Mental operations on number symbols by children*

SUSAN HOFFMAN

University of Florida, Gainesville, Florida 32601

TOM TRABASSO

Princeton University, Princeton, New Jersey 08540

and

MORTON FRIEDMAN

University of California, Los Angeles, California 90024

Children 4 to 6 years of age made choices among colors that had been associated with numbers. Prior to learning four number-color associations, pretraining on counting and ordering objects was given to half of the Ss. Associations were established by either absolute (paired associate) or relative (choice discrimination) methods, and then tests involving "more or less" questions on each of the possible six color pairs were given. Numerical associations were acquired best with absolute training or counting-ordering pretraining. Using the children's justifications for their responses, the test data were fit by Estes's (1966) scanning model; a response strength model based on Luce's (1959) axiom failed. Young children apparently possess an ordered numerical scale, can map this scale onto other nonquantitative symbols, and can generate, compare, and make ordinal decisions using these symbols.

A central question in the development of quantitative concepts is at what age can a child perform ordinal comparisons (cf. Klahr & Wallace, 1973). A related question is whether the child can perform mental operations upon an ordinal scale in answer to questions directly or by inference (Flavell, 1963). This paper reports an attempt to develop a procedure to study these questions, focusing on number as the scale.

In the task that was used, the underlying scale of numerical values was a combination of both equality and inequality relations: $A > B = C > D$. In order to assure that the quantities were symbolic, the child was required to map numerical values onto color names and to use only the colors in making decisions about numerical relations. This was done by substituting a different colored cup for each letter and by assigning numerical values to the colors via two training procedures, either paired associate or choice discrimination.

Let $x = 1$, $y = 3$, and $z = 5$. Then, for the paired associate (absolute) procedure, $A = x$, $B = y$, $C = y$, and $D = z$ represent the initial information. For the choice discrimination (relative) procedure, S was required to choose one color from the set (A,B) or (C,D) where the mapping of colors to numbers was the same as above. After training, the S's knowledge was tested by asking S to make relational choices between all possible pairs of colors.

The experiment was designed to compare two

alternative models: a "response strength" vs a "scanning" model, both of which generate quantitative predictions on the test data. In a related study, Campione (1969) rewarded retarded children for choosing Stimulus A when paired with B and Stimulus B when paired with C and then tested all three pairs, AB, BC, and AC. Stochastic transitivity was obtained, since the choice of A over C was greater than that for A over B or B over C and the data were fit by Luce's (1959) constant ratio rule, implying that the child possesses an underlying ratio scale of response strengths to the different stimuli.

An alternative model (Estes, 1962) is one that involves a memory search which leads to a comparison and decision process on values generated from each color. Application of this "scanning" model by Estes (1966) has been made successfully on number of "reward magnitude" choices by both adult humans and monkeys (data from Meyer et al, 1966, were fit by Estes).

METHOD

Subjects

The Ss were 48 children from the University Elementary School, University of California, Los Angeles. There were eight boys and eight girls in each of three evenly distributed age groups of 4, 5, and 6 years. Each S was tested alone. Two boys and two girls of the same age were assigned to each experimental condition.

Procedure

Four 5-oz paper cups colored blue, green, pink, or yellow were used as stimuli. Plastic trinkets (five of which S was allowed to keep after completion of the experiment) were used to indicate the number associated with each color. The S sat next to E at a table and was told that a game was to be played and that he was to try his best to do well at playing it. The S was

*Based on an undergraduate BS honors thesis by the first author, Department of Psychology, University of California, Los Angeles, this research was supported by USPHS Grant MH-19223 to T. Trabasso. The authors wish to thank Madeline Hunter of the University Elementary School, University of California, Los Angeles for her cooperation in providing Ss. Reprints may be obtained by writing to T. Trabasso, Department of Psychology, Princeton University, Princeton, New Jersey 08540.

Table 1
Mean Number of Errors and Standard Deviations in Training

	Kind of Training				Total
	Absolute		Relative		
	Mean	SD	Mean	SD	
Pretraining	1.50	0.63	1.00	1.67	1.25
No Pretraining	6.92	0.75	3.42	3.43	5.17
Total	4.21		2.21		

shown each cup and asked to name its color. All Ss readily did so.

In order to sensitize Ss to the numerical values of the trinkets, a pretraining exercise on counting and ordering was given to half of the Ss. The S was shown piles of one, three, or five trinkets. Then E pointed to each pile in turn and asked how many trinkets it contained. All Ss answered these questions correctly. Then, as E pointed to pairs of piles, she asked, in turn and in different orders, "Which is more, five or one (three or five, three or one)?" and "Which is less, five or three (one or five, three or one)?" All Ss answered these questions without error, showing that they had ordinal knowledge of those numbers.

Then E set out for all Ss the four colored cups in a row perpendicular to S's view and placed five trinkets in the first (A), three in the second (B) and third (C), and one in the fourth (D) cup. The color-number assignments were, thus, A-5, B-3, C-3, and D-1. These assignments were counterbalanced over colors and Ss.

Each training condition was composed of Ss of mixed pretraining experience: Half of the Ss had the counting ordering practice, while the other half did not. In the absolute training condition, the cups were set out in the A-5, B-3, C-3, D-1 order. Then each cup was presented, in turn from left to right, and S was required to guess how many trinkets it contained. After S responded, the cup's contents were poured out and displayed for about 4 sec. The S was asked whether his or her guess was correct and then, without comment, E returned the trinkets to the cup and replaced it in its position. This paired associate training continued until S anticipated correctly all four amounts twice in succession.

In the relative condition, a discrimination training procedure was used in which S was presented with a pair of cups, either (A,B) or (C,D), and asked to choose the cup that contained the most trinkets. After a choice, the contents of both cups were poured out and displayed separately. The S was asked if his choice was correct and E returned each pile of trinkets to the appropriate cup without further comment. The pairs were alternated in order of presentation until S made eight consecutive correct choices.

The test phase was the same for all Ss. Each S was shown one of the six possible pairs of cups and asked a question. Each pair was tested two times, once with the question "Which cup had more?" and once with the question "Which cup had less?" The 12 test pairs and questions were presented in a different random order for each S. No feedback was given following a choice on a test trial. After making a choice, S was asked to explain his or her response.

RESULTS AND DISCUSSION

Training

All Ss met the training criteria and underwent testing. A 2 by 2 by 3 by 2 (pretraining by type of training by age by sex) analysis of variance on errors showed that pretraining, training, and their interaction were the only reliable effects ($p < .01$). The respective F ratios were

18.60, 71.35, and 10.46, with $df = 1/24$. Table 1 shows the means for these factors.

It may be seen from Table 1 that learning to associate numerical values in the absolute condition was harder than learning to choose a color of higher value in the relative condition. This difference is probably attributable to task requirements: The children in the absolute task had to learn four color-number associations, whereas the relative task required choices between two pairs. Counting and ordering pretraining helped learning, especially in the absolute task, since this task required number responses.

Testing

The B,C tests provide data on the question of relative vs absolute learning. In the A,B test B was "less," while in the C,D test C was "more." If Ss learned these relative values in the B,C tests, they should choose C to the question "more" and B to the question "less." On these tests, E allowed S the option of saying that the values were the same, i.e., no choice could be made. Table 2 gives the frequencies of the three possible responses to the B,C tests for both conditions.

When Ss made choices, they were random (exact χ^2 , $p = .24$). On 57% of the tests, Ss stated that the B,C choices were equal in value (most often they said, "Both cups have three toys"). The fact that these statements were more frequent ($z = 3.07$, $p < .01$) following the absolute training indicates that numerical values were learned better here; it is also the case that the relative training required Ss to make relational choices, contributing to their greater frequency following training in that condition. A similar analysis using pretraining as the criterion variable yielded nearly equal (28 vs 27) frequencies of statements of equivalence, and the distribution of B and C choices was random ($p = .20$).

Errors on the tests with unequal values were analyzed with analysis of variance, adding question (more or less) as an additional factor to those of training. The same factors were statistically significant: pretraining, training, and their interaction (F ratios were 14.75, 4.55, and 7.69, $df = 1/24$, $p < .05$). In addition, age was reliable [$F(2,24) = 9.36$, $p < .05$]. However, since the error rate was only 8.1%, the children were very accurate in their choices. Half of the errors occurred in the relative task with no pretraining; one-fourth and one-eighth each were respectively contributed by the

Table 2
Number of Ss Making Ordinal or Equality Choices on Tests of Equal Values (B,C)

Choices	Training Condition	
	Absolute	Relative
B is more than C	5	13
B is less than C	8	15
B equals C	35	20

other relative and the two absolute groups. The respective proportions contributed by the 4-, 5-, and 6-year-old groups were .50, .375, and .125. Table 3 shows the proportion of correct responses to the tests of unequal value for each condition.

Explanations

The children's reasons for their choices were classified into four categories: (1) ordinal—The numerical value of each color was stated or, less frequently, a relative difference was offered, e.g., "Blue has three and pink has one" or "Blue has more than pink." (2) Equal—The numerical values were stated and were equal or, less frequently, only a statement of equality was offered, e.g., "They both have three toys" or "Green and blue are the same." (3) Unqualified knowledge—The S said that he knew but no further comment was made. (4) No Knowledge—The S said that he did not know.

The proportion of ordinal and equal explanations was .94, and of these, 94% were numerical. The relative group without pretraining accounted for three-fourths of the relative explanations. All unqualified statements of knowing came from Ss who had no pretraining (4% of the explanations). Age was a factor only in the relative group with no pretraining, where the younger Ss made more claims of not knowing values (2% of the explanations).

Table 4 presents the proportion of explanations on the B,C tests. The main findings are that the pretraining reduced statements of unqualified knowledge and only one case of stated equality occurred in the relative group with no pretraining. The pattern of results here suggests that, if one is to obtain full explanations that justify choices made involving numerical estimates, some preliminary training on using these estimates is required. The data also indicate that explanation is dependent upon the task environment, not age or knowledge.

Models

The above results indicate that young children have knowledge of number as an ordinal scale and that they can use this knowledge to make novel judgments of equality or inequality. Two models which provide alternative interpretations of the mechanisms that underlie how the children succeed in these tests will be considered now.

Table 3
Proportion Correct Choices on Tests

Test Pair	Pretraining		No Pretraining	
	Absolute	Relative	Absolute	Relative
A, B	.96	.92	.96	.71
A, C	1.00	.96	.92	.71
A, D	1.00	1.00	.96	.92
B, C	.52	.48	.54	.56
B, D	.96	.88	1.00	.83
C, D	1.00	.96	.96	.80

Table 4
Distribution of Explanations on (B,C) Tests

Condition		Category			
		O	E	K	NK
Pretraining	Absolute	.29	.25	.00	.46
	Relative	.42	.33	.17	.08
No Pretraining	Absolute	.20	.38	.00	.42
	Relative	.46	.04	.17	.33
Total		.34	.25	.09	.32

Note—O = ordinal, E = equal, K = knowledge, NK = no knowledge.

According to the response strength model used by Campione (1969), the probability that stimulus *i* is selected from the test compound (*ij*) is proportional to the response strengths of the respective stimuli, *V_i* and *V_j*, viz

$$P_{i,j} = V_i / (V_i + V_j) \quad (1)$$

Choices on test compounds may be predicted independently from pairs of other choice probabilities. Choices on compound *i,k* may be predicted from choice probabilities for compounds *ij* and *j,k*, and it can be shown that

$$P_{i,k} = \frac{(P_{i,j})(P_{j,k})}{(P_{i,j})(P_{j,k}) + (1 - P_{i,j})(1 - P_{j,k})} \quad (2)$$

Since four colors were used and there were six test compounds, it is possible to make two independent predictions for each choice probability. One set of predictions is mediated by the strengths of colors B and C, i.e., the probability of choosing color A over D, *P_{a,d}*, may be predicted by either *P_{a,b}* and *P_{b,d}* or *P_{a,c}* and *P_{c,d}*. These predictions constitute a test of the hypothesis that the strengths of B and C are equal and constant across patterns. The other set of predictions is mediated by A and D and tests the constant ratio assumption for choices on end values.

Using the appropriate prediction pairs of choice probabilities in Table 3, maximum likelihood estimates of the *V_i* in Eq. 2 were obtained via the logistic function (Bush, 1963). Predictions for all test proportions were found for each condition and are shown in Table 5 along with chi-square tests for goodness of fit on the observed proportions of Table 3. The tests took into account small expected frequencies by combining cells where the expectation was less than five. The *P_{a,d}* and *P_{b,c}* predictions are averages of the two independent predictions within a set.

The response strength model's predictions based upon B and C tests are quite accurate. This result is expected if B and C values are equal and the data above strongly suggested this. However, predictions based on tests involving end values, A or D, are decidedly inaccurate. The discrepancies for the first three conditions listed in

Table 5
Predicted Proportion of Correct Choices on Tests: Strength Model*

Test Pair	Pretraining				No Pretraining			
	Absolute		Relative		Absolute		Relative	
A, B	1.00	.98	.96	1.00	.90	.02	.65	.98
A, C	.96	1.00	.92	.98	.96	.50	.76	1.00
A, D	1.00	—*	1.00	—	1.00	—	.91	—
B, C	—*	.50	—	.45	—	.60	—	.47
B, D	1.00	.98	.96	1.00	.96	1.00	.75	.82
C, D	.88	.50	.88	.98	1.00	1.00	.81	.82
Chi Square	0.78	21.44†	0.40	4.32	0.44	99.53†	1.59	23.22†

*The first column for each condition is based on tests involving B or C; the second is based on tests involving A or D.

† $p < .01$

Table 5 may be attributed to the fact that extreme values are used in estimating strengths, but this rationalization fails on the fourth condition. Where the probabilities are sufficiently different from unity and permit a strong test, the response strength model fails.

The "scanning" model proposed by Estes (1966; Friedman, Rollins, & Padilla, 1968) assumes that Ss make their choices on the basis of comparison of generated values rather than of response strengths. On a test trial, S scans the stimuli that are presented, searches memory, generates expected numerical values, and then selects that stimulus which is higher (if the question is "more"). If S generates values that are equal, he "rescans" and generates a new pair of values until he finds an unambiguous choice.

In the present study, independent prediction of the choice probabilities can be made using S's explanations, since these entailed numerical estimates in the majority of cases. The distribution of number responses for each stimulus is required by the model, and these were constructed using the following rules: If S stated numerical values for both colors in a test pair, one point was assigned to the appropriate cell for the color and number. If S stated a value for one color only, one point was assigned to the stated color and number combination; one-half point was assigned to the other two number values for the other color of the pair. If S indicated no knowledge of the values, then one-third point was assigned to each number for each color. These rules were used to score each test for each S and the sums were converted into proportions. The resulting distributions are shown for each condition in Table 6. The gradients in Table 6 are steep, reflecting the high degree of numerical associations, and a serial position effect is apparent.

To apply the scanning model, we denote the distributions for a test pair (i,j) as follows:

Color	Number:	1	3	5	
i		a	b	c	(3)
j		d	e	f	

Then, the probability of choosing color i as more than color j is:

$$P_{i,j} = \frac{c(d + e) + bd}{1 - (ad + be + cf)} \quad (4)$$

In Eq. 4 the numerator is the joint probability that color i is expected to have a higher numerical value than color j. The denominator is one minus the probability of a tie. For the question "less" the probability is, of course, $1 - P_{i,j}$. Equation 4 and the probabilities in Table 6 were used to predict the choice probabilities on tests (Table 3) and the predicted values are shown in Table 7.

The scanning model's predictions are close for all groups. It should be noted that these predictions, unlike the strength model, do not require estimation of any parameters from the choice data and have at least 1 df more. The scanning model is clearly superior on these data.

In overview, then, young children demonstrate a knowledge of number as an ordinal scale of measurement. Moreover, they can make decisions

Table 6
Probability Distribution of Stating Numerical Values for Each Color

Condition		Numerical Value		
		1	3	5
Absolute-Pretraining	A	.02	.01	.98
	B	.06	.85	.08
	C	.07	.87	.06
	D	.94	.04	.02
Relative-Pretraining	A	.06	.08	.85
	B	.14	.76	.10
	C	.10	.80	.10
	D	.84	.08	.08
Absolute-No Pretraining	A	.06	.04	.89
	B	.06	.88	.07
	C	.09	.82	.10
	D	.96	.02	.02
Relative-No Pretraining	A	.21	.26	.54
	B	.22	.54	.24
	C	.20	.54	.26
	D	.70	.14	.15

Table 7
Predicted Proportion of Correct Choices
on Tests: Scanning Model

Test Pair	Pretraining		No Pretraining	
	Absolute	Relative	Absolute	Relative
A, B	.98	.92	.93	.67
A, C	.98	.92	.93	.66
A, D	1.00	.98	1.00	.86
B, C	.55	.44	.50	.48
B, D	.98	.90	.98	.80
C, D	.98	.91	.98	.81
Chi Square	0.18	2.02	0.59	1.63

involving inequalities and equalities in response to ordinal questions in either direction. The method by which they do this, as indicated by the scanning model, shows that they employ quantitative estimators as output values from information stored in memory, a set of processes consistent with recent information processing accounts of number concepts (Gelman, in press; Klahr & Wallace, 1973).

REFERENCES

- Bush, R. R. Estimation of parameters. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963. Pp. 429-470.
- Campione, J. C. Transitivity and choice behavior. *Journal of Experimental Child Psychology*, 1969, 7, 387-399.
- Estes, W. K. Theoretical treatments of differential reward in multiple-choice learning and two-person interactions. In J. H. Criswell, H. Solomon, and P. Suppes (Eds.), *Mathematical methods in small group processes*. Stanford: Stanford University Press, 1962. Pp. 133-149.
- Estes, W. K. Transfer of verbal discriminations based on differential reward magnitudes. *Journal of Experimental Psychology*, 1966, 72, 276-283.
- Flavell, J. H. *The development psychology of Jean Piaget*. New York: Van Nostrand Reinhold, 1963.
- Friedman, M. P., Rollins, H., & Padilla, G. The role of cue validity in stimulus compounding. *Journal of Mathematical Psychology*, 1968, 5, 300-310.
- Gelman, R. The nature and development of early number concepts. In H. Reese (Ed.), *Advances in child development*. New York: Academic Press, in press.
- Klahr, D., & Wallace, J. G. The role of quantitative operators in the development of conservation of quantity. *Cognitive Psychology*, 1973, 4, 301-327.
- Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.
- Meyer, D. R., LoPopolo, M. H., & Singh, D. Learning and transfer in the monkey as a function of differential levels of incentive. *Journal of Experimental Psychology*, 1966, 72, 284-286.

(Received for publication December 26, 1973;
revision received February 25, 1974.)