SCHOLASTIC APTITUDE TEST MATHEMATICS SUBSCORES AS A SUBJECT VARIABLE IN LEVINE'S BLANK PROBE METHOD OF INVESTIGATING HYPOTHESIS FORMATION

A Thesis

Presented to

the Faculty of the Department of Psychology

Appalachian State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Price Coleman Locke

December 1976

SCHOLASTIC APTITUDE TEST MATHEMATICS SUBSCORES
AS A SUBJECT VARIABLE IN LEVINE'S BLANK PROBE
METHOD OF INVESTIGATING HYPOTHESIS FORMATION

BY

Price Coleman Locke

Approved by:

Frank R. Terrant Jr.
Chairman, Thesis Committee

[signature]
Professor of Psychology

Russell E. Walls
Associate Professor of Psychology

Joyce G. Crouch
Chairperson, Department of Psychology

Richard H. Rupp
Dean of the Graduate School

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Under the assumption that Marvin Levine's blank probe method in hypothesis-testing theory is a viable one, this study attempted to use that method to determine significant differences in the process of learning via hypothesis formation. As a basis of differentiation, 51 undergraduates were divided into three groups corresponding to high, medium, and low scores on the mathematics subtest of the Scholastic Aptitude Test (SAT). Subjects having high SAT scores were expected to show the most systematic problem-solving behavior and those having low scores the least. Little pretraining was given, and a typical Levine configuration of eight problems was given to each subject. In addition to Levin's usual categorization of responses (as being hypothesis-congruous or not), verbal reports were taken from the subject after each blank probe.

No statistically significant results were obtained, thus indicating that performance on the experimental task is independent of performance on the SAT. However, refinement of the data, utilizing the subjects' verbal reports, revealed that the high SAT group showed the most systematic problem-solving solving behavior and that the middle SAT group showed the least systematic problem-solving behavior. Since that degree of systematic behavior which facilitates performance in a task such as taking a Scholastic Aptitude Mathematics Test was not analogously reflected in the performance on the simpler Levine tasks, it was concluded that the mathematics subtest of the SAT is not an efficient measure to use in investigating Levine's blank probe method of hypothesis-testing.

CHAPTER I

INTRODUCTION

In recent years hypothesis-testing theory has been the focus of much research in the area of concept learning. The basic premise of such theory is that the human actively selects and evaluates various hypotheses (Restle, 1962; Levine, 1963) in the formation of a concept. The population of hypotheses which a learner (or subject) samples consists of both relevant and irrelevant hypotheses, and initially the learner simply picks a hypothesis at random from that population. If it so happens that he chooses an irrelevant hypothesis, this choice eventually leads to an incorrect response, which causes the learner to discard that particular hypothesis and choose another. The process continues until correct responses are made repeatedly and consistently under some hypothesis. As an illustration of this process, consider the rectangles shown under "STIMULI" in Figure 1. If a rectangle represents a card, then that card has two characters thereon, and a given character is 1) either an "X" or a "T", 2) either on the left or on the right, 3) either colored or noncolored, and 4) either large or small. Assume that by some prearranged rule exactly one of the two characters per card is a correct choice and that the other character is an incorrect choice (one such rule is that the large character is always correct); this rule, then, is the concept to be learned.

STIMULUS CARDS AND HYPOTHESES

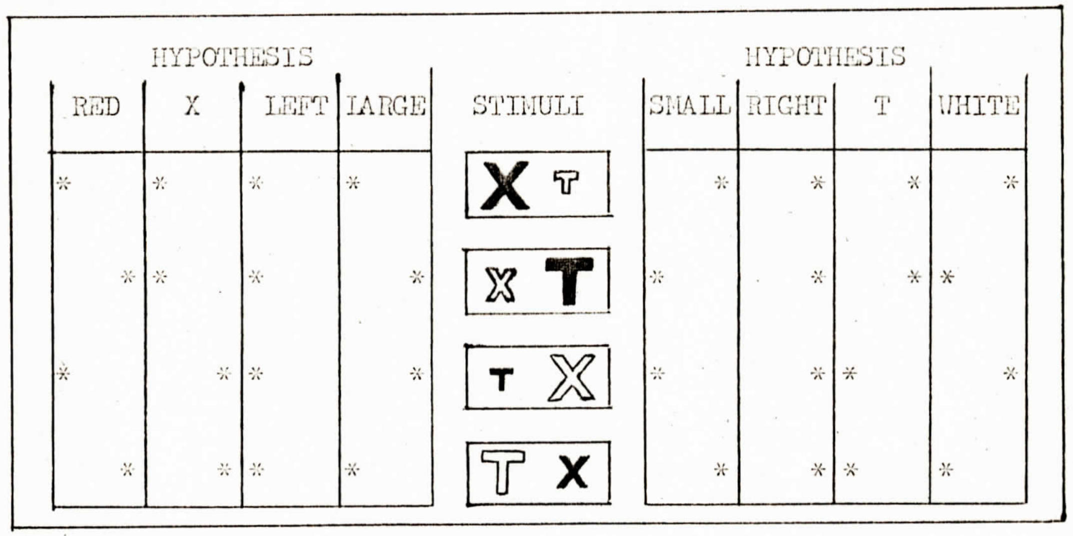

Figure 1

If a subject is presented with a sequence of such cards and is given the task of ascertaining the rule governing which of the two characters is correct on a given card, hypothesis-testing theory predicts that he will formulate a hypothesis (such as, the colored character is always correct) and then choose characters according to that hypothesis until he makes an incorrect choice. The subject will then start using another hypothesis, and the process continues until the subject repeatedly and consistently makes the correct choice, thereby indicating that he is working under the proper hypothesis. Note in Figure 1 that eight hypotheses are shown, four in each "HYPOTHESIS" column. The four asterisks in each of the eight columns simply indicate which side of the card a subject picks. For example, if for the indicated four card sequence a subject chooses the left character on the first and fourth cards and the right character on the second and third cards, it can be inferred that the subject was operating under the hypothesis that the large character is always the correct choice.

A broad overview of hypothesis-testing theory was given by Brown (1974); the fact that 14 of the article's 62 references include Marvin Levine as an author indicate the degree of his involvement in this area. There are several underlying assumptions which are central to Levin's theory; a brief summary will be given here. First, the subject selects a hypothesis from some set of possible hypotheses at the beginning of a trial. For instance, the subject may predict that the stimulus on the left will always be the correct one and choose accordingly. Second, the set of hypotheses from which the subject samples is finite and is known to the experimenter. In the case of the stimuli shown in Figure 1, there

exist exactly eight hypotheses corresponding to simple attributes of each dimension; these are shown in the figure's eight columns. Third, if no feedback (i.e., "right" or "wrong") is given after the subject makes a choice on a trial, the subject maintains the same hypothesis during the next trial. Consequently only one hypothesis will be retained by the subject during a sequence of trials in which no feedback is given to the subject. (Such a series of no-feedback trials is referred to as a "blank-trials" probe, and will be clarified in a later portion of this section.) The validity of this assumption has been shown by Levine, Leitenberg, and Richter (1964) and Levine (1966). Fourth, the pattern of responses during a blank probe will be perfectly correlated with the aspect of the stimulus corresponding to the hypothesis held. Thus if a particular response pattern is uniquely associated with a particular attribute, it can be inferred that the subject's hypothesis held that attribute to be the correct one. Such unique response patterns are shown in the eight "HYPOTHESIS" columns of Figure 1, and are referred to as hypothesis-congruous response patterns. Fifth, there is a very small probability that the subject will erroneously pick a stimulus which is not consistent with his hypothesis held on any given trial. If this does happen, the resultant response will, of course, be inconsistent with any hypothesis of the type illustrated in Figure 1. Such response patterns can be categorized as not being congruous with hypothesis formation, or simply non-congruous.

Although Levine has recently utilized verbal reports by his subjects as a means of revealing hypothesis-forming behavior (Phillips and Levine, 1975), he is most noted for his development and use of the previously mentioned blank-trials probe method (1963, 1966). A typical Levine problem utilized 16 trials (a trial is the presentation of one card) with feedback given to the subject only on the first trial and every fifth trial thereafter. Each of the resultant four-trial sequences without feedback therein was a blank probe. If the response pattern shown by the subject during a blank probe was hypothesis-congruous, the probe was categorized relative to the eight hypotheses shown in Figure 1; otherwise it was simply labeled non-congruous. The discrimination learning which he investigated always involved multidimensional stimuli such as those shown in Figure 1; one of the two stimuli on a card was always correct during a given problem.

Levine has spent time researching the validity and usefulness of his blank probe method. For instance, the blank probe's predictive power is indicated by the fact that the probability of a subject's showing the same hypothesis during a blank probe after being correct on the previous blank probe is .95 (Levine, 1966). Similarly, the probability of a subject's showing the same hypothesis after being wrong is .02. Also, Levine has consistently found a remarkably high total percentage of hypothesis-congruous responses to blank probes in his experiments; he has obtained percentages ranging from 92.4% (1966) to 95.4% (Levine, Miller, and Steinmeyer, 1967). Other experimenters, however, report somewhat lower values, although the difference may conceivably be attributed to amount and intensity of pretraining and experimental design. For instance, Wells (1972) and Coltheart (1971) obtained values of 72% and 76%, respectively, in their studies. Both experimenters' designs differed from Levine's in that their four-dimension stimuli were presented singly (successive discrimination) rather than in pairs (simultaneous discrimination). Wells

utilized only two problems (as opposed to Levine's usual 16), and Coltheart presented only one problem.

Several developmental studies have been made with blank probes. Eimas (1969, 1970) and Ingalls and Dickerson (1969) consistently found that performance (as measured by percentage of hypothesis-congruous responses) increased with age. The typical experiment compared two or three age levels in children with university students used as controls. For example, in his four-dimension study Eimas (1969) obtained a value of 88%; four preliminary practice problems were given. An eight-dimension experiment (Eimas, 1970) yielded a Levine-like figure of 92%; eight practice problems were given. Also, Ingalls and Dickerson (1969), in a four-dimension study with six practice problems, found that 89% of the blank probe responses were consistent with hypothesis formation. In support of Levin's 1966 paper they found the probability of a hypothesis being repeated was .95 when the feedback immediately preceding was "right" and .02 when the feedback preceding was "wrong."

Virtually no work has been done to investigate the effect of non-developmental subject variables on performance during blank-trials probes. It may be that one or more subject traits influence performance during the blank probes. Furthermore, such a trait might be of importance in general problem-solving. In particular, by the very nature of the task, those subjects who tend to be more systematic and logical in their thinking may perform substantially better than other subjects. This possibility has not been investigated and is the focus of the present study.

Similarly no results have been discussed to this point relative to blank probe performance in successive problems for a given subject, although there is data to support the notion that hypotheses are shown by the subject with greater probability during successive trials of any one given problem (Levine, 1969; Frankel, Levine, and Karpf, 1970). It seems reasonable to assume that a learning set may influence performance, and that the high percentages of hypothesis-congruous responses reported by Levine (relative to Wells and others) may well be attributable to the extensive pretraining which is incorporated in Levine's procedure.

The Problem

The intent of this study was to utilize Levine's 1966 experimental procedure to investigate the influence of a subject trait on hypothesis-congruous responses during blank-trials probes. The primary method of this study was to compare the hypothesis-formation behavior of three groups formed as strata (high, middle, and low) on the basis of Scholastic Aptitude Test (SAT) mathematics subscores. The SAT score was taken to be a measure of a subject's analyticity, for according to P. H. Dubois in Buros' The Seventh Mental Measurements Yearbook (1972), ". . . the mathematical material [of the SAT] now depends less on formal knowledge and more on 'logical reasoning and . . . the perception of mathematical relationships.'" Those subjects having high SAT scores were expected to show the highest percentage of hypothesis-congruous responses in a series of eight experimental problems.

CHAPTER II

METHOD

Subjects

The subjects for this study consisted of 58 volunteer undergraduate students who were selected from classes at Appalachian State University, and who were ranked according to their scores on the mathematics section of the Scholastic Aptitude Test (SAT). Due to subjects' being at various academic ranks there was unavoidable variation in the time the SAT had been administered (generally, the subjects had taken it during their senior year in high school). Seven subjects were discarded due to the unavailability of official test score reports; as a result three groups of 17 subjects each (high, middle, and low) were formed. The range in SAT scores for each group was as follows: the high group ranged from 530 to 690, the middle group ranged from 450 to 520, and the low group ranged from 320 to 450. For the entire group the median was 500, the mean was 493.6, and the standard deviation was 81.7.

Apparatus

Each subject received a four-trial pretraining problem (with feedback given after each trial) and then eight experimental problems, all in one session. As shown in Figure 1, the discrimination involved four dimensions (position, shape, color, and size) with two values within dimensions (left or right, X or T, red or white, and large or small, respectively). The blank probe stimulus cards were an internally orthogonal set in which values of exactly two of the four dimensions changed from trial to trial, so that the subject's four consecutive responses were an indication of the hypothesis held. Two such internally orthogonal sets exist. The 24 permutations of the set shown in Figure 1 were utilized in the blank probes (one set per probe), and nine distinct permutations of the complementary set were used in the feedback and pretraining trials. Presentation of stimulus cards for any given problem was made in accordance with Table 1, with the subject receiving feedback after his or her response to cards in trials 1, 6, 11, and 16 of a single concept problem. The actual testing was carried out in a secluded cubicle with minimal distractions. Each subject was seated at a table across from the experimenter, and background information was solicited from the subject immediately prior to the testing session.

Design and Procedure

Each subject performed on eight consecutive problems, each containing three blank trial probes. For each subject, the procedure was the same, except for randomization of the feedback. The feedback sequence in a given problem was predetermined by random selection and was independent of the response pattern shown by the subject. The feedback on trial sixteen in a given problem was always "right" and the sequence of feedback for trials one, six, and eleven in a given problem was one of the eight permutations using "right" and "wrong" as responses. Each subject received each permutation exactly once per session; the order of presentation was randomized. Furthermore each of the 24 permutations of the four stimulus cards in a blank probe was presented to the subject exactly once per session.

Table 1

Trial Presentation

Trial	Category
1	Feedback
2–5	Blank Probe
6	Feedback
7–10	Blank Probe
11	Feedback
12–15	Blank Probe
16	Feedback

At the beginning of each session the subject was shown a few stimulus cards and was read the following instructions:

> In this experiment you will be presented with several easy problems. Each problem consists of a series of cards like these. Each card will always contain two letters, and the letters will be of two colors. You will also notice that the letters are of two different sizes, and of course, that one letter is on the left and one is on the right. Every card will be like these except that the letters and colors will be changed around. One of the two letters on each card is correct in the sense that I've marked it here on my sheet. For each card I want you to tell me which of the two letters you think is correct and I'll tell you whether you're right or wrong. Then you go on to the next card, again you make a choice, and again I'll tell you whether you are right or wrong. As we continue in this way you can learn the basis for my saying right or wrong. The object for you is to figure this out as fast as possible so that you can choose correctly as often as possible.

The subject was then given the pretraining problem, which consisted of four trials (i.e., four separate card presentations). After this pretraining, the subject was instructed as follows:

> In the last problem I said right or wrong after each card. For the remaining problems I won't always tell you whether you are right or wrong. After most cards I'll say nothing. But don't let that disturb you; still try to be right all the time.

The eight experimental problems were then presented and the subject's choice for each trial was marked on a data sheet by the experimenter. Also recorded was the subject's response to being asked at the end of each probe what particular approach (if any) he or she had taken to the problem during that probe.

CHAPTER III

RESULTS

Figure 2 shows the relationship between subjects' SAT mathematics scores and their total number of hypothesis-congruent responses (of a possible total of 24). In this case the SAT score is treated as the dependent variable Y so that a linear regression model for the data ($Y = a + bX$) might be investigated. Note that there is no immediate effect to be seen in the graph; the data points are spread out with no obvious pattern. The line of regression was calculated to be $Y = 504.6 - 3.7613X$ ($\underline{r} = -.094$). However, as might have been expected, the null hypothesis that $b = 0$ was accepted ($\underline{F} = .45$, $\underline{df} = 1, 49$; see Hoel, Port, and Stone, pp. 152-153). The consequent inference is that the two measures used are independent of one another.

This lack of a statistically significant relationship between SAT scores and percentage of hypothesis-congruous responses is also shown by an analysis of variance in which the dependent variable used is the number of hypothesis-congruous response sequences shown by the subject during a given problem. Thus each subject received a score of zero or one for each blank probe and therefore a score of zero to three, inclusive, for an entire problem. Cell, row, and column means are given in Table 2. There were no statistically significant results, although the mean number of hypothesis-congruous responses tended to increase as the subject progressed through the problems. Table 3 summarizes the results of this two-factor

SAT Scores vs. Number of Hypothesis-Congruous Responses

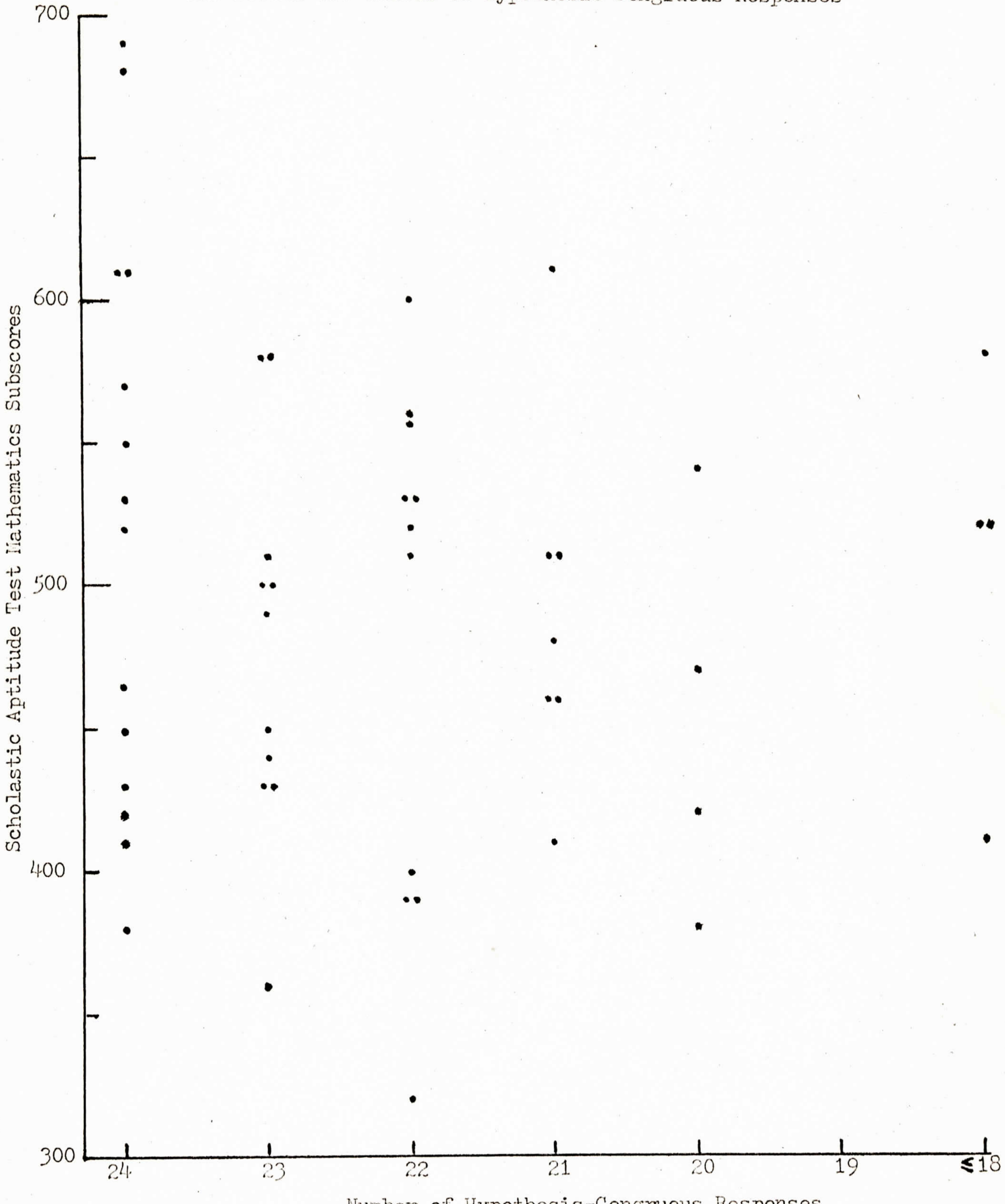

Figure 2

Table 2

Cell, Row, and Column Means

for Analysis of Variance in Table 3

Analyticity	1	2	3	4	5	6	7	8	$\bar{X}_{.j.}$
Low	2.647	2.647	2.765	2.765	2.647	2.882	3.000	2.882	2.779
Medium	2.529	2.765	2.765	2.588	2.765	2.765	2.588	2.706	2.684
High	2.765	2.765	2.765	2.941	2.706	2.882	2.706	3.000	2.816
$\bar{X}_{..k}$	2.647	2.725	2.765	2.765	2.706	2.843	2.765	2.863	$\bar{X}_{...}$ = 2.760

Note: j denotes the jth row
k denotes the kth column

Table 3

Analysis of Variance of Number of Hypothesis-Congruous Responses for the Eight Problems

Source	df	Mean Square	F
Between subjects	50		
Analyticity Groups (A)	2	.6348	1.3139
Subjects within groups	48	.4831	
Within subjects	357		
Problems (B)	7	.2507	.9652
A x B	14	.2118	.8155
B x Subjects within groups	336	.2597	

analysis of three levels of SAT scores, the between-subjects variable (17 subjects per cell), and eight levels of problems, the within-subjects variable. None of the F ratios are significant, indicating that there is no difference between the analyticity groups, no difference across problems, and no interaction between the two.

However, an interesting effect is noted in investigating the frequency of subjects (relative to the three analyticity groups) above and below the median number of hypothesis-congruous responses made by each subject. By means of a Median Test, these data (Table 4) seem to indicate that the high analyticity group performs the best and that the middle analyticity group performs the worst ($\chi^2 = 5.86$, $p < .06$). Note that whereas the low analyticity group is divided equally about the median, there are striking imbalances in the division of the other two groups.

As a consequence of the foregoing results the data were reanalyzed in an attempt to ascertain what effects, if any, SAT scores might have on performance in the task. Accordingly, the subjects' verbal reports were used in categorizing responses, although the categories of hypothesis-congruous (HC) and non-congruous (NHC) as defined and used by Levine were retained. These latter two response categories were therefore subdivided into the following three subcategories, based on the subjects' accompanying verbal reports: 1) responses which were described as following from one of the eight basic (B) hypotheses investigated by Levine, 2) responses which were described as following from no (N) rule or guessing, and 3) responses which were described as following from a hypothesis different from (and often more complex (C) than) the basic eight. This subdivision of verbal responses associated with HC and NHC response patterns is summarized in Table 5.

Table 4

Obtained and Expected Frequencies for

Six Categories of Analyticity Effects

Analyticity	Above Median	Below Median
Low	O = 8	O = 9
	E = 7.67	E = 9.33
Medium	O = 4	O = 13
	E = 7.67	E = 9.33
High	O = 11	O = 6
	E = 7.67	E = 9.33

$\chi^2 = 5.86$

$p < .10$

Note: $p = .05$ $\chi^2 = 5.99$

Table 5

Frequency of Verbal Responses Associated with Hypothesis-Congruous and Noncongruous Choice Patterns

Low Analyticity						Medium Analyticity						High Analyticity					
Hyypothesis-Congruous Responses			Non-Congruous Responses			Hypothesis-Congruous Responses			Non-Congruous Responses			Hypothesis-Congruous Responses			Non-Congruous Responses		
B	N	C	B	N	C	B	N	C	B	N	C	B	N	C	B	N	C
24	0	0	0	0	0	24	0	0	0	0	0	24	0	0	0	0	0
24	0	0	0	0	0	24	0	0	0	0	0	24	0	0	0	0	0
24	0	0	0	0	0	23	0	0	0	0	1	24	0	0	0	0	0
24	0	0	0	0	0	22	1	0	1	0	0	24	0	0	0	0	0
23	0	0	0	1	0	21	0	0	3	0	0	24	0	0	0	0	0
23	0	0	1	0	0	21	0	0	3	0	0	24	0	0	0	0	0
22	0	0	1	1	0	21	0	2	0	1	0	23	0	0	0	0	1
22	0	0	2	0	0	21	0	0	3	0	0	23	0	0	1	0	0
21	1	0	0	1	1	20	3	0	1	0	0	22	0	0	1	1	0
19	0	4	1	0	1	20	0	0	2	0	2	22	0	0	2	0	0
19	0	5	0	0	0	16	0	5	0	1	2	22	0	0	1	0	1
17	0	6	0	0	1	16	1	5	0	0	2	21	0	1	0	0	2
16	4	0	1	0	3	15	0	7	0	1	1	18	0	0	6	0	0
16	0	5	0	0	3	14	1	6	0	2	1	17	0	4	1	0	2
15	3	4	0	1	1	14	0	9	1	0	0	16	4	4	0	0	0
13	0	7	0	0	4	14	0	1	1	3	5	15	0	7	0	1	1
9	1	7	3	0	4	12	0	6	0	4	2	12	0	8	1	0	3

Note: B denotes Basic N denotes No Rule C denotes Complex

Ignoring SAT groups, an interesting effect is observed. An error may be defined as any response which was not both hypothesis-congruous (HC) and concomitantly reported as hypothesis-contruous (B). These errors were, of course, etiher HC or NHC. For a given subject, let G denote the category of having made more HC errors than NHC errors and let L denote the category of having made fewer HC errors than NHC errors (or the same number). Table 6 presents the frequency of subjects (relative to the categories of G and L) in two error frequency groups. Note that subjects who made few errors (zero to three, inclusive) tended to make more NHC errors, while subjects who made several or many errors (four or more) tended to make more HC errors ($\chi^2 = 27.0$, $p < .001$).

Table 7 presents the percentage of errors that were HC and NHC, respectively, for four error frequency groups. Note that the effect just mentioned is also observed here. As the total number of errors made increases, the type of error made tends to swing from being NHC to being HC.

Investigation of this data array also revealed variation in the approach to the problems between the three analyticity groups with consequent differences in performance. In general, the high analyticity group tended to show the most hypothesis-congruous responses and the medium group the least. Such behavior is evident, for instance, on consideration of only those hypothesis-congruous responses (HC) which the subject concomitantly reported as following from the subject's holding one of the basic eight hypotheses (B). Table 8 shows that whereas the low analyticity group is divided equally, there are imbalances in the division of the other two groups.

Table 6

Obtained and Expected Frequencies for Four Categories of Number of Total Errors and Type of Error Predominating

Total Errors	L	G
0-3	O = 28 E = 19.33	O = 1 E = 9.67
4	O = 6 E = 14.67	O = 16 E = 7.33

$\chi^2 = 27.0$, $p < .001$

Note: G denotes category of having made more hypothesis-congruous errors than non-congruous errors

L denotes category of having made fewer hypothesis-congruous errors than non-congruous errors (or the same number)

Table 7

Distribution of Type of Error Made by

Subjects Grouped According to Total Errors

Total Errors	Percentage of Errors Which Were: HC	NHC
0-2	6	94
3-5	43	57
6-8	63	37
>8	62	38

Table 8

Obtained and Expected Frequencies for Six Categories of

Analyticity and Number of Hypothesis-Congruous

Responses Also Verbally Hypothesis-Congruous

Analyticity	Number of Hypothesis-Congruous Responses Also Verbally Hypothesis-Congruous: 22-24	< 22
Low	O=8 E = 7.67	O = 9 E = 9.33
Medium	O = 4 E = 7.67	O = 13 E = 9.33
High	O = 11 E = 7.67	O = 6 E = 9.33

$\chi^2 = 5.86$

$p < .10$

Note: $p = .05$ $\chi^2 = 5.99$

There are other indications of the high/low/medium analyticity groups (in that order) showing progressively decreasing degrees of dealing efficiently and systematically with the problems. For example, that order is seen when ranking the three groups relative to frequency of subjects therein who showed (and at the same time reported) hypothesis-congruous responses for all the probes. A similar order is noted in the frequencies of subjects in each group who reported following no rule on one or more probes during the session. This order also holds in the total number of non-congruous responses shown by each group (NHC). These data are summarized in Table 9.

Yet another measure of the differences in approach between the three groups appears on consideration of which (if any) of the four basic characteristics (position, shape, color, and size) a subject investigated in the first probe of each problem. Each subject was categorized as to how many characteristics he or she thus investigated. Table 10 presents the frequency of subjects (relative to the three analyticity groups) in two such (cumulative) categories. These data tend to indicate that the high analyticity group was most diverse in its approach and that the medium group was least diverse.

Table 9

Measures Indicating Performance

Differences Relative to Analyticity

	Analyticity		
	Low	Medium	High
Percentage of Subjects Who Both Showed and Reported only Hypothesis-Congruous Responses	23.5	11.8	35.3
Percentage of Subjects Who Followed No Rule on One or More Probes	35.3	52.9	17.6
Total Number of Non-Congruous Responses over All Subjects	30	43	25

Table 10

Obtained and Expected Frequencies for Six Categories of Analyticity and Number of Characteristics Investigated at Least Once

Analyticity	Number of Characteristics Investigated at Least Once: 1-2	3-4
Low	O = 6 E = 6	O = 11 E = 11
Medium	O = 9 E = 6	O = 8 E = 11
High	O = 3 E = 6	O = 14 E = 11

$\chi^2 = 4.64$

$p < .10$

CHAPTER IV

DISCUSSION

The simple categorization utilized by Levine proved to be inadequate to reveal any differences in performance between the analyticity groups. This situation may stem largely from the fact that 92.0% of all responses in this study were hypothesis-congruous (by Levine's criteria), a somewhat unexpected occurrence due to the purposely limited pretraining given the subjects. It seemed that for some subjects the reactive task of verbalizing the bases of their approaches caused them to focus on the problems more than if they had not been required to conceptualize their responses.

Incorporating the subjects' verbal responses into the data array resulted in a refinement of the data which allowed differences between the three groups to emerge. Although the high analyticity group showed the most systematic behavior, the most striking aspect of the picture is the fact that the middle analyticity group showed the least systematic behavior. The reasons for this situation are not immediately clear.

However, there are obvious differences in the two tasks under consideration, i.e. the Scholastic Aptitude Test and the blank probes task utilized in this study. The Levine probe method is designed to detect general hypothesis-formation behavior and thus is simpler in nature than the problems on the Scholastic Aptitude Test. Consequently it might actually require less analytical ability (or even a different type) to perform well therein, which is why the low analyticity group (who might be described as simple rule-followers) did well. The medium analyticity group, on the other hand, suffered a decrement in systematic performance as measured by not recognizing the basic simplicity of the task, which may be accounted for by their failing to use their greater analytical ability appropriately, or perhaps their higher intellectual level allowed <u>complex</u> competing responses to occur. The high analyticity group, of course, appropriately channeled their ability and performed best of all. This model is shown graphically in Figure 3. Perhaps a little analyticity is a dangerous thing.

In conclusion, this study has failed to show that the subject variable in question (analyticity) has an effect on performance in Levin's blank probe method of investigating hypothesis formation. However, there was a difference in performance between the three groups (which was evident on refinement of the data) which indicates that analyticity of the subject interacts with the nature of the task. In the case at hand, that degree of analyticity which facilitates performance in a task such as taking a Scholastic Aptitude Mathematics Test was not analogously reflected in the performance of a task simpler in nature.

Future studies in this area will probably be more fruitful if the verbal report classification is refined somewhat. In particular the responses which were both HC and B made up the bulk of the responses, consequently limiting the subsequent degree of analysis. Perhaps increasing the number of dimensions and/or values would influence the variability of subjects' responses. More complex concept tasks, e.g. those involving conjuctive or disjunctive rules, may also be more sensitive to variations of analyticity among subjects. At the same time, the lack of clear-cut results in this study indicate that SAT scores as a measure of analyticity may not be the most appropriate instrument to use in investigating Levine's blank probe method of hypothesis-testing.

Performance Model for Simple Task-Solving

Figure 3

REFERENCES

REFERENCES

Brown, A. S. Examination of hypothesis-sampling theory. Psychological Bulletin, 1974, 81, 773-790.

Buros, O. K. (Ed.). The seventh mental measurements yearbook. Highland Park, N. J.: Gryphon Press, 1972.

Coltheart, V. Memory for stimuli and memory for hypotheses in concept identification. Journal of Experimental Psychology. 1971, 89, 102-108.

Eimas, P. D. A developmental study of hypothesis behavior and focusing. Journal of Experimental Child Psychology, 1969, 8, 160-172.

Eimas, P. D. Effects of memory aids on hypothesis behavior and focusing in young children and adults. Journal of Experimental Child Psychology, 1970, 10, 319-336.

Frankel, F., Levine, M., and Karpf, D. Human discrimination learning: A test of the blank trials assumption. Journal of Experimental Psychology, 1970, 85, 342-348.

Hoel, P. G., Port, S. C., and Stone, C. J. Introduction to statistical theory. Boston: Houghton Mifflin Company, 1971.

Ingalls, R. P., and Dickerson, D. J. Development of hypothesis behavior in human concept identification. Developmental Psychology, 1969, 1, 707-716.

Levine, M. Mediating processes in humans at the outset of discrimination learning. Psychological Review, 1963, 70, 254-276.

Levine, M. Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 1966, 71, 331-338.

Levine, M. Neo-noncontinuity theory. In G. H. Bower and J. T. Spence (Eds.), The psychology of learning and motivation. Vol. 3. New York: Academic Press, 1969.

Levine, M., Leitenberg, H., and Richter, M. The blank trials law: The equivalence of positive reinforcement and nonreinforcement. Psychological Review, 1964, 71, 94-103.

Levine, M., Miller, P., and Steinmeyer, C. H. The none-to-all theorem of human discrimination learning. Journal of Experimental Psychology, 1967, 73, 568-573.

Phillips, S., and Levine, M. Probing for hypotheses with adults and children: Blank trials and introtacts. Journal of Experimental Psychology: General, 1975, 104, 327-354.

Restle, F. The selection of strategies in cue learning. Psychological Review, 1962, 69, 329-343.

Wells, H. Effect of intertrial interval duration on component processes in concept learning. Journal of Experimental Psychology, 1972, 94, 49-51.

APPENDIX

APPENDIX

The experimental data for each subject are summarized below. A subject received a score of 1 if his response during a blank probe was hypothesis-congruous or 0 if the response was non-congruous, and each of the eight columns represents one problem presented to the subject. Each row presents the data for one subject, and the three analyticity groups are labeled.

Low Analyticity Group

SAT Score	Problem Number I	II	III	IV	V	VI	VII	VIII
320	1 1 1	0 1 1	1 1 1	1 0 1	1 1 1	1 1 1	1 1 1	1 1 1
360	1 1 1	1 1 1	1 1 1	0 1 1	1 1 1	1 1 1	1 1 1	1 1 1
380	1 1 0	1 1 0	1 1 0	1 0 1	1 1 1	1 1 1	1 1 1	1 1 1
380	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
390	1 1 1	1 1 1	1 1 0	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0
390	1 1 1	1 1 0	1 1 1	0 1 1	1 1 1	1 1 1	1 1 1	1 1 1
400	0 1 1	1 1 1	1 1 1	1 1 1	1 1 0	1 1 1	1 1 1	1 1 1
410	0 0 0	1 0 1	1 1 1	1 1 1	1 0 0	1 1 1	1 1 1	0 1 1
410	1 1 1	1 1 1	1 0 0	1 1 1	1 1 1	1 1 0	1 1 1	1 1 1
410	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
420	1 1 1	0 1 1	1 1 1	1 1 1	1 0 0	0 1 1	1 1 1	1 1 1
420	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
430	1 1 1	0 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
430	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
430	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0	1 1 1	1 1 1	1 1 1
440	0 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
450	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1

Medium Analyticity Group

SAT Score	Problem Number I	II	III	IV	V	VI	VII	VIII
450	1 1 1	1 1 1	1 0 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
460	1 1 0	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0	1 0 1
460	1 1 1	1 1 1	1 1 1	0 0 1	0 1 1	1 1 1	1 1 1	1 1 1
464	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
470	0 0 1	1 1 0	1 1 1	1 1 1	1 1 1	1 1 1	1 0 1	1 1 1
480	1 1 1	1 1 1	1 1 1	1 0 0	1 1 1	0 1 1	1 1 1	1 1 1
490	1 1 1	1 1 1	1 1 1	1 1 1	0 1 1	1 1 1	1 1 1	1 1 1
500	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0	1 1 1	1 1 1	1 1 1
500	1 1 1	0 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
510	0 0 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 0 1	1 1 1
510	1 0 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	0 1 1
510	1 1 1	0 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 0 1	1 1 0
510	1 1 1	1 1 1	0 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
520	1 1 1	1 1 1	1 0 1	1 1 0	1 1 1	1 1 1	1 1 1	1 1 1
520	0 1 0	0 1 1	1 1 1	1 1 0	1 1 1	1 1 1	1 1 1	1 0 0
520	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
520	1 1 1	1 1 1	0 1 1	0 1 1	1 1 0	0 0 0	0 0 0	1 1 1

High Analyticity Group

SAT Score	I	II	III	IV	V	VI	VII	VIII
530	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 0 1	0 1 1	1 1 1
530	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
530	1 1 1	1 1 1	1 0 1	1 1 1	1 1 1	1 1 1	0 1 1	1 1 1
540	1 0 0	1 1 1	1 1 1	1 1 1	1 0 1	1 1 1	0 1 1	1 1 1
550	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
559	1 1 1	0 1 1	1 1 1	0 1 1	1 1 1	1 1 1	1 1 1	1 1 1
560	1 1 1	1 0 0	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
570	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
580	1 1 1	1 1 1	1 1 1	1 1 1	0 1 1	1 1 1	1 1 1	1 1 1
580	1 1 1	1 1 0	1 0 1	1 1 1	0 0 1	1 0 1	1 1 0	1 1 1
580	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	0 1 1	1 1 1
600	1 1 0	1 1 1	1 0 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
610	0 1 1	1 1 1	0 1 1	1 1 1	1 0 1	1 1 1	1 1 1	1 1 1
610	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
610	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
680	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
690	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1