Estimates and Estimate-Based Inferences in Young Children

KURT E. HECOX AND JOHN W. HAGEN[1]

University of Michigan

The ability of children to perform in estimation-based inference tasks was studied in two experiments. In the first study, children in the CA range 5–7 years were tested for their ability to make increasingly accurate estimates of proportions as a function of CA. A simple visual judgment task which did not involve either verbal stimuli or responses was utilized. It was found that there was above-chance level performance at all CA levels and that performance improved as a function of CA. In the second study, measures of the ability to make quantitative inferences were also studied in children in the CA range 6–8 years. It was hypothesized that: (1) performance would be above-chance level at all ages; (2) inferential performance would improve with increasing CA; (3) conservatism, defined here as the constriction of responses to the middle of the response scale, would vary inversely with performance level; and (4) attention strategies would be a significant determiner of performance. All of the hypotheses except (2) were supported. The results were discussed in terms of cognitive models which consider inference strategies and the role of language in the development of these strategies. The studies provide evidence that accurate but nonlogical modes of inference operate in the performance of children within the CA range tested.

The ability to make estimates, whether of size, shape, or relative frequency, underlies a great deal of everyday behavior. An example of this fact has been pointed out by Polyani (1968):

> The way a judgment of true coherence is arrived at can be seen when a jury faces a difficult decision. The jurors may see a pattern of circumstances pointing to the accused. But it is always conceivable that this pattern may be due to chance. How unlikely a chance should they admit to be possible? Or what degree of coincidence should be considered not to be reasonably believable? (p. 28).

In this study the performances of children at various age levels are compared on both a quantitative estimation and a quantitative estima-

tion-based inference task. A nonverbal response mode was developed to measure performance on a task which involved making judgments about visual stimuli which contained an uncertainty component. Two major questions were investigated: the ability of children to form accurate and consistent estimates of proportions based on a visual display, and the ability to revise estimates based on the presentation of new information.

In order to reach a decision in uncertain situations an individual must be able to formulate and operate upon subjective estimates of likelihoods. The development of statistical decision theory has relied heavily on the principle that people are capable of dealing in terms of probabilities of events (Edwards, 1954; Tanner, 1965; Edwards, Lindman, & Phillips, 1965). Further evidence shows that perceptual processes may be susceptible to these same kinds of analyses (Swets, Tanner, & Birdsall, 1960; Tanner, Swets, & Green, 1956). At least one approach to language learning has taken this probabilistic viewpoint also (Lenneberg, 1957). Subjective probabilities are, however, only one type of quantitative estimation.

The ability to use imperfectly valid sensory cues in a cue weighting task is another instance where the manipulation of quantitative estimates is required. The utilization of such imperfect cues seems to be a well-established capability of human beings (Peterson, Hammond, & Summers, 1965; Brunswik, 1939; Wallace, 1941). The importance of replicating these results in a variety of experimental contexts to obtain "representative" results of natural contexts has been indicated by Brunswik (Brunswik, 1947; Brunswik & Herma, 1951). His focus was primarily on the acquisition of cues in a given experimental situation while the focus of this study is the development of imperfect cue utilization in children.

Given that the environment is uncertain and changing, its representation must be based on approximation and estimation. It is difficult to see how the application of logic, the passing to strict conclusions from strict premises, is justified in these circumstances. These factors led Polyani to underline the necessity of some nonstrict logical structures in saying, "that any attempt to eliminate nonstrict modes of inference or even to reduce them to insignificance, must be misleading" (1968, p. 44). The contention is not that deductive logic operations are irrelevant to human thought, but rather, inappropriate for the complete characterization of cognition in uncertain task situations. This perspective is somewhat similar to that of probabilistic functionalism (Brunswik, 1955). It is important, therefore, to trace the development of what Polya has termed, "patterns of plausible inference" (1954). Further, it is assumed the possession of a skill or concept is not "all-or-none" but a

question of degree. To determine the degree to which a subject exhibits some behavior, it is necessary that he be able to make more than one type of error. The implication for estimation studies, then, is that the response repertoire must be greater than two.

A consideration in the development of a methodology for dealing with estimation studies is the role played by language. The difficulties in using nonverbal stimuli while retaining a verbal or verbal dependent response mode have been pointed out by Tajfel (1963):

> There would be, no doubt, some difficulties in inferring that . . . children are as yet quite incapable of manipulating discontinuous quantities in an abstract form. One such difficulty is the inadequacy of verbal instructions and questions for eliciting from the child the most effective use of numerical concepts which it might already have at its disposal (p. 113).

Considerable evidence has accumulated supporting the pivotal role played by language in the development of higher thought processes (Dietze, 1955; Kendler & Kendler, 1961; Luria, 1961; Reese, 1962). It seems that by the ages of 5–7 years language plays an important role in higher cognitive processes (White, 1965). Therefore, language assumes the dual role of methodological and theoretical determinant in the study of higher thought. One immediate implication of this dual role is that until the methodological component of language is effectively controlled, the theoretical impact of language development will be difficult to assess. In an effort to limit the effects of language on the methodological aspects of this study, the stimuli and response materials were nonverbal. However, the instruction set was verbal so that some linguistic-based variation could have occurred. The necessity of controlling for instructional set effects and response learning has been well documented (Mandler, 1967; Bruner & Rodriquez, 1953). A training sequence in which all subjects had to demonstrate their understanding of the instructions and response materials was used to control for response learning and instruction variables.

Two studies were conducted to investigate children's ability to make estimates and inferences. In Study I, three groups of children at CA (chronological age) levels 5, 6, and 7 years were tested individually to determine the development of the ability to make quantitative estimates. In Study II, the hypotheses, based on the results of Study I, were tested to investigate the development of the ability to make inference judgments and to explore the relation between performance and certain factors which affect performance in children at CA 6, 7, and 8 years.

EXPERIMENT I

The hypotheses were: (1) performance at the CA levels of 5, 6, and 7 years on a task which requires quantitative estimates of nonverbal

visual stimuli is above-chance level; and (2) performance improves as a function of increasing CA. The first study was conducted as a necessary prerequisite to the study of quantitative inferences in prelogical children.

Method

Subjects. Fifty-two children enrolled in a private suburban school in Northern Illinois served as subjects. The composition of the three age groups was as follows: CA 5.6, 9 males and 7 females; CA 6.5, 10 males and 8 females; and CA 7.6, 11 males and 7 females. Intelligence test scores were obtained. Mean Kuhlmann–Anderson IQ scores were: CA 5.6, 107.5; CA 6.5, 111.5; and CA 7.6, 106.5. Subjects were all judged to be normal in terms of behavior and personality characteristics. All were at the normal grade level except for three who were advanced one grade level each.

Stimulus materials. Visual fields of 100 red and black dots, arranged as 10×10 arrays were presented on slides. The stimuli varied in the number of red dots and the proportions were: .1, .2, .25, .3, .4, .5, .6, .7, .75, and .8 red. Slides were projected by a Kodak Carousel projector on a screen placed about 10 ft from where the subjects were seated. For all but the .2 red stimuli, the colors were randomly interspersed. There were three .2 red stimuli, varying in configuration (see Table 1).

TABLE 1
FIELD CONFIGURATION OF .2 RED SLIDES[a]

X-Design	Block design	Random
✗○○○○○○○○✗	○○○○○○○○○○	✗○○○○○○○○✗
○✗○○○○○○✗○	○○○○○○○○○○	○○○○○✗○○✗
○○✗○○○○✗○○	○○○○○○○○○○	○✗○○✗○○✗○○
○○○✗○○✗○○○	○○✗✗✗✗✗○○○	✗○○○○○○○○○
○○○○✗✗○○○○	○○✗✗✗✗✗○○○	○○○○○○○✗○○
○○○○✗✗○○○○	○○✗✗✗✗✗○○○	○○○○○✗○✗○○
○○○✗○○✗○○○	○○✗✗✗✗✗○○○	○○✗○○○○○○✗
○○✗○○○○✗○○	○○○○○○○○○○	✗○○○○○○○○○
○✗○○○○○○✗○	○○○○○○○○○○	○○✗○✗○○○○○
✗○○○○○○○○✗	○○○○○○○○○○	○○○○○✗✗✗○✗

[a] ○ = Black dot; ✗ = Red dot.

Response materials. Each subject sat in front of the apparatus pictured in Fig. 1. The essential features of the apparatus are: the viewer through which a subject sees ten columns of ten dots each, and a lever by which the subject can change the proportion of red and black columns appearing in the viewer. The viewer was 12×8 in. The following responses were possible with the given apparatus: 0, .1, .2, .3, .4, .5, .6, .7, .8, .9, and 1.0 red.
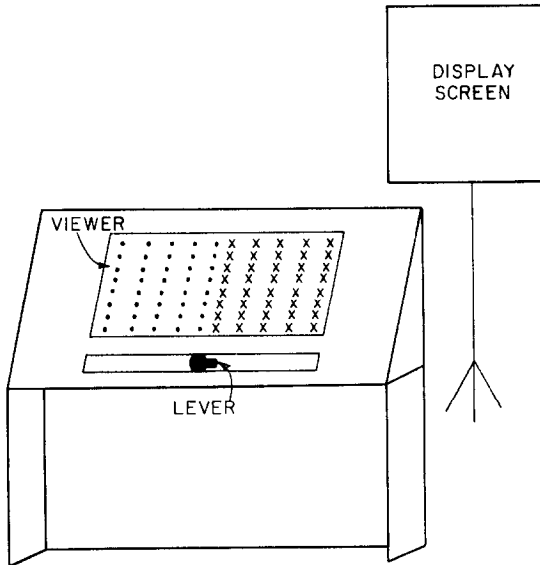
FIG. 1. Response apparatus.

Procedure. The subjects were tested individually. They were instructed in the use of the response apparatus until they could use it properly. They were then told:

> If you are ready, we will begin the game. Shown on the screen in front of you will be a picture of red and black dots, just like the ones you see in your picture (the viewer). The idea of this game is for you to make your picture have just as many red and black dots as there are on the screen in front of you. There will be too many dots to count so I don't want you to try. This is a guessing game. Now let's do some practice ones together.

The training sequence included two slides of each of the following proportions, .2, .8, .1, and 1.0 red in a highly ordered configuration. If the subject correctly performed the last three of the training sequence, the test sequence began. If unsuccessful, then the training sequence was repeated until criterion was met. All subjects met criterion. Upon attainment of criterion subjects were told:

> Now we are going to begin the real game. If you try hard, you will win a prize. Here are the prizes. Would you like to choose one to work for now? (Prizes were candy bars, gum, and small toys.) Once again, the idea of this game is for you to make your picture have just as many red and black dots as there are on the screen in front of you. Do you understand?

Total time for the test sequence was about 8–10 min. Slides were shown for about 20 sec each with 10-sec intervals between slides. The experimenter stood behind subjects changing the slides and recording responses. Each subject received the prize he had chosen. There was no feedback to the subjects concerning their performance during the test sequence. Each subject saw the slides in the following order; .4, .2, .7, .5, .3, .75, .1, .2, .5, .6, .25, .8, .4, and .2 red.

Results and Discussion

The data were analyzed in terms of the absolute difference between mean scores of subjects' responses and the actual proportion for a given slide. These error values were averaged over all slides or different groups of slides, in the different analyses. A different measure of accuracy was used for some analyses which was based on the number of correct answers (direct hits).

Age and accuracy. Average errors per subject were computed across the 14 trials. CA was found to be a significant variable ($F = 7.0$, $df = 2,49$, $p < .02$). Only the difference between the CA 5 and 7 year groups was significant ($t = 2.33$, $p < .02$), and the older children performed more accurately. The mean errors per age group were .133, .104, and .089 for the CA 5, 6, and 7 groups, respectively. All age groups performed considerably better than at chance level for the task. The average error expected for randomly determined responding would be .354.

Stimulus type and accuracy. The stimuli were grouped into high ($>.7$), middle (.4–.6), and low ($<.3$) proportion red dots, and an analysis of variance with repeated measures on one factor was performed. The difference between these stimulus groupings was significant ($F = 3.3$, $df = 2,33$, $p < .05$), with more errors on the high proportion group than the middle ($t = 7.5$, $df = 24$, $p < .01$) and the low ($t = 9.3$, $df = 24$, $p < .01$). These results are shown in Fig. 2. The interaction variance was not significant. If subjects actually based their judgments on the proportion of both exhibited colors there would be no apparent reason for the disparity between performance on the high and low proportions. One possible explanation for the difference is that subjects were attending only to the red dots. As the number of red dots increased so did the difficulty in judging their numerosity. It may be that attention strategies are important even for such relatively simple estimation tasks.

In the above analyses of slide type effects only the randomly arranged configuration of the .2 red slides was included. In an analysis of variance, repeated measure design, no differences were found among the three .2 stimulus configurations. When a direct hit analysis was applied to the data the effect of slide type was significant ($X^2 = 9.52$, $df = 4$,
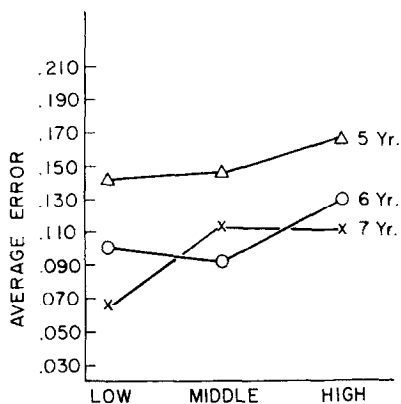
Fig. 2. Mean error as a function of slide type for each age level.

$p < 0.5$). The "X" design accounted for 43%, the random design 38% and the block design 19% of the direct hits. Although it would be difficult to make any strong conclusions with regard to field complexity and judgmental difficulty from a sample of three slides, the results suggest that this variable deserves additional research.

Response distribution. It is usually assumed that the proportion of responses is distributed similarly among various experimental groups in a decision-making situation. The data of this study indicate that this assumption is not met when subjects are compared at the CA levels tested here. The proportion of total responses in each response category is shown in Fig. 3. The slides were grouped into high, medium, and low proportion red, and a chi-square goodness-of-fit between CA groups was performed. There was a significant difference between both CA 5 and
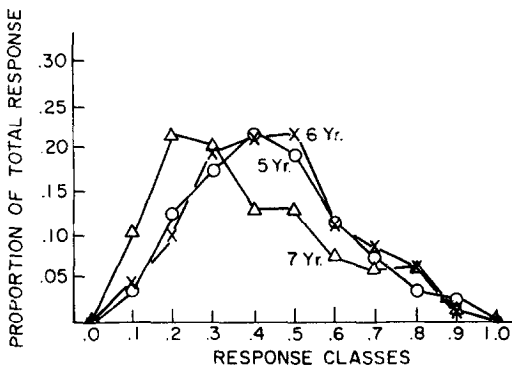

Fig. 3. Response distributions for each age level.

7 ($X^2 = 13.32$, $df = 2$, $p < .001$) and CA 6 and 7 ($X^2 = 14.64$, $df = 2$, $p < .001$).

Conservatism. Response conservatism refers to the tendency of both high and low stimuli to be judged towards the middle of the response scale rather than the extremes. To measure the degree of conservatism, the number of responses in the two response classes adjacent to the correct response class and away from the middle, was compared to the number of responses in the two response classes adjacent to the correct class but towards the middle of the response range. A group is considered conservative if the number of responses towards the middle is significantly greater than the number of responses away from the middle. This analysis was performed using the .2, .4, .6, and .8 stimuli for each age group separately.

Both the CA 5 and CA 6 year groups showed significant conservatism ($X^2 = 9.8$, $df = 1$, $p < .01$; $X^2 = 12.04$, $df = 1$, $p < .01$, respectively), while the CA 7 year group did not. When subjects were divided into high and low performance groups (median break on overall error score), the high performance group was significantly less conservative than the low performance group ($X^2 = 4.3$, $df = 1$, $p < .05$). Because of the small number of 5-year-olds above the median and the low number of 7-year-olds below the median, it was possible to test high versus low performance with age held constant only for the 6-year-olds. The high performers were not conservative while the low group was conservative ($X^2 = 20.25$, $df = 1$, $p < .001$) for the CA 6 group. The difference between the two groups was significant ($X^2 = 13.32$, $df = 1$, $p < .001$). When the medians of the 5- and 7-year-olds were used to determine within-age high and low performance groups, differences were not significant.

It seems then, that response conservatism and performance level are related. Age differences cannot adequately account for the obtained differences while performance levels can. It is not clear, however, whether conservatism determines performance or those who do not perform well are more conservative.

Relation to standard measures of ability and achievement. The subjects had been administered the Kuhlmann–Anderson Intelligence Test 6 months prior to the experiment. An analysis of covariance between task performance and IQ scores for all ages showed a significant component independent from that measured by IQ ($F = 14.1$, $df = 1,29$, $p < .001$).

The subjects in first and second grade had also been given the Stanford-Binet Achievement Test. Correlations between the subtests and task scores are shown in Table 2. It should be noted that these are correlations with errors, not accuracy; positive relations are therefore in-

TABLE 2

CORRELATION BETWEEN STANFORD-BINET ACHIEVEMENT SUBTESTS AND AVERAGE
ERROR ON ESTIMATION TASK AS A FUNCTION OF AGE

CA level	Achievement subtests			
	L1	L2	L3	A1
5 years	na[a]	na	na	na
6 years	−.40	−.20	−.29	−.23
7 years	−.09	−.14	+.16	+.38

[a] na = not available.

dicated by negative correlations. The four subtests used were word meaning (L1), paragraph meaning (L2), language (L3), and arithmetic (A1). None of the correlations reached significance. The lack of significant correlations for the language subtests is consistent with the original purpose to control the language-dependent aspects of cognitive functioning in this setting.

Additional analyses. An analysis of variance between age, sex, and average error showed no significant differences for sex ($F = 3.21$, $df = 1,46$) or with respect to interaction variance ($F = 1.83$, $df = 4,46$). An analysis of variance between morning and afternoon performances was not significant ($F < 1$).

EXPERIMENT II

Developmental trends in estimation performance were established in Experiment I. Experiment II investigates the developmental changes in ability to use those estimates in the formation and revision of quantitative inferences. A set of four tasks was given to CA groups of 6, 7, and 8 years to measure these abilities. Task 1 required subjects to estimate the proportion of red and green boxes appearing in a visual display. In Task 2, subjects were asked to estimate the proportion of red and green balls that would be found in a displayed set of closed red and green boxes, if every green box contained a green ball and every red box contained a red ball. Tasks 3 and 4 differed in that red boxes contained varying proportions of red and green balls, while the green boxes still contained only green balls.

The specific hypotheses tested were: (1) children are able to adjust their estimates in the correct direction, at a level superior to chance; (2) the degree to which subjects accurately adjust their estimates to match the altered stimulus displays is a function of CA; (3) high performance subjects exhibit significantly less conservatism than low performance subjects; and (4) attention strategies or characteristics are important determiners of performance levels.

Method

Subjects. Forty-one subjects enrolled in the same private school as in Experiment I served as subjects. The composition of the three age groups was as follows: CA 6.1, 9 males and 4 females; CA 7.2, 6 males and 7 males; and CA 8.1, 7 males and 8 females. All were at the normal grade level except for two who were advanced one grade level each. Eight of the CA 6.1, 8 of the CA 7.2, and 13 of the CA 8.1 groups had participated in Experiment I, 9 months earlier.

Stimulus materials. The stimuli consisted of pictures of arrays of closed red and green boxes (test slides) or opened red and green boxes (clue slides). Each of five different test slides consisted of 48 closed boxes differing in the proportion of red and green boxes present. The color of the boxes was randomly determined within the limitations of the specific proportions. Each of the clue slides consisted of 24 open green boxes or 24 open red boxes. Open green boxes always contained green balls, while the proportion of red balls in open red boxes varied with the series. The various types of "test" and "clue" slides are shown in Table 3. Slides were projected onto a screen about 12 ft away from the subjects.

Response materials. The response apparatus from Experiment I was used. In order for the subject to be able to match the proportions demanded in this setting, the picture appearing in the viewer showed either 10 columns of closed boxes or 10 columns of colored balls. Once again there were 11 possible responses for the apparatus.

Procedure. The subjects were tested individually. Instructions for the use of the response apparatus were the same as in the first experiment. The subjects were then told that in this game they were to make their picture have just as many red and green boxes as the picture on the screen in front of them. They were allowed to choose a prize to work for during the session. Prizes were candy bars and small toys, and were given them at the end of the entire session. They were told that they should do their best in order to obtain the selected prize.

The first series consisted of 10 test slides, which were persented in order: .5, 1.0, .25, .75, 0, .75, .25, 1.0, 0, and .5 red closed boxes. Total time for test sequence was 2–3 min; pictures were shown for 10 sec with about 5-sec intervals between presentations.

For the second series, the picture in the viewer was changed from columns of boxes to columns of balls. It was then indicated to the subjects that hereafter they were to guess the number of red and green balls they would find in the red and green closed boxes they were to be shown. In order to help them guess, each series showed two clue slides simultaneous with the test slide; one indicated the contents of the red boxes, the other the contents of the green boxes. For the second series the clue

TABLE 3
"Clue" and "Test" Slide Proportions for All Tasks

| Task | Visual display elements | | | |
| | Boxes | | Balls | |
	Red	Green	Red	Green
Task 1				
Test slide 1	48	0	—	—
Test slide 2	36	12	—	—
Test slide 3	24	24	—	—
Test slide 4	12	36	—	—
Test slide 5	0	48	—	—
Task 2				
Test slides 1–5	Same as above	Same as above	—	—
Clue slide 1	24	0	24	0
Clue slide 4	0	24	0	24
Task 3				
Test slides 1–5	Same as above	Same as above	—	—
Clue slide 2	24	0	12	12
Clue slide 4	0	24	0	24
Task 4				
Test slides 1–5	Same as above	Same as above	—	—
Clue slide 3	24	0	6	18
Clue slide 4	0	24	0	24

slides indicated a perfect one-to-one correspondence between color of box and ball.

Series 3 differed from Series 2 with respect to the red box clue slide. On this series half of the red boxes contained red balls and the remaining contained green balls, while the green boxes all contained green balls. Once the change had been indicated to the subjects and they had verbalized the change, the same test slide sequence as for Series 1 and 2 was shown, in the same order.

On the fourth series the red box clue slide changed again. This time only one-fourth of the red boxes contained red balls, the remainder of the boxes being filled by green balls. The change was indicated to the subject and he was again asked to verbalize it. The test slide sequence was the same for this series as it was for the others, and in the same order.

There was no feedback to subjects concerning performance during the test sequences.

Results and Discussion

The four series in Experiment II formed an interrelated sequence. In order to obtain a pure measure of the amount of error subjects make in using one-to-one correspondence (Series 2) it is necessary to subtract the amount of error due to the estimation alone (Series 1). Series 2, 3, and 4 differ from one another in that the likelihood of finding a red ball in a red box becomes smaller with each series. The amount of information the subject obtains from the changing clue slide determines the degree to which his estimates are altered. By comparing the responses in successive series we may obtain a measure of how much the subject changed his estimates in accord with the new information.

Age and accuracy. Analysis of variance between total average error per subject and age for the first series showed no significant differences. This finding was in agreement with the first experiment, in which 6- and 7-year-olds did not differ on this same measure.

Analysis of variance between age and average error for the one-to-one correspondence series showed a significant difference ($F = 6.50$, $df = 2,38$, $p < .01$). Subsequent tests indicated that there was a significant mean difference between CA groups 6 and 7 ($t = 3.60$, $p < .01$) and 6 and 8 ($t = 2.93$, $p < .01$). The difference between the error distributions for the first two series when grouped according to age and slide type showed no significance, for a chi-square goodness of fit. When the amount of error due to the first series was subtracted from the second series error there were no longer any significant age differences. It may be concluded that any significant differences on one-to-one correspondence tasks are not due to CA ability differences in the use of one-to-one correspondence, but rather its combination with the slight developmental differences in ability to form simple estimates.

Stimulus type and accuracy. In order to perform the third and fourth series tasks, a subject must attend to the fact that a clue slide has changed from simple one-to-one correspondence, and shift his estimates accordingly. The test slide showing only red boxes provides an index of subjects' adjustment to these changes in clue slide. In Series 4, where the clue slide contains only one-fourth red balls in red boxes, only two subjects estimated that all the closed red boxes would contain red balls. In other words, subjects clearly adjusted their estimates. In Series 3, however, such an adjustment was made far more frequently by older children than younger ones. Both the difference between 6- and 7-year-olds ($X^2 = 8.25$, $df = 1$, $p < .01$) and 6- and 8-year-olds ($X^2 = 4.51$, $df = 1$, $p < .05$) was significant. About 42% of the 6-year-olds estimated that the closed red boxes would contain all red balls despite the clue

TABLE 4

ANALYSIS OF 1.0 RED SLIDE RESPONSE FOR SERIES 3 AND 4

	Task 3		Task 4	
CA level	All red responses	Adjusted to cue change	All red responses	Adjusted to cue change
6 years	11	15	1	25
7 years	2	24	0	26
8 years	5	25	1	29

slides, while 8% of the 7-year-olds and 16% of the 8-year-olds responded in this fashion. Table 4 represents these data.

Since the clue slide for the closed green boxes never changed, the proper response to the all green test slide was all green balls. The analysis of the all green test slides showed an interesting trend. As Table 5 shows, a significant proportion of subjects did not respond properly to that slide. This type of error might be called overgeneralization of cue uncertainty, since subjects adjusted their responses to all red and all green test slides in parallel fashion. The difference between the 6- and 8-year-olds is significant $(X^2 = 4.71, df = 1, p < .05)$, the older subjects performing poorer than the younger on Series 3. The difference between performances for the 8-year-olds on Series 3 and 4 is significant $(X^2 = 7.12, df = 1, p < .02)$. This overgeneralization is not simply associated with attending to the uncertainty of the red clue slide, for the 7-year-olds were able to attend to the all red case as well as the 8-year-olds but did not exhibit the same pattern of overgeneralization. A possible explanation is that 6-year-olds did not attend to the uncertainty at all, while the 7- and 8-year-olds did. The difference between the 7- and 8-year-olds might be that the CA 8 focused on just the red colored clue slides, thereby not attending to the all green clue slides. These hypotheses cannot be unequivocally evaluated from the present data.

TABLE 5

PROPORTION OF OVERGENERALIZED RESPONSES ON ALL GREEN TEST SLIDES FOR EACH CA LEVEL ON TASKS 3 AND 4

	Task 3		Task 4	
CA level	Over-generalized	Correct	Over-generalized	Correct
6 years	4	22	5	21
7 years	6	20	8	18
8 years	12	18	3	27

If a subject understood the elements of the experimental task and attended to those elements, responses to the 1.0 and .0 slides in Series 1 and 3 should be perfect. In Series 3 and 4, the same slides provide us with more information about whether the subject is attending to the uncertain nature of the clue, than his ability to revise his estimates. Including both the 1.0 and .0 test slides in an analysis of estimate-revision performance might be inappropriate.

After removing the 1.0 and .0 test trials from the analysis of error, performance on the remaining slides shows a significant difference between the 6- and 8-year-olds ($t = 2.36$, $p < .05$). When the all red and all green slides are included in the analysis no developmental trends appear. It would seem that the overgeneralization error, found primarily in the 8-year-olds, and not found in the 6-year-olds, would have obscured the developmental differences.

It was hypothesized that attention would be a major variable in the developmental differences. The preceding analyses of the all red and all green response trials have shown to what degree and in what direction these differences tend. In the first study, it was found that the degree of attention focussing could explain some of the developmental differences in response distributions. In this study we have a similar finding in that attention focussing is able to account for the patterns of errors across the age levels. Corroborative evidence was found in another study. Vurpillot (1968) studied eye scanning of picture stimuli and task performance in children at four age levels from 3–9 years. It was found that the subjects under 6 used limited scanning and made inappropriate judgments, while older children adopted a task-appropriate eye-scanning strategy and their performance was also better.

Conservatism. Conservatism was measured in the same manner as in Experiment I. The only significant difference between age levels was that 7-year-olds were significantly less conservative than 8-year-olds ($X^2 = 3.95$, $df = 1$, $p < .01$). Irrespective of age group, a significant difference was found between high and low performers on Series 2 responses ($X^2 = 3.68$, $df = 1$, $p < .01$). When the same procedure of using the median as the cut-off point for high and low performers was used for the third and fourth series, no differences were found. However, if we use the same absolute value for the cut-off as was used in Experiment I (.110) the difference between the degree of conservatism for the performance levels is significant ($X^2 = 21.04$, $df = 1$, $p < .001$). This finding suggests that a subject could be conservative on one task and not so on another. In other words, conservatism seems to be related to the subjects' understanding of the task and less to some type of pervasive cognitive style. Subsequent testing is needed to substantiate these hypotheses.

Adjustment of estimates. Instead of determining the absolute error in judgment for the final two series, an analysis of whether the estimate of a subject was changed in the correct direction was undertaken. For example, if the subject's responses for the .25 test slides in Series 2 averages .35, then any response for that subject that is less than .35 is scored correct. The scores on Series 3 were used as a basis for evaluating the correctness of change for a response on Series 4. This method of analysis allows a subject who has a high average error to score well despite that average error value. In other words, it tests consistency of change relative to a subject's own estimates.

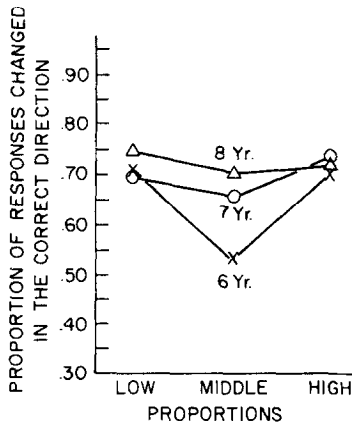The data are depicted for Task 4 in Fig. 4. There were no significant

FIG. 4. Proportion of responses altered in the correct direction as a function of slide type for Task 4.

differences with regard to age for Task 3 or 4. Any differences between slide types might be difficult to interpret because of the possible effect of response conservatism. More interesting than any differences between the groups is the fact that all age levels adjust their estimates in the correct direction at a level significantly above chance. In other words, although subjects might not extract all of the information from their clue slides, they are able to alter their estimates in the correct direction when faced with imperfectly weighted cues.

CONCLUSIONS AND IMPLICATIONS

It was hypothesized that children in the age range tested possessed a better than chance level of functioning on estimation and estimation-based inferences. This hypothesis was confirmed. It was further hypothesized that the CA range represented a period of change with respect

to the degree that subjects adhered to optimal models. This prediction was confirmed in the case of estimations and inferences about estimations when differences due to attention only were controlled. The hypothesis that significant changes would occur before the age of 7 was not conclusively confirmed in the case of estimations. However, the performance of the 6-year-olds on the inference task showed that the ability to adjust inferences or estimates must develop before 6. This finding casts some doubt on the position that young children are not capable of manipulating numerical quantities in an abstract form. Further, it is usually advanced that spatial reorganization interferes with the ability of a child to employ one-to-one correspondence and to exhibit conservation of quantity. In all of the estimation tasks subjects were required to make their responses via a highly ordered configuration, although the stimulus sets were presented in random arrays. The requirement of representational reorganization did not interfere with the ability of subjects to perform on the one-to-one correspondence task.

One is tempted to relate these findings to certain aspects of Piaget's theory. Flavell (1963) points out that Piaget dealt with the development of probability concepts in a series of experiments (p. 341). The conclusion is that preoperational children do not seem to differentiate between chance and nonchance or between the possible and the necessary. Although our study did not include an evaluation of the children's development in terms of concrete operations, their CA would suggest they were not fully at this level; yet their performance indicated they could make this differentiation between the possible and the necessary. Since the relevant Piagetian issues are complicated and only marginally related to the major issue here, the matter will not be pressed further.

Two elements of cognitive style were suggested by the first experiment and found to be of importance in the second experiment. The first was attention strategy, while the second was response conservatism. Both factors seem to warrant continued research. Some specific hypotheses that were suggested by this study were: (1) response conservatism is a function of the relation between the subject and his ability to cope with environmental tasks, and not a pervasive style or trait; and (2) when presented proportion estimation tasks there is a developmental trend towards translating the problem into a numerosity judgment.

Some problems in the analysis of quantitative estimation data were indicated. First, it cannot be assumed that the response distributions of young children can be equated. Secondly, it cannot be assumed that since a stimulus set is varying only in some quantitative dimension, responses will vary likewise. There can be significant qualitative shifts in response strategies with quantitative shifts in stimuli.

There is ample evidence in this study that young children are able to deal with uncertain situations requiring extralogical abilities. These techniques are assumed to develop because they are needed. However, at this point we have very little information about the limitations of instructing children in such techniques, or even about the degree to which such abilities are possessed at each age level. These issues are relevant to many forms of judgment processes, and should receive additional research.

REFERENCES

BRUNER, J., & RODRIQUEZ, J. Some determinants of apparent size. *Journal of Abnormal and Social Psychology*, 1953, **48**, 17–24.

BRUNSWIK, E. Perceptual characteristics of schematized human figures. *Psychological Bulletin*, 1939, **36**, 553. (Abstract)

BRUNSWIK, E. *Systematic and representative design of psychological experiments: with results in physical and social perception.* Berkeley: University of California Press, 1947.

BRUNSWICK, E. In defense of probabilistic functionalism: a reply. *Psychology Review*, 1955, **62**, 236–242.

BRUNSWIK, E., & HERMA, H. Probability learning of perceptual cues in the establishment of a weight illusion. *Journal of Experimental Psychology*, 1951, **41**, 281–290.

DIETZE, D. The facilitating effect of words on discrimination and generalization. *Journal of Experimental Psychology*, 1955, **50**, 255–260.

EDWARDS, W. The theory of decision making. *Psychological Bulletin*, 1954, **51**, 380–417.

EDWARDS, W., LINDMAN, H., & PHILLIPS, L. Emerging technologies for making decisions. In *New directions in psychology II*. New York: Holt, Rinehart, and Winston, 1965. Pp. 265–325.

FLAVELL, J. *The developmental psychology of Jean Piaget.* Princeton: Van Nostrand, 1963.

KENDLER, H., & KENDLER, T. Effect of verbalizations on reversal shifts in children. *Science*, 1961, **134**, 1619–1620.

LENNEBERG, E. Probabilistic approach to language learning. *Behavioral Science*, 1957, **2**(1), 1–12.

LURIA, A. *The role of speech in the regulation of normal and abnormal behavior.* New York: Liveright, 1961.

MANDLER, G. Verbal learning. In *New directions in psychology III*. New York: Holt, Rinehart, and Winston, 1967. Pp. 3–50.

PETERSON, C., HAMMOND, K., & SUMMERS, D. Multiple probability learning with shifting cue weights. *American Journal of Psychology*, 1965, **78**, 660–663.

POLYA, G. *Patterns of plausible inference.* Princeton: University Press, 1954.

POLYANI, M. Logic and psychology. *American Psychologist*, 1968, **23**(1), 27–43.

REESE, H. Verbal mediation as a function of age level. *Psychological Bulletin*, 1962, **59**, 502–509.

SWETS, J., TANNER, S., & BIRDSALL, T. The evidence for a decision-making theory of visual detection. Technical Report No. 40, 1960, Electronic Defense Group, University of Michigan.

TAJFEL, H. Classification and quantitative judgements. *British Journal of Psychology,* 1963, 54(2), 101–114.

TANNER, S. Statistical decision processes in retention and recognition. Report for the Sensory Intelligence Laboratory, University of Michigan, 1965.

TANNER, S., SWETS, J., & GREEN, D. Some general properties of the hearing mechanism. Technical Report No. 30, 1956, Electronic Defense Group, University of Michigan.

VURPILLOT, E. The development of scanning strategies and their relation to visual differentiation. *Journal of Experimental Child Psychology,* 1968, 6, 632–650.

WALLACE, R. Apparent personality traits from photographs varied in bodily proportions. *Psychological Bulletin,* 1941, 38, 744–745.

WHITE, S. Evidence for hierarchical arrangement of learning processes. In *Advances in child development and behavior,* Vol. 2. New York: Academic Press, 1965. Pp. 187–220.