224 subjects from three grade levels participated: 66 fifth graders (M = 10.5 years, range = 10–12 years), 70 eighth graders (M = 13.25 years, range = 13–14.8 years), and 88 college students (M = 19.25 years, range = 17.9–22 years). The numbers of subjects in the intuitive–first condition were 34, 37, and 44, for fifth, eighth, and college, respectively. The experiment was conducted in groups of approximately 20. The fifth and eighth graders were students in a local public school, which reported nonsystematic assignment of students to classes. Those students who received parental permission participated in their regular classroom groups. Groups were randomly assigned to order conditions. The college students were undergraduates in introductory psychology who received extra credit points for their participation. They participated outside regular class time in groups of approximately 20. One additional fifth grader was eliminated from the analysis because she did not complete the numerical condition of the task, and a second fifth grader was eliminated due to failure to follow instructions.

Procedure and Design

Stimuli were presented pictorially to avoid the mechanical difficulties of dealing with real water. The stimuli were two drawings of glasses, and a drawing of a pitcher with a handle and spout. All drawings were 12 × 26 in. (30.5 × 66 cm) in size, and constructed of felt board. Solid blue felt strips represented water. Each drawing was accompanied by a 16-in. (40.6-cm) schematic thermometer with a movable marker to indicate temperature. The subjects marked their responses in booklets, which contained reduced pictures of the thermometers, using a new picture on each trial.

All subjects participated in both the intuitive and numerical conditions, with order manipulated between groups. For each condition the drawings of the containers and thermometers were set on a table in front of the group. Two of the containers, which will be referred to as the standards, were always approximately half full of blue felt water. The two standards were constant throughout each condition. One standard always contained 40° water, and was described as "cool" water in the intuitive condition. The second standard always contained 60° water, and was described as "warm" water in the intuitive condition. One thermometer, marking the temperature of the water, accompanied each standard. In the intuitive condition the thermometers were labelled with a drawing of a snowman at the bottom, and of a fire at the top, and were not graduated. The answer booklet thermometers were also not graduated in the intuitive condition. Numbers were present only on the backs of the thermometers in the intuitive condition for the experimenter's use. In the numerical condition the thermometers were labelled from 0 to 80 in 5-degree increments on both the stimuli and the answer booklets.

On each trial the experimenter changed the quantity and temperature of the water in the pitcher. The subjects were asked to predict the final temperature of one of the standards after the water from the pitcher had been added to one of the standards. After four practice trials each condition consisted of 24 trials generated by the 2 × 4 × 3 (Standard Temperature × Added Temperature × Quantity of Added Water) ANOVA design shown in Figure 1.

The added quantities were described as 1, 2, or 3 cups of water in the numerical condition, whereas the standards were described as 3 cups. In the intuitive condition the experimenter simply said "this much water" and pointed to the felt strips on the drawings. The four added temperatures were 20°, 40°, 60°, and 80°. The numerical values were given orally and on cards displayed next to the drawing in the numerical condition. In the intuitive condition the temperatures were described as "cold" (20°), "cool" (40°), "warm" (60°), and "hot" (80°). The subjects were given the quantity of the water in the pitcher, the temperature of the water in the pitcher, and standard to which the water was added. The subjects were asked, "what would be the temperature of the water in this glass, after I've poured in this water from the pitcher?" The experimenter always pointed to the pitcher and relevant standard as she stated the question. The subjects marked their answers on the thermometer pictures in their booklets. In the numerical version they were instructed to try to use math and to write the number that represented the temperature in addition to marking the schematic thermometer. At the end of each version of the task, subjects were asked to write a description of how they found their answers.

RESULTS

Mean Component Scores

The mean component scores are plotted in Figure 2 as a function of age group, task, and task order. Because the raw component scores had different maximums and minimums, all components were linearly transformed to a percentage scale to facilitate comparison. Chance performance is 50%, except for the range component for which chance is 22%. The standard errors are included in Figure 2.

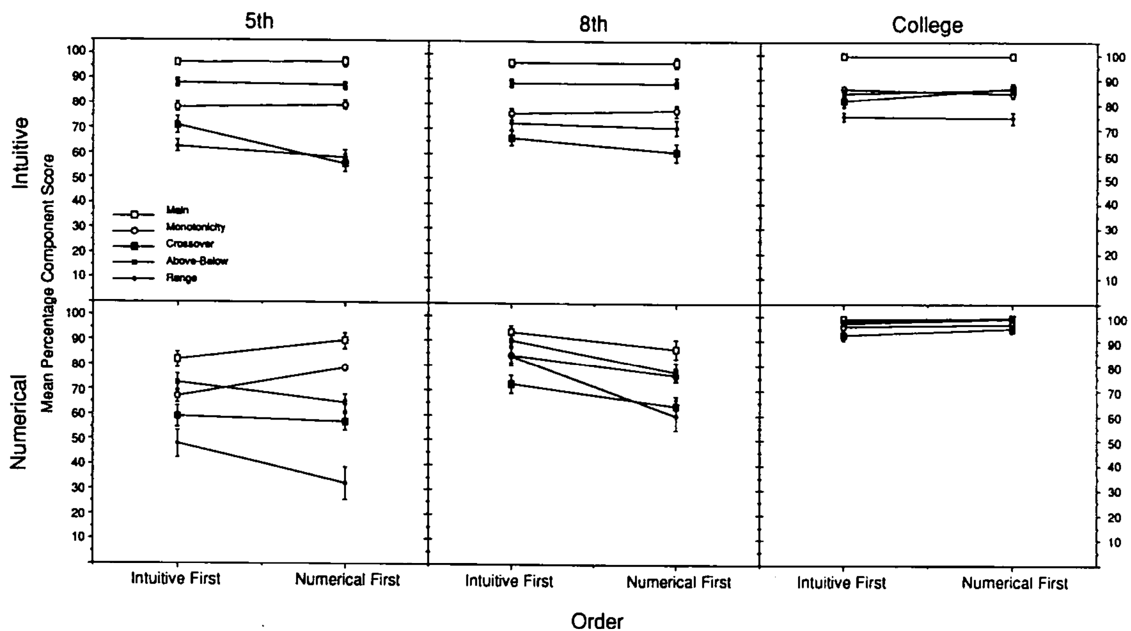

Figure 2. Mean component scores plotted as a function of grade. The top panels show the results for the intuitive task, and the bottom panels show the results for the numerical task. Bars represent ±1 standard error.

Order of tasks had a larger influence on the numerical than on the intuitive component scores. In a 2 × 2 × 3 × 5(Order × Task × Grade × Component) ANOVA, a significant interaction was found, $F(8, 872) = 2.54$, $p < .01$. A separate ANOVA of the numerical task alone showed a significant Order × Grade × Component interaction, $F(8, 872) = 4.24$, $p < .01$, a significant Order × Component interaction, $F(4, 872) = 10.32$, $p < .01$, an effect of order, $F(1, 218) = 4.89$, $p < .05$, and a Grade × Component interaction, $F(8, 872) = 32.11$, $p < .01$. Examination of the lower panels of Figure 2 shows that for the fifth and eighth graders the range and above–below components are higher in the intuitive–first order than in the numerical–first order, whereas there is no order effect for the college students. For the eighth graders all of the components were slightly worse in the numerical-first order than in the intuitive-first order. For the fifth graders, however, two of the components were slightly better in the numerical-first order (main effect and monotonicity). This result might be expected if the fifth graders are likely to add temperatures in the numerical-first order. Numerical addition would be expected to produce more consistently high main effect and monotonicity scores than estimation.

In contrast, for the intuitive task in the upper panels of Figure 2, it can be seen that there is very little effect of order on the component scores for any of the age groups. In an ANOVA of the intuitive task alone, the Order × Grade × Component interaction was significant, $F(8, 872) = 2.67$, $p < .01$, as was the Grade × Component interaction, $F(8, 872) = 16.43$, $p < .01$, but there was no Order × Component interaction, $F(4, 872) = 2.14$, $p > .05$, and no effect of order, $F(1, 218) = 2.29$, $p > .10$. The upper panels of Figure 2 show little effect of order except on the crossover component for the fifth graders. It appears that for the fifth graders, performing the numerical task first interfered with their intuitive understanding of the crossover component.

The manipulation of task order influenced performance primarily in the numerical task. The performance of the fifth and eighth graders in the numerical task was improved considerably by experiencing the intuitive task first. In contrast, there was little effect of task order on performance in the intuitive task itself or on the performance of the college students.

Fuzzy Set Analysis

The overall analyses of the age groups showed significant effects of the task-order manipulation on the numerical task. The fuzzy set method provides a way of examining individual differences within age groups. Each person's component profile represents the structure of the person's understanding of the task as expressed in task performance. As a measure of the degree of fit, or degree of membership in each fuzzy set, we calculated the Euclidean distance from each subject's component profile to each of the six fuzzy set prototypes presented earlier. A person can be said to be "in" the fuzzy set to which he or she is closest. Based on previous research, we expected that there would be more subjects falling closest to the adding and everything-wrong prototypes in the numerical task than in the intuitive task. The everything-wrong prototype represents chance performance on the task. If a subject is confused or experiments with ineffective numerical strategies, then the component profile should be close to the everything-wrong prototype. Such a pattern should be more likely in the numerical than in the intuitive task. Any response pattern in which temperature is treated as an extensive variable should be close to the adding prototype. In addition, to the extent that subjects are able to use their intuitive understanding to regulate their numerical performance, task order should influence the frequencies of subjects in the adding group in the numerical condition.

Table 2 presents the numbers of subjects falling closest to each fuzzy set prototype for the intuitive and numerical tasks when these tasks were performed first. This is a between-subjects comparison. The chi-square statistic was used to test the hypothesis of homogeneous distributions across the intuitive and numerical tasks (Marascuilo & Serlin, 1988), and was significant, $\chi^2(5, N = 224) = 108.71$, $p < .01$, showing that the distributions differ. The differences are approximately as expected based on the results of Moore et al. (1991). In the intuitive task, a high proportion of subjects fell closest to the everything-but-range and everything-but-range-and-crossover prototypes, whereas, in the numerical task, a high proportion of subjects fell closest to the adding, everything-wrong, and everything-right prototypes. The fuzzy set analysis shows that in the intuitive task subjects were unlikely to treat temperature as an extensive quantity.

Figures 3 and 4 (pp. 94–95) present the mean temperature judgments for the four fuzzy set groups with the largest numbers of subjects in the intuitive and numerical tasks, respectively. These results show that the fuzzy set approach

Table 2. Numbers of Subjects Closest to Each Fuzzy Set Prototype

	Task	
Fuzzy Set	Intuitive	Numerical
1. Everything-right	13	50
Between 1 and 2	3	4
2. Everything-but-crossover	11	8
3. Everything-but-range	41	1
Between 3 and 4	3	2
4. Everything-but-range-and-crossover	39	5
5. Adding	1	23
6. Everything-wrong	4	15

Note. For the chi-square test, those equidistant between two categories were placed in the higher numbered category. Data are for all age groups combined for the task performed first.

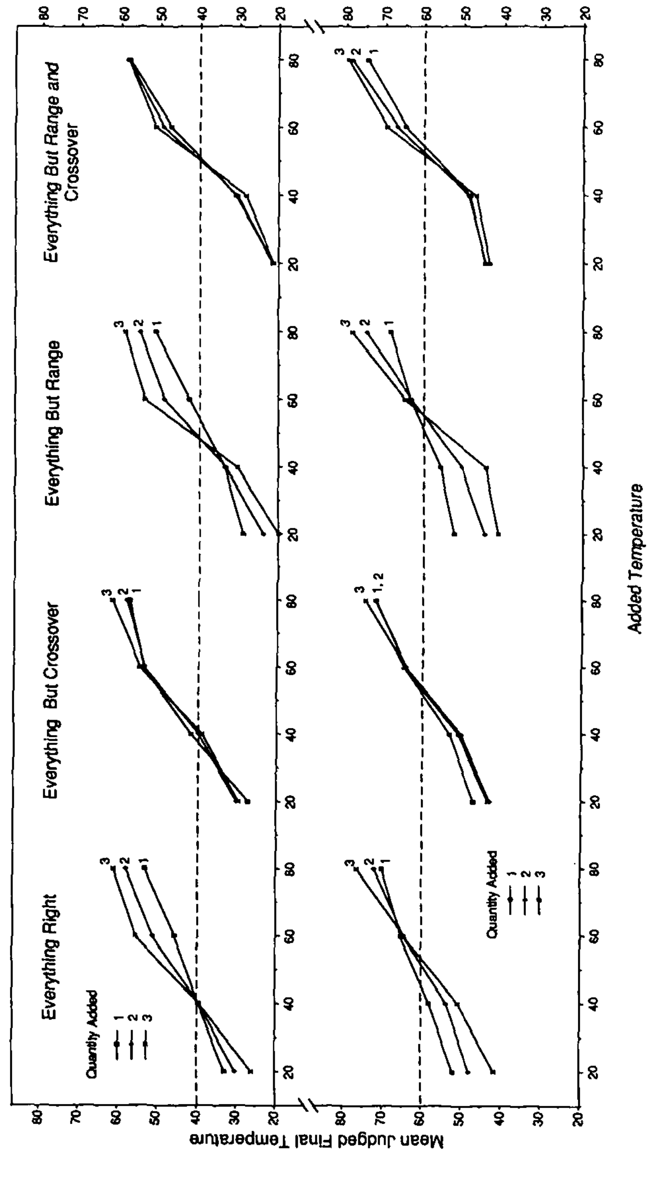

Figure 3. Mean judged final temperatures for the four most populous fuzzy set groups in the intuitive task. The upper panels show the results for the 40° standard, and the lower panels show the results for the 60° standard.

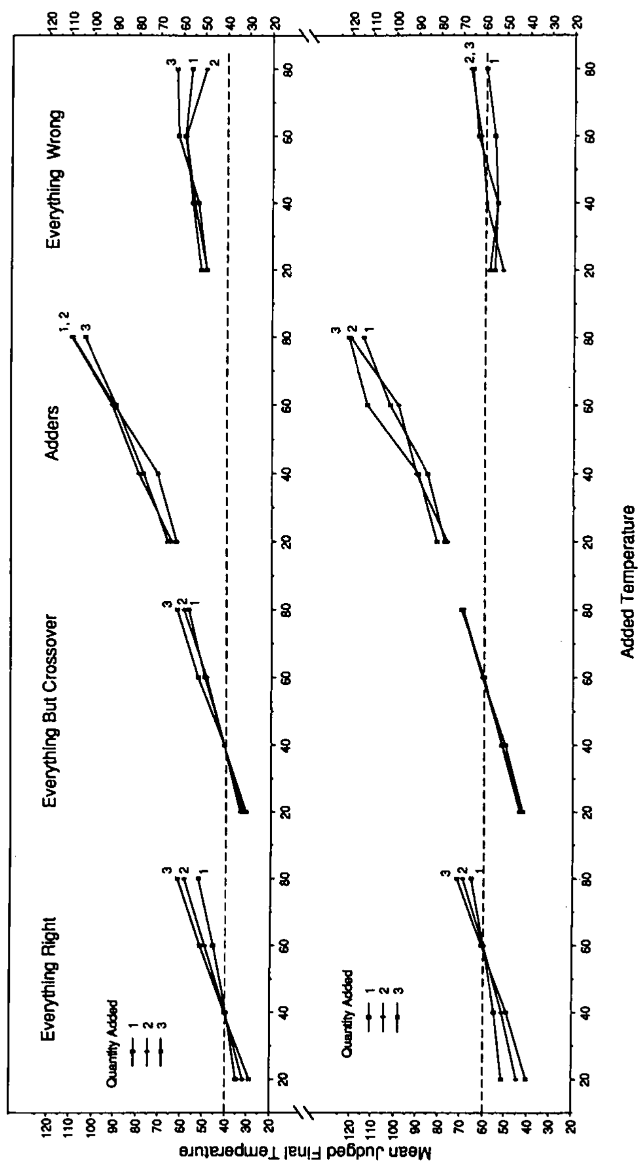

Figure 4. Mean judged final temperatures for the four most populous fuzzy set groups in the numerical task. The upper panels show the results for the 40° standard, and the lower panels show the results for the 60° standard.

captures groups of individuals with qualitatively different response patterns. Note that the four largest groups differ between the task versions. Each fuzzy set group has a distinct pattern of judgments that is consistent with its prototype definition. An ANOVA performed on the four most numerous fuzzy set groups in each task version showed a significant Group × Quantity × Temperature interaction, $F(18, 1254) = 6.61$, $p < .01$, $F(18, 1206) = 3.96$, $p < .01$ for the intuitive and numerical task versions, respectively. These interactions verify that the pattern of responses differed significantly among the fuzzy set groups.

First consider the intuitive groups in Figure 3. The mean responses of the everything-right group are very close to the correct answers and show a large significant Quantity × Temperature interaction with the expected crossover form, $F(6, 162) = 26.72$, $p < .01$. The everything-but-crossover group shows almost no discernible effect of quantity, and although the Quantity × Temperature interaction is significant, $F(6, 156) = 2.23$, $p < .05$, it is small and did not have the correct form. For the everything-but-range group, there was a large significant interaction of Quantity × Temperature, which showed the correct crossover form, $F(6, 426) = 56.09$, $p < .01$. However, the means of this group have a wider spread than the correct answers, consistent with the prototype definition for this group. The results for the everything-but-range-and-crossover group also showed a significant interaction of Quantity × Temperature, $F(6, 510) = 4.55$, $p < .01$, but it also did not have the correct crossover form.

The patterns of means for the numerical fuzzy set groups in Figure 4 also map neatly onto the prototype definitions. The mean responses of the everything-right group are extremely close to the correct answers, and the Quantity × Temperature interaction is large with the correct crossover form, $F(6, 636) = 272.70$, $p < .01$. For the everything-but-crossover group, the quantity curves are virtually colinear reflecting the fact that quantity was ignored by these subjects, Quantity × Temperature, $F(6, 186) = 2.86$, $p > .10$, although the effects of standard, $F(1, 31) = 483.24$, $p < .01$ and temperature, $F(3, 93) = 639.63$, $p < .01$, were significant. The adding group shows a pattern in which all the means are above the values of the standards (horizontal dashed lines), as expected if the subjects treat temperature as an extensive quantity. There were significant effects of standard, $F(1, 31) = 34.29$, $p < .01$, and temperature, $F(3, 93) = 101.23$, $p < .01$, but no other effects or interactions. Finally, the everything-wrong group shows a nonsystematic pattern of responses. Only the effect of temperature was significant, $F(3, 99) = 4.97$, $p < .01$, and compared to the other groups it was quite small. These subjects appear to have been confused by the numerical version of the task. This type of nonsystematic responding was very rare in the intuitive version of the task in any of the age groups.

Age Differences in Fuzzy Set Membership

Figure 5 presents the age distributions of those individuals who are closest to each fuzzy set prototype in the two versions of the task. These results show that

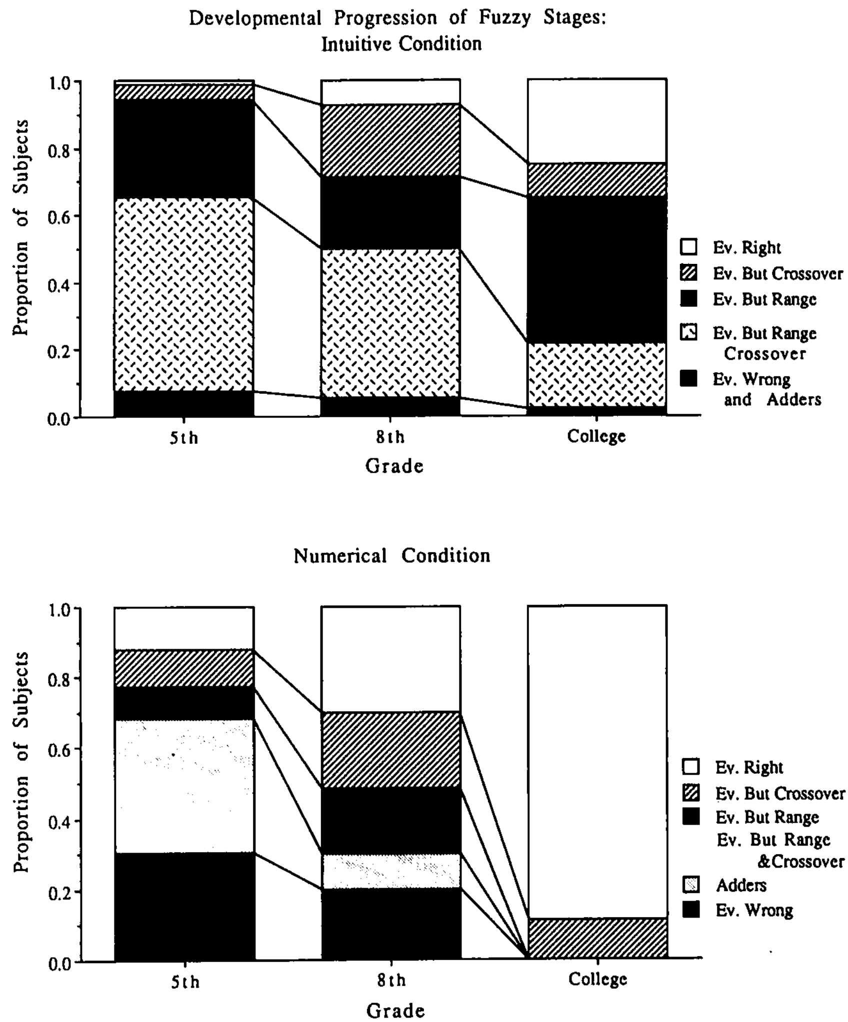

Figure 5. Percentage of subjects in each age group located closest to each fuzzy set prototype.

two different developmental pictures emerge depending on task format, demonstrating the importance of distinguishing between intuitive and numerical proportions. In the intuitive task (top panel of Figure 5) a large proportion of the college students were closest to the everything-right and everything-but-range prototypes. The largest proportions of the fifth and eighth graders were closest to the everything-but-range and everything-but-range-and-crossover prototypes. A chi-square test on the frequencies showed significant differences in membership across age groups, $\chi^2(8, N = 224) = 51.07, p < .01$. (The adding and everything-wrong groups were combined because of small expected values.) For the numerical task in the bottom panel of Figure 5 there were also significant age differences, $\chi^2(8, N = 224) = 136.06, p < .01$. (The two smallest groups, everything-but-range and everything-but-range-and-crossover, were combined.) In the numerical task it can be seen that the proportion of subjects closest to the everything-right prototype increased with age, whereas the proportions closest to the adding and everything-wrong prototypes declined with age. It is clear that college students perform exceedingly well when they are given numerical values. This may be because the college students use the numbers to calibrate their responses very accurately, so that their responses are all within range in the numerical task. In contrast to the excellent performance of the college students, the majority of the fifth graders are closest to the adding and everything-wrong prototypes in the numerical task.

Individual Differences and Developmental Paths

The fuzzy set approach provides both fine-grained information about individual understanding and coarse-grained information about the sequences of developmental levels. As a consequence, the fuzzy set approach encourages description of possible alternate developmental pathways to mature performance in a domain. For the intuitive task, the fact that subjects are found closest to both the everything-but-range and the everything-but-crossover prototypes decreases the plausibility of a universal invariant sequence for the development of the crossover and range components. The data are consistent with the interpretation that there are two developmental paths: Crossover can be acquired before the range component, or vice versa.

An important aspect of the fuzzy set approach is that it provides a measure of degree of membership, or goodness of fit, of each subject to each fuzzy set prototype. A more detailed look at the distances to the prototypes helps to show other aspects of the developmental paths for the task. The rank order of each subject's distances to the fuzzy set prototypes provides a simplified profile locating the individual with respect to all the prototypes. Figure 6 presents a sketch of the developmental paths for the intuitive task. The interior part of Figure 6 shows the positions of those subjects closest to the four most populous intuitive prototypes which form a rectangle in a plane. The rectangle is partitioned into eight regions, each defining a different rank order of distances to the four prototypes, which are labeled on the edges (Coombs, 1964). The number inside each slice of the rectangle shows the number of subjects who had that particular distance rank order. The two numbers exactly on the dashed lines are subjects exactly halfway between prototypes. This provides a diagram of the positions of the subjects in the developmental space. For example, those subjects closest to the everything-but-range-and-crossover prototype are divided between being next-closest to everything-but-crossover ($n = 47$) or everything-but-range ($n = 19$). These

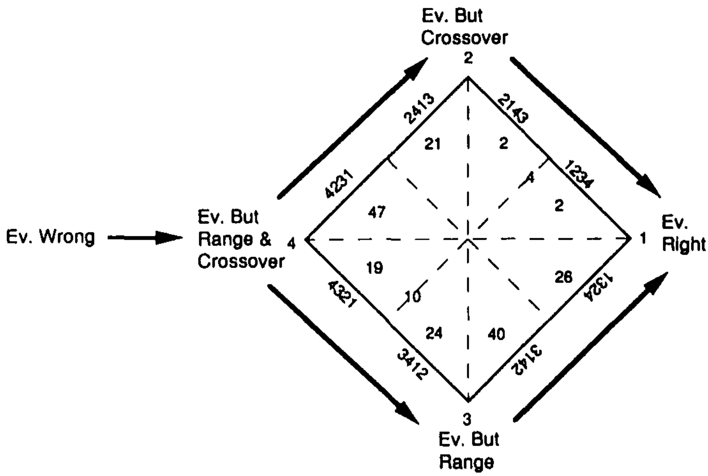

Figure 6. Developmental paths for the fuzzy set groups of the intuitive task. The interior of the diagram maps the positions of individuals with respect to four of the fuzzy set prototypes based on the rank orders of their distances to the prototypes. The sequences of numbers around the edge correspond to the rank order of distances to the prototypes for each region. The numbers at the vertices label the four prototypes. Only those subjects closest and next-closest to these four fuzzy sets are included. Prototypes are numbered as in the tables.

might represent the beginnings of the two alternative developmental paths. Thus, examination of the membership profiles shows the developmental pathways in more detail.[4]

Task Order

Because all college students in the numerical task were closest to either the everything-right or everything-but-crossover prototypes, their data are not pertinent to the hypothesized effects of task order. Table 3 (p. 100) presents the frequency of fifth- and eighth-grade subjects closest to each fuzzy set prototype in the two order conditions of the numerical task. The distributions differed

[4]One objection is that developmental pathways cannot be determined without longitudinal data. Our view is similar to that of Coombs and Smith (1973), Davison (1983) and Froman and Hubert (1980), who all presented methods of testing developmental sequence hypotheses applicable to cross-sectional data. The basic idea behind the approaches of Coombs and Smith (1973) and Davison (1983) is to use the subject's response pattern to construct a "preference order" for the various stages. The preference orders can then be used to test the hypothesis that there is a single developmental pathway. Longitudinal data provide the most definitive evidence about developmental pathways, and are essential if rate of change is of interest. For determining developmental priority, cross-sectional data are sufficient.

Table 3. Numbers of Subjects in Numerical Task Closest to Each Fuzzy Set Prototype

	Task Order	
Fuzzy Set	Intuitive First	Numerical First
1. Everything-right	19	10
2. Everything-but-crossover	14	8
3. Everything-but-range	1	2
4. Everything-but-range-and-crossover	9	7
5. Adding	9	23
6. Everything-wrong	19	15

Note. Data are from the fifth and eighth graders. For the chi-square test, Groups 3 and 4 were combined.

significantly, $\chi^2(4, N = 136) = 10.83, p < .05$. A post hoc test using Goodman's method (Marascuilo & Serlin, 1988) showed that the proportions in the adding group differed significantly across task-order conditions, $z = 3.186, p < .05, s^* = 3.081$. Thus, experiencing the intuitive task first decreased the frequency of subjects closest to the adding prototype. In contrast, for the intuitive task, there was no effect of task order, $\chi^2(4, N = 136) = 4.78$, n.s. Thus, task order influenced the likelihood that temperature would be treated as an extensive quantity in the numerical task.

Relationship Between Intuitive and Numerical Performance

A remaining issue is the relationship between level of intuitive understanding and treatment of temperature as an extensive quantity in the numerical task. In order to examine the relationship between the intuitive and numerical tasks, we conducted several analyses. Table 4 presents the numbers of subjects closest to each fuzzy set prototype for the two versions of the task. First, Cohen's kappa was calculated for Table 4 as a measure of the exact correspondence of fuzzy set categorization across tasks. The value was significant, but small and negative, $K = -.0864, z = -2.564, p < .05$. This shows a significant difference between fuzzy set groupings on the two tasks. Second, a chi-square test of independence on the data in Table 4 was significant, $\chi^2(16, N = 224) = 45.12, p < .01$, showing that there is a relationship between performance in the two tasks, although it is not one of exact correspondence. (The adding and everything-wrong groups were combined for the chi-square test because of small expected values.) As shown in Table 4, those subjects who were closest to the adding and everything-wrong prototypes on the numerical task (Categories 5 and 6) were most likely to be closest to the everything-but-range-and-crossover prototype (Category 4) on the intuitive task, and were very unlikely to be closest to the everything-right prototype (Category 1) on the intuitive task. Thus, those subjects who showed either nonsystematic response patterns, or who treated tem-

Table 4. Numbers of Subjects Closest to Each Fuzzy Set Prototype

	Intuitive Prototype					
Numerical Prototype	1	2	3	4	5	6
1. Everything-right	19	16	43	27	2	0
2. Everything-but-crossover	6	5	7	13	0	1
3. Everything-but-range	0	0	2	1	0	0
4. Everything-but-range-and-crossover	1	3	5	4	1	2
5. Adding	1	1	5	23	0	2
6. Everything-wrong	1	2	10	18	0	3

Note. Prototypes are numbered as in Tables 1 and 2. Subjects equidistant between two categories were placed in the higher numbered category.

perature as an extensive quantity in the numerical task, were most likely to be those with relatively poor intuitive understanding.

It is important to note, however, that those subjects closest to the everything-but-range-and-crossover prototype in the intuitive task do have above-chance understanding of the above–below component. An above-chance score on the above–below component implies that temperature is being treated as an intensive quantity. Thus, an intuitive response pattern which is close to the everything-but-range-and-crossover prototype (Category 4) is *not* consistent with the treatment of temperature as an extensive quantity represented by the adding prototype. This implies that, even for those subjects closest to the everything-but-range-and-crossover prototype on the intuitive task, there can be an influence of task order on their numerical performance. That is, the intuitive understanding of subjects in the everything-but-range-and-crossover group is good enough that adding numerical temperatures should seem counterintuitive *if* the subjects use their intuitive understanding to guide their numerical performance.

In summary, this experiment showed both a significant relationship between intuitive and numerical task performance, and an effect of task order on numerical task performance. The overall pattern is consistent with the interpretation that subjects are capable of using even incomplete intuitive understanding to regulate their numerical task performance.

DISCUSSION

Influence of Intuitive Task on Numerical Task Performance

This experiment shows that performing the intuitive version of the temperature task first improves performance in the numerical task. This was shown in the mean age-group differences in component scores as a function of task order, and in the differences in the distribution of subjects across fuzzy set prototypes as a function of task order. The fuzzy set analysis provides evidence that even incomplete intuitive understanding can be used to regulate responses in the numerical task.

There have been very few empirical studies in which subjects have performed the same task presented in both intuitive and numerical forms (Brunswik, 1956; Budescu, Weinberg, & Wallsten, 1988; Haines, 1988; Hammond et al., 1987). The primary goals of those previous studies have been either to compare the accuracy of performance in the two conditions, or to chronicle the characteristics of performance in the intuitive and numerical modes. The results of the previous studies do not clearly point to the superiority of either numerical or intuitive performance in adult subjects. Hammond et al. (1987) found that highway engineers performed more accurately when judging highway capacity if the task were presented numerically and they were encouraged to work out explicit equations for making their judgments, than if the task were presented pictorially without numbers. This finding seems similar to the superior performance of the college students in the numerical task as opposed to the intuitive task here. However, Hammond et al. (1987) also found that extreme errors were more likely to occur for numerically calculated answers than for intuitively based answers. Budescu et al. (1988) found that, in a gambling task, college students won slightly less when probabilities were given verbally than when they were given pictorially or numerically, but decision time was slightly faster in the pictorial condition than in the other conditions. Erev, Bornstein, and Wallsten (in press) found that when subjects made numerical assessments of probability, their choices between gambles were less optimal than when they made choices without numerically evaluating probabilities. Thus, whether adult performance in a numerical task is superior to performance in an intuitive task depends partly on the criteria used and perhaps other task variables.

In contrast to past research, in this study the goal was not to determine whether performance is "better" in the numerical versus intuitive condition, but rather to explore the ability of subjects to use even partial intuitive understanding to direct their numerical performance. The results showed that performances in the two versions of the task are related and, for the fifth and eighth graders, performing the intuitive task first improves performance in the numerical task. Thus, subjects do use even incomplete intuitive task knowledge to guide their numerical solutions. This study stands as one of only a very few empirical studies addressing the way intuitive understanding is used during numerical problem solving, or during what Brunswik (1956) called "analytic" thought.

Galotti's (1989) review of the literature on everyday and formal reasoning found very little research that has addressed the relationship between these two modes of thought. There is also, currently, little theory available regarding the interconnections between intuitive and numerical reasoning. In the viewpoint of Inhelder and Piaget (1958), the construction of a numerical proportion represents more advanced thought than does a qualitative proportion, even though the ability to think qualitatively persists after the development of quantitative proportions. For Inhelder and Piaget, both modes of thought may exist simultaneously, but qualitative understanding is a necessary prerequisite for quantitative proportions. Brunswik (1956), on the other hand, viewed intuitive and analytic (or formalized numerical) thought as opposite poles of a cognitive continuum. According to Brunswik, adult thought varies along the continuum between intuitive and analytic, and neither pole of the continuum is inherently more characteristic of mature thought than the other.

Hammond (1982; Hammond et al., 1987) developed a theory of the way that task characteristics influence the extent to which a person will perform a task intuitively versus analytically. Although a theory of how task characteristics influence one's position on Brunswik's intuitive-analytic continuum is important, it does not completely explain the effects of task order found in our study. The question addressed in our research is the extent to which subjects use knowledge that is available to them in the intuitive cognitive mode to guide their performance in a more analytic cognitive mode. For a given task it is possible that subjects may alternate among different cognitive modes, using an estimate generated in the intuitive mode to constrain the type of computational solutions generated in the analytic cognitive mode (Dixon & Moore, 1991). We propose that one important aspect of development is the ability to shift along the cognitive continuum and use intuitive cognition to guide analytic cognition. The task-order effects for the fifth and eighth graders in this study suggest that these age groups do not spontaneously make such shifts unless the memory availability of their intuitive representations is high.

Proposed Relationships Between Intuitive and Numerical Cognition

Under what conditions would a strong relationship between intuitive and numerical performance of the same task be expected? The relationship between the two is likely to depend on (a) the availability in memory of intuitive knowledge, individual mathematical operations, and mathematical formulas; (b) the ability to evaluate computed numerical answers against intuitively generated estimates, the quality of the intuitive estimates, and the likelihood of conducting this evaluation; and (c) the ability to use analogical reasoning to map one's intuitive understanding onto mathematical solutions learned in other settings. These are all processes in which it is likely that developmental change will be found.

The order effect in this study shows that increasing the memory availability of intuitive knowledge increases the correspondence between intuitive and analytic thought for younger age groups. Once intuitive knowledge is available in memory, however, it must be used to evaluate the results of one's calculations. Such comparisons require that a person make an intuitive estimate *separately* from the process of calculating a numerical answer, and then compare the two. Even when intuitive and numerical answers are generated separately and compared appropriately, if intuitive understanding is poor it may lead to the acceptance of an incorrect numerical strategy. If intuitive knowledge is adequate and is compared correctly with answers generated by numerical calculations, then a calculation scheme that is qualitatively incorrect should, at a minimum, be rejected. Once a calculation scheme is seen to be incorrect, then the subject needs to modify it or develop a new one.

In order to select computational schemes that are promising and to apply those computational schemes in a way that is consistent with one's intuitive knowledge, analogical reasoning skills are also needed. Rejecting incorrect numerical approaches does not in itself yield a successful solution. For example, a person may know how to calculate a weighted average, and might even consider that the weighted average formula would apply to temperature mixture. However, without being able to connect the variables in the formula to the quantities and temperatures in a temperature-mixture task, the correct solution will not be achieved. Gentner (1983) called the process of connecting variables across tasks "structure mapping." Intuitive knowledge of a problem provides a relational structure among the variables, and so the quality of intuitive knowledge should influence the structure-mapping process.

From this analysis of the relationship between intuitive knowledge and numerical problem solving, it is clear that there are many ways in which subjects may fail to connect their intuitive and numerical knowledge. Based on this analysis, the relatively weak relationship between intuitive and numerical performance in this study is not surprising. Further research using task isomorphs is needed to examine the proposed connections and disconnections between intuitive knowledge and numerical problem solving.

Developmental Levels As Fuzzy Sets

This study used a new approach for representing the type of developmental changes that are often viewed as "qualitative" changes. A similar approach was proposed by Davison (1983) in which the distances of individuals from the ideal response patterns of a stage are measured. As in any developmental theory, in the fuzzy set approach, the investigator hypothesizes theoretically significant developmental landmarks. The emphasis in the fuzzy set approach is *not* on whether the hypothesized developmental steps form a universal invariant sequence, describe saltatory developmental changes, or are qualitatively distinct. Instead, the emphasis is on describing the developmental paths which are observed.

By using the components of understanding to describe individuals as having different degrees of membership in the fuzzy developmental levels, a rapprochement between continuity and discrete stage theories might be achieved. For example, in a longitudinal study, progress toward a given fuzzy set would be reflected in an increase in the distance from the fuzzy set prototype to which that person was initially closest, and a decrease in distance to the next fuzzy set prototype in one of several possible developmental sequences. It becomes clear from this style of analysis that whether developmental levels are regarded as discrete or continuous is a matter of the grain of analysis. Using the coarse-grained measure of nearest prototype, development can be viewed as discrete. Using the fine-grained measure of distance to the prototype, development can be viewed as continuous and there can be considerable differences between individuals who are "in" the same fuzzy developmental level. As noted by Allen and Starr (1982), the grain of analysis can determine what structures one sees in research, and furthermore, different levels of analysis may be appropriate for different purposes. The fuzzy set approach encourages investigators to be explicit about the level of analysis they choose.

Flavell (1971) concluded that it is inadequate to view developmental stages as discrete categories between which subjects move in sudden leaps. Such a "caricature" of a developmental stage theory implies that people spend more time statically being in a stage than dynamically developing. By viewing developmental stages as categories with fuzzy boundaries, the field of developmental psychology can move beyond contentless arguments over whether developmental stages "really exist" (see Chapman, 1988, for an interesting discussion of philosophical issues related to this point). The fuzzy set approach provides methods for measuring degree of membership, rate of change between prototypes, and the pathways of development. This experiment illustrates how the effect of an experimental manipulation can be studied using the fuzzy set approach. The fuzzy set approach provides an informative way to represent simultaneously individual differences within groups and qualitative developmental sequences.

REFERENCES

Allen, T.F.H., & Starr, T.B. (1982). *Hierarchy*. Chicago: University of Chicago Press.

Anderson, N.H., & Cuneo, D.O. (1978a). The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General, 107*, 335–378.

Anderson, N.H., & Cuneo, D.O. (1978b). The height + width rule seems solid: Reply to Bogartz. *Journal of Experimental Psychology: General, 107*, 388–392.

Bart, W.M., & Mertens, D.M. (1979). The hierarchical structure of formal operational tasks. *Applied Psychological Measurement, 3*, 343–350.

Bogartz, R.S. (1978). Comment on Anderson and Cuneo's "The height + width rule in children's judgments of quantity." *Journal of Experimental Psychology: General, 107*, 379–387.

Brunswik, E. (1956). *Perception and the representative design of experiments*. Berkeley: University of California Press.

Budescu, D.V., Weinberg, S., & Wallsten, T.S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 281–294.

Chapman, M. (1988). *Constructive evolution: Origins and development of Piaget's thought*. New York: Cambridge University Press.

Chapman, R.H. (1975). The development of children's understanding of proportions. *Child Development, 46*, 141–148.

Coombs, C.H. (1964). *A theory of data*. New York: Wiley.

Coombs, C.H., & Smith, J.E.K. (1973). On the detection of structure in attitudes and developmental processes. *Psychological Review, 80*, 337–351.

Davison, M.L. (1983). *Multidimensional scaling*. New York: Wiley.

Dixon, J.A., & Moore, C.F. (1991). Intuitive understanding constrains generation of mathematical strategies. *Abstracts of SRCD Biennial Meeting, 8*, 234.

Erev, I., Bornstein, G., & Wallsten, T.S. (in press). The information-reduction assumption and applications to decision theory. *Organizational Behavior and Human Decision Processes.*

Flavell, J.H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology, 2*, 421–453.

Froman, T., & Hubert, L.J. (1980). Application of prediction analysis to developmental priority. *Psychological Bulletin, 87*, 136–146.

Galotti, K.M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin, 105*, 331–351.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Gigerenzer, G., & Richter, H.R. (1990). Context effects and their interaction with development: Area judgments. *Cognitive Development, 5*, 235–264.

Goldberg, S. (1966). Probability judgments by preschool children: Task conditions and performance. *Child Development, 37*, 151–168.

Haines, B.A. (1988). *Intuitive and computational problem-solving strategies: A developmental perspective.* Unpublished dissertation, University of Wisconsin–Madison.

Hammond, K.R. (1982). *Unification of theory and research in judgment and decision making.* Boulder: University of Colorado, Center for Research on Judgment and Policy.

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.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11*, 123–141.

Kun, A., Parsons, J., & Ruble, D. (1974). Development of integration processes using ability and effort information to predict outcome. *Developmental Psychology, 10*, 721–732.

Lopes, L.L. (1976). Model-based decision and inference in stud poker. *Journal of Experimental Psychology: General, 105*, 217–239.

Marascuilo, L.A., & Serlin, R.C. (1988). *Statistical methods for the social and behavioral sciences.* New York: Freeman.

Martorano, S.C. (1977). A developmental analysis of performance on Piaget's formal operations tasks. *Developmental Psychology, 13*, 666–672.

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.

Noelting, G. (1980a). The development of proportional reasoning and the ratio concept. Part I: Differentiation of stages. *Educational Studies in Mathematics, 11*, 217–253.

Noelting, G. (1980b). The development of proportional reasoning and the ratio concept. Part II: Problem structure at successive stages, problem-solving strategies and the mechanism of adaptive restructuring. *Educational Studies in Mathematics, 11*, 331–363.

Oden, G.C. (1977). Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 565–575.

Piaget, J. (1960). *The child's conception of geometry.* New York: Basic Books.

Piaget, J. (1965). *The child's conception of number.* New York: Norton. (original work published 1941)

Reed, S.K., & Evans, A.C. (1987). Learning functional relations: A theoretical and instructional analysis. *Journal of Experimental Psychology: General, 116*, 106–118.

Siegler, R.S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481–520.

Siegler, R.S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46*(2, Serial No. 189).

Siegler, R.S., & Simon, E. (1975). Stages in children's understanding of the balance scale problem (C.I.P. Working Paper No. 307). Pittsburgh, PA: Carnegie-Mellon University.

Siegler, R.S., & Vago, S. (1978). The development of a proportionality concept: Judging relative fullness. *Journal of Experimental Child Psychology, 25*, 371–395.

Smith, E.E., & Medin, D.L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Strauss, S., & Bichler, E. (1988). The developmental of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education, 19*, 64–80.

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.

Surber, C.F. (1980). The development of reversible operations in judgments of ability, effort and performance. *Child Development, 51*, 1018–1029.

Surber, C.F. (1984). Inferences of ability and effort: Evidence for two different processes. *Journal of Personality and Social Psychology, 46*, 249–268.

Surber, C.F., & Haines, B.A. (1987). The growth of proportional reasoning: Methodological issues. In R. Vasta & G. Whitehurst (Eds.), *Annals of child development*, (Vol. 4, pp. 35–87). Greenwich, CT: JAI.

Wohlwill, J.F. (1973). *The study of behavioral development*. New York: Academic.

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.

APPENDIX

Components of Understanding for Temperature Mixtures

Main Effect

Description. A coarse-grained measure of understanding the principle that the higher the added temperature, the higher the final temperature.

Scoring. For each quantity, the answers for the highest and lowest values of the added temperature are compared. One point is added to this component if the responses are ordered correctly, one point is subtracted if they are ordered incorrectly.

Monotonicity

Description. A fine-grained measure of understanding the principle that the higher the added temperature, the higher the final temperature.

Scoring. For each quantity, the ordering of the answer for adjacent values of added temperature was examined. One point was added to this component if the order was correct, and one point was subtracted if the order was incorrect.

Above–Below

Description. A measure of understanding the principle that the final temperature must be above the standard if the added temperature was above the standard, and that the final temperature must be below the standard if the added temperature was below the temperature of the standard.

Scoring. For each added temperature above the standard, one point was added to this component if the response was above the standard, and one point was subtracted if the response was below the standard. Responses to added temperatures below the standard were scored analogously.

Range

Description. A measure of understanding the principle that the final temperature must be within the range of the temperature of the added and standard quantities.

Scoring. For each response that fell within the specified range, one point was added to this component; one point was subtracted for responses out of range. Stimuli for which the added temperature was equal to the standard were excluded.

Crossover

Description. A measure of understanding the principle that the change produced by a given added temperature depends on the quantity of added water.

Scoring. For each added temperature, the ordering of the responses to the highest and lowest quantities was examined. One point was added to this component if the ordering was correct, and one point was subtracted if the ordering was incorrect.