1964

The Application of the Additivity Theorem to Subjective Probability

Branton K. Holmberg
Central Washington University

THE APPLICATION OF THE ADDITIVITY THEOREM TO SUBJECTIVE PROBABILITY

A Thesis

Presented to

the Graduate Faculty of

Central Washington State College

In Partial Fulfillment

of the Requirements for the Degree

Master of Education

by

Branton K. Holmberg

August 1964

APPROVED FOR THE GRADUATE FACULTY

Jack J. Crawford, COMMITTEE CHAIRMAN

Eldon E. Jacobsen

Persis T. Sturges

TABLE OF CONTENTS

CHAPTER I

THE PROBLEM AND DEFINITION OF TERMS

The purpose of this study is to examine the validity of applying the additivity theorem to subjective estimates of probability. Additivity, as used herein, refers to the process of adding the probability of the occurrence of an event and the probability of the non-occurrence of the same event and arriving at a total of <u>one</u>. The additivity theorem as it is described in this paper is derived from notions of mathematical probability similar to that of Peatman (11:150) which states: Let n be the number of 1) exhaustive, 2) mutually exclusive, and 3) equally likely cases of an event under a given set of conditions. If m of these cases are known as event A, then the probability of Event A under the given set of conditions is: $P(A) = \frac{m}{n}$ which if carried to its mathematically logical conclusion results in $P(A) + P(\bar{A}) = 1$.

From the mathematical probability concept described above, Savage (13:142) and Hunt (9:364) have developed the following additivity theorems which they feel are applicable to subjective probability. According to Savage, the additivity theorem asserts "the summed portions of exclusive, exhaustive categories of a set of events must total to one." Hunt describes his theorem as "the subjective probability of success is for the same person more or less inverse to the probability of failure, that is, probability of success plus the probability of failure equals one."

Subjective probability, as typically used by decision theorists, etc., refers to the psychological counterpart of mathematical probability, which, though typical, is by no means the wholly accepted definition of subjective probability. Thus, in order to clarify and conjoin the major concepts in this study, an exploration of the various meanings and interrelationships of probability, subjective probability, and the additivity theorem is necessary.

The term "probability" has a variety of meanings. Five major definitions are listed by Good (7:344-45) and are briefly summarized as follows: 1) Classical probability which is any definition of probability that utilized the aspect of equally probable cases; 2) Subjective probability (also referred to as personal or logical probability) which depends on the given information as well as on the event whose probability is to be estimated [P (E/F) which reads, the probability of E given F] ; 3) Physical probability which is the probability of a "success" given the "experimental setup"; 4) Inverse probability which proposes that the final probability of a hypothesis is proportional to its initial probability times its likelihood;

and 5) Neoclassical probability (Good attributes the creation of this definition to himself) which states that a theory of subjective probability is general enough to account for physical probabilities, but not conversely. Although a physical probability can be regarded as something that is not subjective, its numerical value can be equated to the limiting value of a subjective probability when experiments are repeated indefinitely under essentially constant circumstances.

The emphasis on subjective probability which predominates Good's descriptions is presently under scrutiny in the field of psychology. At this time, no consensus of an acceptable definition of subjective probability has been arrived at. Although definitions vary widely in specificity most have a common base which is mathematical (sometimes referred to as objective) probability. For example, Von Neumann and Morgenstern (16:83) propose that subjective probability is equal to objective probability; Howard (8:335) claims psychological (subjective) probability may be defined as perceived mathematical probability; Edwards (6:479) says subjective probability is the name for a transformation on the scale of mathematical probabilities which is somehow related to behavior; and Thrall, Coombs, and Davis (15:258) state that subjective probability refers to perceptions of relative frequency with objective frequencies implicit, or, in other words, subjective probabilities are psychological values for explicit objective probabilities.

Due to the fact that subjective probability has been modeled after mathematical probability, many researchers have found it necessary to assume that subjective probability functions on an additive basis. Several models of subjective probability have been proposed, all of which either directly or indirectly depend upon the empirical operation of the additivity principle. Examples of such models are 1) "we can denote full belief by 1, full belief in the contradictory by 0, and equal beliefs in the proposition and its contradictory by 1/2. From this it follows that the degree of belief in P + the degree of belief in $\bar{P}$ (not P) = 1; and the degree of belief in P given Q + the degree of belief in $\bar{P}$ given Q = 1. Ramsey (12:171) 2) "Expected utility = Probability x Utility of success + (1 - probability) x Utility of failure." Siegel (14:253) 3) If $\langle T, X, \mathfrak{F} \gtrsim \rangle$ is a weak subjective probability structure, then there is a unique function S defined on $\mathfrak{F}$ such that for every E and F in $\mathfrak{F}$:

i) $s(E) \geq 0$

ii) $s(X) = 1$

iii) $s(E) + s(\bar{E}) = 1$

iv) if $E \subseteq F$ then $s(E) \leq s(F)$

v) $x, y \approx (E)\ u, v,$ if and only if

$xs(E) + ys(\bar{E}) = us(E) + vs(\bar{E})$

Explanation of symbols:

a) - T is a finite interval of real numbers

b) - $\langle X, \mathfrak{F} \rangle$ is a field

c) - $s(E) = \frac{v - y}{v - y} + (x-u)$ if and only if $x, y \approx (E)\ u, v$.

d) - X is in $\mathfrak{X}$

e) - In the case $E \subseteq F$ then $E = F$

f) - There are numbers x, y, u, v, in T such that
$x, y \approx (E)\ u, v$.

g) - $\approx$ means "is approximately equal to" (5:39).

4) " Ψ is a function attaching to an event i a subjective probability, $\Psi(i)$. It is assumed there exists an event i_o such that $\Psi(i_o) = \Psi(\text{not } i_o)$ so that $\Psi(i_o)$ is equal to 1/2." Note: the symbol Ψ denotes subjective probability (3:385). 5) "if we study the relationships between the subjective probabilities of different outcomes of the same event and were to represent the chance of winning a raffle or lottery by p, then 1 - p is the chance of not winning, $p + (1 - p) = 1$. In any situation where there are many possible outcomes, only one of which can actually happen, the probabilities of the separate outcomes add up to 1" (2:87).

Not all of the above-mentioned models have been experimentally tested. Of those that have, the following types of results have occurred. Davidson, Suppes, and Siegel (5:70) in a study utilizing a six-sided die with two sets of 3-letter nonsense syllables, each set printed on three different sides, reported strong evidence supportive of additivity model 3) described above. Their method of measurement was as follows: "for every subject a prediction is possible for all options of the following form, no matter what amount of money x may be:

	option 1	option 2
E	-4¢	6¢
$\bar{E}$	X	11¢

the prediction is as follows: if x $\leq$ 18¢, the subject will choose option 1; if x $\leq$ 14¢, he will choose option 2; if x is between 14¢ and 18¢, no prediction can be made." Their evidence for support was based on how the subject's choice compared with the choice that would have been made if the subject were maximizing actuarial value (in other words picking the objective maximum in each bet). It is not clear exactly how such comparisons were made; however, they do allude to the fact that their results were mathematically derived through a rather complex organization of axioms and postulates of which step iii) in model 3) was an integral part. No evidence is cited which indicates that a subjective estimate of proportion was made of the complementary event of any given bet, which would have made valid their assumption that $s(E) + s(\bar{E}) = 1$.

In order to test model 4) above, Coombs and Komorita (3:387) developed a gambling experiment in which "each bet in the experiment consisted of two amounts of money; an amount to be won and an amount to be lost. The probability of winning a given bet was kept constant at .50 and the probability of losing was its complement.

Consequently, it was equally probable that a S would win or lose a given bet and it is assumed in this study that the subjective probabilities were also equal." By making the assumption that the subjective probabilities of a bet were equal, the experimenters stated they had gained results which supported the additivity principle. At this point a question arises which relates to their basic assumption that the subjective probability of a bet is .50 and its complement is .50 which necessarily totals to 1. Though they apparently utilized this assumption, there is no indication that an attempt was made to measure the subjective estimate of proportion of the complementary event.

That is, although there was, apparently, a test made of the subjective estimate of probability regarding the initial estimate of a given bet, no subjective estimate was made of the complementary event of the same bet to validate the fact that together these subjective probability estimates total 1 as was assumed.

Cohen and Hansel (2:90-91), in model 5) above, devised a study in which 13-year-old boys were asked to place three hypothetical boys running a race as to which one would win, which would come in second, and which would come in third. Each subject was asked to scale the certainty, on a 0 to 10 scale, to which they felt the winner he had selected to win would win. For their second and third place choices, they were asked to scale the certainty that these choices would not win.

The investigators then subtracted the scaled figure for not winning from 10 to determine the judgment for winning. By adding the figures for each non-winner to each subject's direct judgment of the winner, the results are reported as supporting the additivity principle. These results, although consistent with the additivity principle, fail to give it direct support inasmuch as no direct measure was made of the complementary event regarding the winner, i.e., a direct estimate of the non-winning event of each subject's number one choice.

As has been pointed out, all of the studies described assume that the subjective probability of any event and its complement add up to one. Yet they have failed to measure the subjective estimate of probability of the complementary event in order to empirically test the validity of the assumption. Furthermore, Coombs and Komorita (3:384) confounded their work by incorporating utility along with subjective probability.

The present study, in order to test the validity of the basic assumption described above and to avoid the pitfalls of confounding probability estimates with utility, value judgments, expected risks, etc., has adopted the method of direct psychophysical judgment. This method requires a subject to estimate the proportion of one type of element in a display that has stimulus elements of two types. It is further felt that the concept of "relative frequency" which is inherent in (and synonymous with) mathematical probability needs clarification.

Blommers and Lindquist (1:193) describe relative frequency as follows:

> To begin, consider a collection or universe of objects. We shall designate this universe as U. Now suppose that the objects comprising U are of several different kinds. Let <u>one</u> of these kinds or classes of objects be called W. Then the probability of an object of type W in the universe of objects, U, is by definition the relative frequency (expressed as either a common or a decimal fraction) with which type W objects occur in this universe.
>
> For example, suppose the objects of the universe are the individual cards comprising an ordinary 52-card deck of playing cards. Then the probability (relative frequency)* of a spade in this universe is one fourth or .25, since 13 of the 52 cards involved are spades.
>
> * "Relative frequency" was inserted by the writer of this study.

The significance of this principle lies in the fact that it necessarily requires an individual to base his expectations on knowledge of repeated sets of like circumstances observed thus far. That is, one's expectations are extrapolated from past experiences to future estimates in that if a particular set of circumstances continues to arise, a given event will happen the same fraction of times in the future that it has in the past. Two types of estimates of proportion are possible: (1) an estimation of the proportion of an event observed in the past; and (2) an estimation of the proportion of an event expected in the future. The latter type of estimation corresponds closely to the relative frequency view of probability and was used in the following experiment.

On the basis of what has been described thus far, the following pilot study was run.

METHOD AND RESULTS

Subjects:

A sample population of 214 college students (enrolled winter quarter, 1964, at Central Washington State College) was tested. Group size was determined in accordance to the size of the particular classes made available to the experimenter.

Apparatus:

A Revere 808, 2 x 2 slide projector which utilizes 32 slot slide trays was used to present all stimulus material. Each stimulus slide picture was projected onto a 5 1/2' square Radiant (wall model) viewing screen which is part of each classroom's standard equipment. Each slide was presented for a five-second interval and immediately followed by another slide. Timing and changing of each slide was handled through the use of a stop watch and a manually operated electric slide changer (push-button, remote control switch).

The stimuli consisted of stimulus categories selected (refer to Appendix I which specifies the stimulus categories and describes the method of selection) from an array of 2 x 2 slides available, at the time of study, in the Central Washington State College Audio-Visual library. The specific stimulus categories presented were pictures of: birds, flowers, modern ceramic art, and names of flowers. The pictures of names were photographed from 5" x 8" white cards, which had the name typed in regular pica-type, and mounted in a standard 2 x 2 cardboard slide frame.

Procedure:

The general procedure was the same for all groups. Each stimulus series was presented and then followed by a printed questionnaire (refer to Appendix II for examples of the questionnaires). The specific procedure for the first group was as follows: (Note: Each group of subjects was tested in the regular classroom during the regular class hour. The projector was set up and the viewing screen exposed prior to the entrance of any of the students.) Once the group was seated, the instructor introduced the experimenter and he (the experimenter) read the following instructions:

Introduction

> Today you are going to participate in a short experiment on visual perception. All that it requires initially is that you settle back and watch some pictures which I will project on the screen here in the front of the room. I want you to observe the pictures carefully, as you will be asked to fill out a questionnaire as soon as the lights come back on. Please put all books and papers on the floor or under your seat. You are to take <u>no</u> notes while watching these pictures, just pay close attention. Are there any questions?
>
> Before we start, please fold the writing arm of your seat down so it is out of the way.

The lights in the room were then turned off and a series of 72 slide pictures shown. The first series for group one consisted of

36% pictures of airplanes and 64% pictures of airplane diagrams. Each slide was shown for a 5-second period with the next slide appearing as quickly as the machine exchanged slides (approximately one second). When the series finished, the lights were turned on and questionnaires were passed out with the following being read aloud:

> You are to put your name, the name of your instructor, and the date in the appropriate spaces at the top of your questionnaire. Please work as quickly as you can while answering the questions and please use your own best judgment. Do not ponder over the questions; just use your best guess. Please answer all of the questions. As soon as you have finished, please turn your paper face down on your desk, and they will be collected as soon as everyone has finished.

When all questionnaires had been collected, the following set of instructions was read aloud:

> Now I am going to show you another set of pictures. The procedure here is the same as it was for the first set. Just settle back and watch the pictures closely. Make sure all papers and books are on the floor. Remember you are to take <u>no</u> notes while watching these pictures, just pay close attention. Are there any questions?
>
> Please fold the writing arm of the desk down so it is out of the way.

The lights in the room were again turned off and the second series of 72 slide pictures was shown. This series, consisted of 64% pictures of birds, and 36% pictures of bird names. The procedure here was the same exactly as for series #1. When the series finished, the lights were turned on, and a second set of questionnaires was passed out with the following set of instructions being read aloud:

The procedure for filling out this questionnaire is the same as before. You are to put your name, the name of your instructor, and the date in the appropriate spaces at the top of your questionnaire. Please work quickly as you answer the questions and please use your own best judgment. Use your best guess. Please answer all of the questions. As soon as you have finished, please turn your paper face down, and they will be collected as soon as everyone has finished.

When all questionnaires had been collected, the group was thanked for their participation, and told to contact the experimenter at a later date for information concerning the experiment.

The procedure was the same for all groups except for variations in the percentage category shown first and the particular stimulus categories shown. (Appendix I states the method for selecting stimulus categories, and indicates the various groupings of individual estimates of related and unrelated stimuli.) The three groups were presented two series of pictures and asked questions about each series in the following sequence. Note: The following abbreviations will be made: unrelated (UR); related (R); airplanes (A); airplane diagrams (AD); birds (B); bird names (BN); flowers (F); flower names (FN); ceramics (C).

	Group I (R)	Group II (UR)	Group III (UR)
Trial #1	36% (A) - 64% (AD)	64% (C) - 36% (F)	64% (F) - 36% (C)
Trial #2	64% (B) - 36% (BN)	36% (B) - 64% (FN)	36% (FN) - 64% (B)

Table I illustrates the findings of the various statistical measures used to analyze the data. Two out of the three groups tested significantly overestimated or underestimated the objective total of 100%, according to t-tests and confidence interval computations (refer

TABLE I

PILOT STUDY (2 SETS OF STIMULI)

	N	Mean	S. D.	χ^2	t-test	Confidence Intervals
Group #1	72	91.888	27.39	40.50*	2.547*	85.56 to 98.40**
Group #2	71	100.338	13.35	23.676*	.2133	97.242 to 103.434
Group #3	71	110.718	25.41	44.16*	4.054*	105.568 to 115.868**

* Significant at the .05 level or below

** Indicates ranges not containing the objective total of the two relative frequencies asked for.

χ^2 Tested for a significant variation from 50% of the sample population accurately estimating the objective total of 100% and 50% of the sample population not accurately estimating it.

t-test tested for a significant mean variation from the objective total of 100%

to Table I for an explanation of the types of statistics used). Furthermore, chi square tests on all three groups indicate that most of the subjects tested were unable to accurately estimate the relative frequency total of 100%, which was the additive total of the two relative frequencies asked for.

The results of this pilot study summarily indicate that the summed subjectively estimated proportions of a projected total event (in this case a subjective estimate of the first event plus a subjective estimate of a comparable event which was the empirical complement of the first event) add to more or less than one (or 100%) which negates the applicability of the additivity theorem.

For the purposes of the present study the following hypothesis will be tested:

A significant proportion of subjective estimates of the expected proportions of a stimulus class and its complement, from a universe composed of two stimulus classes, will total to some figure different from one or 100%.

CHAPTER II

DESIGN, METHOD, AND PROCEDURE

Subjects:

The sample for this study consisted of 212 college undergraduates (enrolled summer quarter, 1964, at Central Washington State College). Group size was determined in accordance to the size of the particular classes made available to the experimenter and varied from $18S_s$ to $44S_s$ in number.

Apparatus and Procedure:

The apparatus and procedure used for this study was the same as that used in the pilot study except for changes in stimuli presented, percentage presented, and number of categories shown. The following changes were made in the present study.

(1) Only unrelated stimuli categories were presented. (Refer to Appendix III for information concerning the selection of the unrelated categories.)

(2) A control group was used to act as a check on consistency of estimating the same proportion from

two different (unrelated) stimulus categories.

(3) A one-set series of stimuli was presented with two questionnaires being answered (the first questionnaire being completed and returned to the experimenter before the second questionnaire was passed out) upon completion of the showing of the one set. The purpose of the single set of stimuli was to determine if a significant variation from the objective total of 100% would occur when the initial estimate and its complement were contained in the same set of stimuli. The "one" and "two" set methods of stimulus presentation are discussed fully in Chapter IV.

The method of presentation for changes (2) and (3) above are as follows:

(2) Note: Each group of subjects were seated in six rows to facilitate the handling of the questionnaires; also the following abbreviations will be made: Prehistoric Animals (PA); Flowers (F); Birds (B); Modern Ceramic Art (MC).

Procedure:

Experimental group I

1 - The first series of 72 pictures consisting of 17% (PA) and 83% (F) was presented.

2 - When the series had been shown, the questionnaires were passed out and counterbalanced in the following manner: (a) the questionnaires asking for the percentage of (PA) were passed to rows 1, 3, and 5 from front to back. (b) The questionnaire asking for the percentage of (F) were passed to rows 2, 4, and 6 from front to back.

3 - The questionnaires were collected.

4 - The second series of 72 pictures consisting of 83% (B) and 17% (MC) was presented.

5 - When the series had been shown, the questionnaires were passed out and counterbalanced in the following manner: (a) the questionnaires asking for the percentage of (B) were passed to rows 1, 3, and 5 from front to back. (b) The questionnaires asking for the percentage of (MC) were passed to rows 2, 4, and 6 from front to back.

This procedure was the same for all groups except for the following variations in the percentage of each category shown. The following outline illustrates the specific changes made for experimental groups 2, and 3, and the control group.

Procedure:

Experimental group two

1.	first series shown	64% (MC) and 36% (B)
2.	second series shown	36% (F) and 64% (PA)

Procedure:

Experimental group three

1. first series shown 50% (B) and 50% (MC)

2. second series shown 50% (PA) and 50% (F)

Procedure:

Control group (Same as experimental group two)

1. first series shown 64% (MC) and 36% (B)

2. second series shown 36% (F) and 64% (PA)

(Refer to Appendix IV for examples of the questionnaires used.)

(3) The procedure for experimental groups 4 and 5 utilized the same instructions concerning the introduction to the experiment and the instructions concerning each set of questionnaires. The only difference being, groups 4 and 5 were presented only one set of stimuli. The procedure was as follows:

Procedure:

Experimental group four

1. Present 72 pictures consisting of 36% (B) and 64% (MC).
2. Pass out the first questionnaire to all subjects asking for the percentage of birds and collect when finished.
3. Pass out the second questionnaire to all subjects asking about the percentage of ceramics and collect when finished.

Procedure:

Experimental group five: (same as above except for change in step one)

1. Present 72 pictures consisting of 83% (B) and 17% (MC) (Refer to Appendix V for examples of questionnaires used.)

CHAPTER III

RESULTS

The main question proposed by this study was whether a subjective estimate of expected proportion of a first event <u>plus</u> a subjective estimate of expected proportion of a comparable second event, which was the empirical complement of the first event, add to more or less than the total of the sums of the relative frequency of both events. Frequency polygons indicating the expected proportions estimated by each group are shown in Charts 1, 2, 3, 4, and 5.

Three sets of data were analyzed for relevance to this question:

(1) Group performance of the three experimental groups tested with two sets of stimuli.

(2) Group performance of the control group tested with two sets of stimuli.

(3) Group performance of the two experimental groups tested with one set of stimuli.

Each set of data was assessed through the use of three methods of statistical analysis.

CHART 1

FREQUENCY OF ESTIMATES--GROUP I--17%-83%--(2 SETS OF STIMULI)

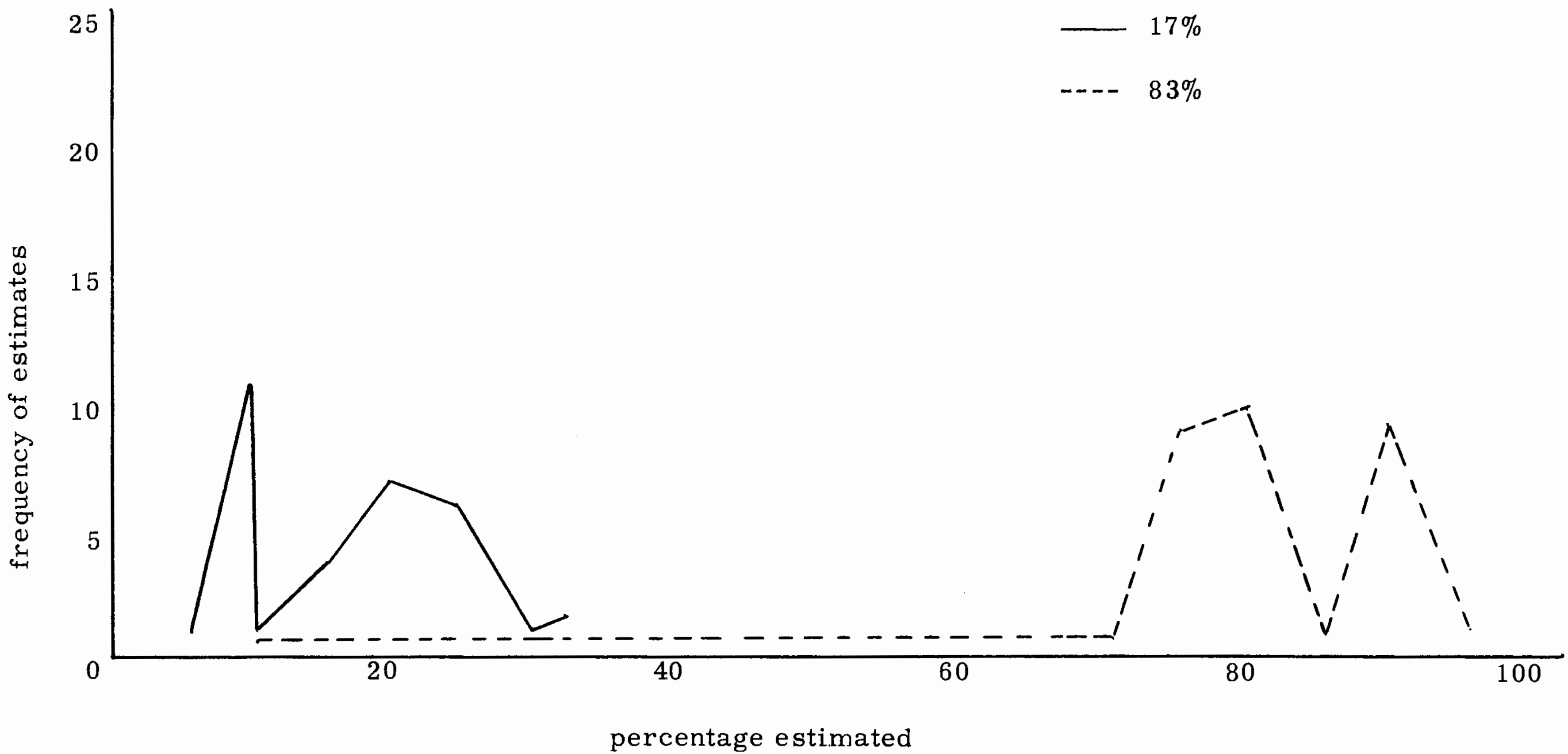


CHART 2

FREQUENCY OF ESTIMATES--GROUP II--36%-64%--(2 SETS OF STIMULI)

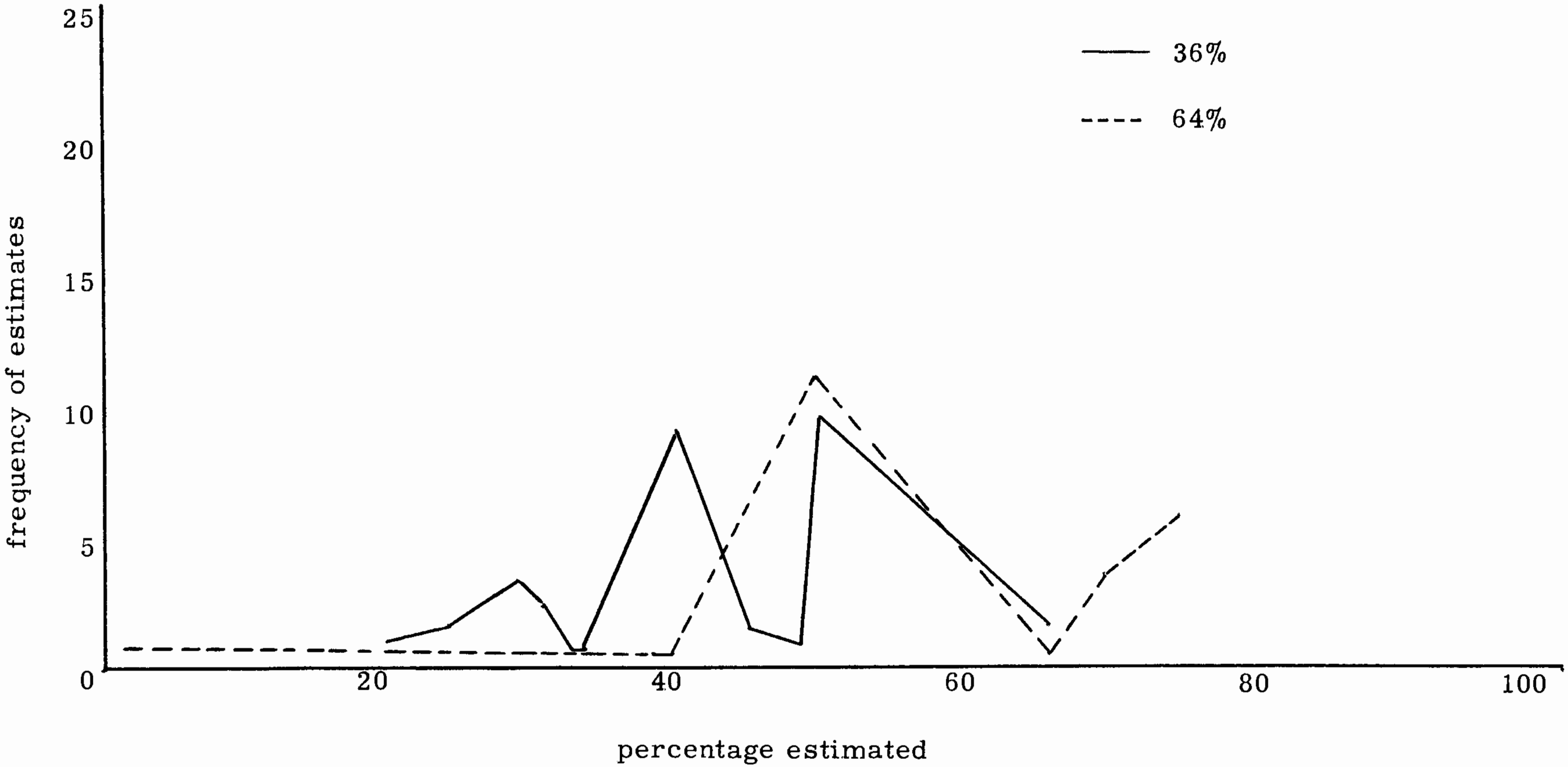


CHART 3

FREQUENCY OF ESTIMATES--GROUP III--50%-50%--(2 SETS OF STIMULI)

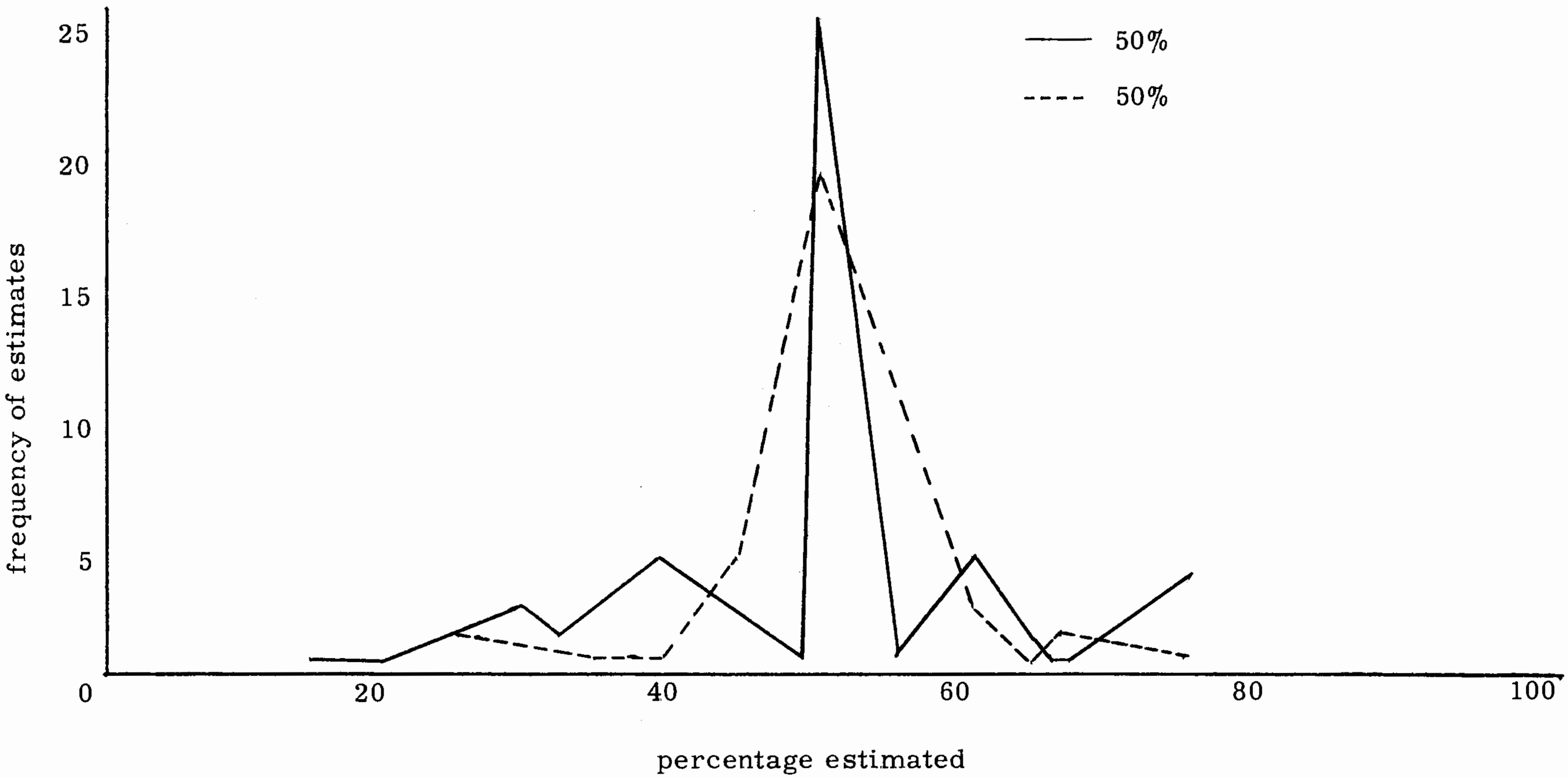


CHART 4

FREQUENCY OF ESTIMATES--GROUP IV--36%-64%--(1 SET OF STIMULI)

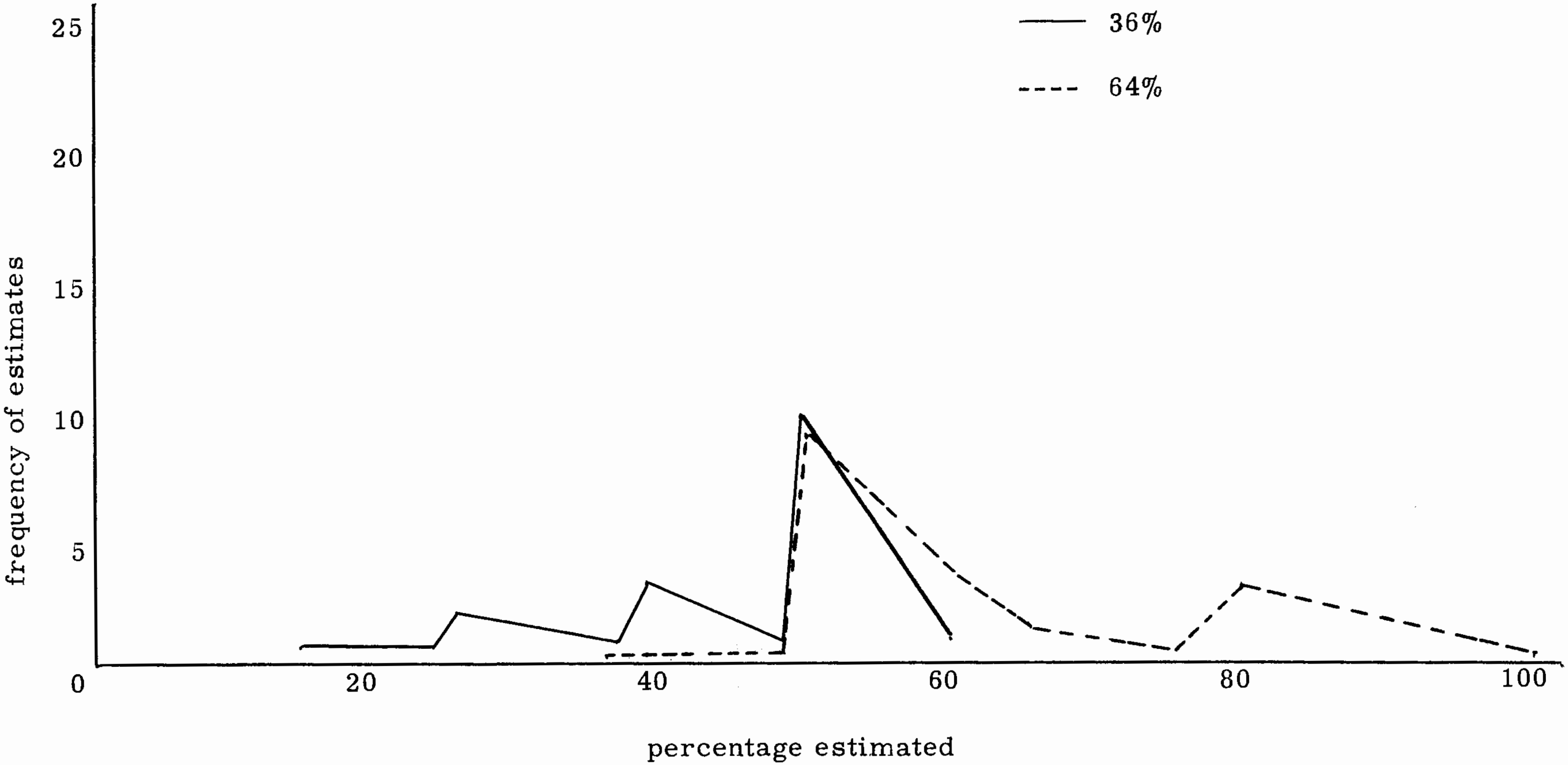
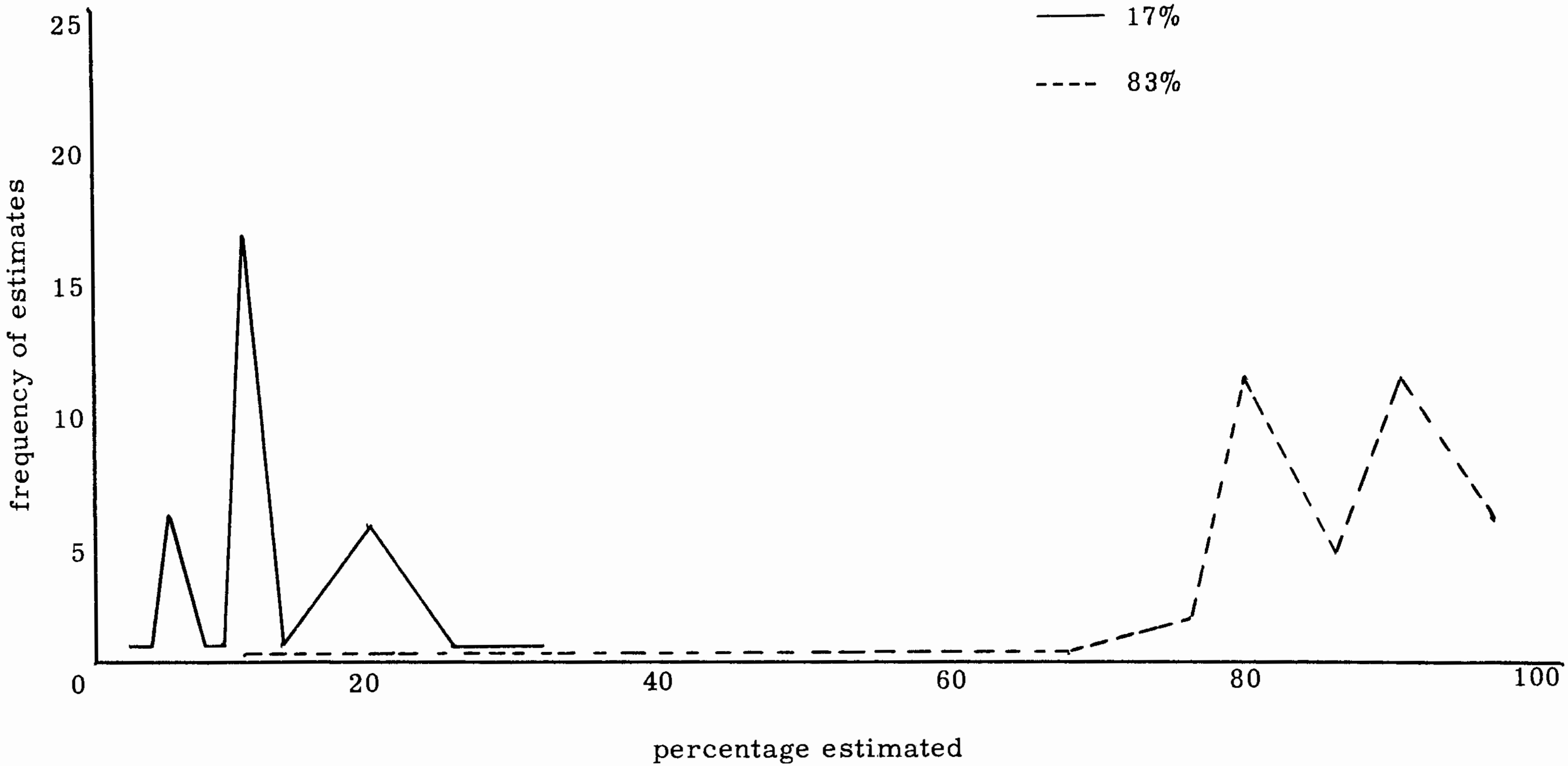


CHART 5

FREQUENCY OF ESTIMATES--GROUP V--17%-83%--(1 SET OF STIMULI)



(1) A chi square test which tested the assumption that by chance 50% of the subjects in any particular group tested would estimate the additive total of the relative frequency of both events, and 50% of the subjects would not.

(2) A t-test for a difference between means which tested for a significant difference between the subjectively expected total of the subjective estimate of the first event plus the subjective estimate of a second event, which was the empirical complement of the first event, versus the additive total of the relative frequency of both events.

(3) A calculation of a confidence interval designed to illustrate whether the additive objective proportion of both events fell within the range of probable population hypotheses estimated by the sample.

The performance of Experimental groups I, II, and III, though tested by all three methods of statistical analysis, attained significant departures from additivity only by a chi square analysis (50% of the Ss would estimate a total of 100% and 50% of the Ss would not). Levels of 9.000 for the 64% (MC) - 36% (PA) and 4.2632 for the 36% (B) - 64% (F) estimates of Group II; and 4.5090 for the 50% (B) - 50% (PA) and 4.5090 for the 50% (MC) - 50% (F) estimates of Group III were derived. In both groups, significance was attained by the fact that the majority of the subjects did <u>not</u> subjectively estimate the expected total 100%.

No significant evidence from either the t-test or the confidence intervals was derived from the performance of these three experimental groups. That is, there is no empirical evidence to support the assumption that group performance varied significantly from the mathematically expected relative frequency of 100%. A complete statistical analysis of the data is illustrated in Table II.

Control group performance revealed chi square levels which proved significant. Chi square levels of 20.8000 for control group (A) which made estimates of the 36% categories of two comparable events, and 18.000 for control group (B) which made estimates of the 64% categories of two comparable events were derived. The chi square test used in this stage assumed that 50% of the Ss tested would estimate 72% (the additive total of estimate one plus estimate two) and 50% would not for group (A), and that 50% of the Ss tested would estimate 128% (the additive total of estimate one plus estimate two) and 50% would not for group (B). Significance was attained by the fact that none of the subjects in either group subjectively estimated the expected relative frequency total of estimate one plus estimate two.

Furthermore, a t-test (as described above) resulted in a significant difference from additivity for both control groups. Additivity in this sense refers to the summed objective proportions of the two categories not necessarily a total of 100%. In the case of group (A) the mean of the sums of estimates one and two was significantly above the additive total of the relative frequencies of the categories representing those estimates. In the case of group (B) the mean of the sums of estimates one and two was significantly below the additive total of the relative frequencies of the categories representing those estimates. For both groups the confidence intervals further support the significant difference between subjectively estimated totals and additive totals in that neither confidence interval contains the additive total for the specific group it relates to. A complete analysis of the data is illustrated in Table III, Section A.

The final stage of the analysis deals with experimental groups IV and V who were shown only one set of stimuli. The results from group V alone show a significant mean difference from additivity. A t-test for the difference between the group mean for the subjectively expected total of the sums of the estimates of the two categories shown and the relative frequency total of those two categories yielded a t of 2.134 (significant at the .05 level) resulting from a subjectively estimated mean significantly below the relative frequency total. The significance of this finding was further supported by confidence intervals not containing the total of the relative frequencies of both categories. All statistical measures, other than those cited above, for significant differences between total subjective estimates and total objective proportions proved insignificant. Refer to Table V, Section A for a complete illustration of the statistical analysis.

The second important question is that of intra-group differences between subjective estimates of expected proportion and the objective proportion of stimulus categories. The statistical methods are the same as those above except that in this case each group will be analyzed in relation to each percentage level they were tested on, rather than the additive total of the relative frequency of both events.

Experimental Groups I, II, and III showed significant chi square levels, which assumed that 50% of the subjects would estimate the expected relative frequency of the proportion they were asked to estimate and 50% would not. Chi squares of 33.000 (83% relative frequency), χ^2 = 33.000 (17% relative frequency), χ^2 - 35.000 (64% relative frequency), and χ^2 = 35.000 (36% relative frequency) were derived. The significance of these chi square levels can be summarized by stating that none of the subjects in either Group I or II estimated the actual relative frequency of the percentage level they were asked to estimate.

In relation to the 36% category which group II was asked to estimate the relative frequency of, a t-test for the difference between means yielded a t of 3.1403 (significant at the .01 level) which indicates that the mean subjective estimate at that percentage level was significantly above the relative frequency presented. Confidence intervals not containing the actual relative frequency further support the significance of this difference. A complete analysis of the data is presented in Table IV.

Analysis of the control group's performances showed significant chi square levels of $\chi^2 = 20.800$ (first estimate-control group A), $\chi^2 = 20.800$ (second estimate-control group A), $\chi^2 = 18.000$ (first estimate-control group B), and $\chi^2 = 18.000$ (second estimate-control group B). As was true in the case of Groups I and II in stage one, none of the subjects in either group subjectively estimated the correct objective proportion of any stimulus category.

The performance of the control groups also yielded significant t's of 4.1876 for the first estimate of control group A, and 2.9187 for the first estimate of control group B (probability $< .05$). These differences indicate that the subjective estimates of relative frequency for subjects in the control groups, on their first estimates, deviated significantly above (36%--group A) and below (64%--group B) the actual relative frequencies of the stimulus categories presented. Both significant mean differences were further supported by confidence intervals which failed to contain the actual relative frequency of the category estimated. Table III, Section B illustrates a complete analysis of the data.

Responses of Experimental groups IV and V show the following chi square levels, all significant at the .05 level or beyond, of χ^2 = 23.000 for group IV (36%), χ^2 = 23.000 for group IV (64%), χ^2 = 38.000 for group V (17%), and χ^2 = 38.000 for group V (83%). Again the significance attained here was due to the fact that none of the subjects in either group subjectively estimated the actual relative frequency of the stimulus categories presented.

A t-test based on the 17% objective proportion category, yielded a t of 4.9705, significant $< .01$. This indicates that the groups mean subjective estimate of the relative frequency of the category presented was significantly different (below) from the actual relative frequency of the category. This difference was again further supported with confidence intervals which did not contain the actual relative frequency. A complete analysis of the data is provided in Table V, Section B.

Summarily, the statistical analysis of question two yielded many significant results in relation to the chi square tests used but offered relatively few significant results regarding the t-tests and confidence intervals. Of the fourteen t-tests computed in this stage of analysis only four proved to be significant. However, this is in part an artifact due to averaging individual estimates of proportion. The chi square analysis based on the results of each individual estimate indicates that most subjects never accurately estimated the objective proportion.

It can generally be stated from these results that group mean subjective estimates at each percentage level, for the most part, did not deviate significantly from the actual relative frequency at that level. Tables III, Section B, IV, and V, Section B provide a complete analysis of the data pertaining to question two.

To further analyze the results of the control group a chi square test for consistency versus non-consistency was run. The results of these tests were as follows:

1. The consistency between the first estimate and second estimate for control group A (36%) was negative at a chi square level of 5.76 (probability $<$ .02). That is, the first and second estimates of Group A were not consistent.
2. The same was true for Group B (64%). Chi square equalled 10.88 which is significant at the .01 level and beyond.

The final question analyzed in studying the results of this experiment pertains to intra-group variability concerning the relationship between estimations of the specific objective percentage level and the additive total of 100%. Specifically, do Ss who accurately estimate the objective proportion at each percentage level also accurately estimate the additive total of 100%? The following results were derived from chi square tests for correlated proportions. (Note: All chi square results are significant at the .01 level or beyond.)

1. Two sets of stimuli:

Group I - 17% estimate:

Chi square = 10.000

	17	$\overline{17}$	
$\overline{100}$	0	23	23
100	0	10	10
	0	33	30

Group I - 83% estimate:

Chi square = 10.000

	83	$\overline{83}$	
$\overline{100}$	0	23	23
100	0	10	10
	0	33	33

Group II - 36% estimate:

Chi square = 7.000

	36	$\overline{36}$	
$\overline{100}$	0	28	28
100	0	7	7
	0	35	35

Group II - 64% estimate:

Chi square = 7.000

	64	$\overline{64}$	
$\overline{100}$	0	28	28
100	0	7	7
	0	35	35

2. One set of stimuli:

Group IV - 36% estimate:

Chi square = 14.000

	36	$\overline{36}$	
$\overline{100}$	0	9	9
100	0	14	14
	0	23	23

Group IV - 64% estimate:

Chi square = 14.000

	64	$\overline{64}$	
$\overline{100}$	0	9	9
100	0	14	14
	0	23	23

Group V - 17% estimate:

Chi square = 16.000

	17	$\overline{17}$	
$\overline{100}$	0	22	22
100	0	16	16
	0	38	38

Group V - 83% estimate:

Chi square = 16.000

	83	$\overline{83}$	
$\overline{100}$	0	22	22
100	0	16	16
	0	38	38

These results indicate that none of the subjects were able to estimate the relative frequency of the stimulus category presented. However, many subjects arrived at a total relative frequency estimate of 100%, but none of them did so by accurately estimating the relative frequency of the two stimulus categories.

On the other hand, significant results (.01 level of significance) of a different character were shown in the estimates of Group III which was asked to estimate categories with objective relative frequencies of 50%. The results for Group III are as follows:

1. Group III - one 50% estimate:

Chi square = 6.400

	50	$\overline{50}$	
$\overline{100}$	9	23	32
100	11	1	12
	20	24	44

2. Group III - a second 50% estimate:

Chi square = 8.333

	50	$\overline{50}$	
$\overline{100}$	11	21	32
100	11	1	12
	22	22	44

These results indicate that subjects who failed to estimate the objective proportion of the specific percentage level presented also failed to estimate the total relative frequency presented and vice versa those who estimated the objective proportion of the specific percentage level presented also estimated the total relative frequency presented.

In summary, the over-all results fail to give unequivocal support to the proposed hypothesis.

TABLE II

GROUP PERFORMANCE VS. OBJECTIVE TOTAL OF 100% (2 SETS OF STIMULI)

		N	Mean	Standard Deviation	Chi square	t-test	Confidence Intervals
GROUP I	Set #1 83%-17% F MC	33	99.0000	10.7540	2.0000	.03952	94.0320 to 103.9680
	Set #2 17%-83% PA B	33	91.0666	21.3659	3.2666	.61977	81.8541 to 101.8791
GROUP II	Set #1 64%-36% MC PA	35	93.5000	25.2454	9.0000 **	.01299	81.1299 to 105.8701
	Set #2 36%-64% B F	35	104.8421	16.4807	4.2632 *	1.2806	97.4316 to 112.2526
GROUP III	Set #1 50%-50% B PA	44	107.0909	17.1505	4.5090 *	1.8920	99.9242 to 114.2576
	Set #2 50%-50% MC F	44	93.1818	20.6578	4.5090 *	1.7691	84.5496 to 101.8140

* significant $< .05$

** significant $< .01$

TABLE III

SECTION A. CONTROL GROUP (2 sets of stimuli) GROUP PERFORMANCE VS. OBJECTIVE TOTAL OF ESTIMATE 1 PLUS ESTIMATE 2

			N	Mean	Standard Deviation	Chi square	t-test	Confidence Intervals
CONTROL GROUP A	36	36 (72)	21	85.3333	12.9595	20.8000 **	** 4.4787	79.7906 to 90.8760*
CONTROL GROUP B	64	64(128)	18	111.6666	26.2364	18.0000 **	* 2.4133	99.5460 to 123.7872*

SECTION B. CONTROL GROUP (2 sets of stimuli) GROUP PERFORMANCE VS. THE OBJECTIVE PROPORTION OF EACH % LEVEL

CONTROL GROUP A	1st est. vs. 36% obj. prop.	21	45.7619	9.016	20.80 **	** 4.18766	41.9058 to 49.6180*
	2nd est. vs. 36% obj. prop.	21	39.5714	13.1234	20.80 **	1.2471	33.9582 to 45.1846
CONTROL GROUP B	1st est. vs. 64% obj. prop.	18	50.4444	21.2368	18.00 **	** 2.9187	40.6334 to 60.2554*
	2nd est. vs. 64% obj. prop.	18	61.2222	11.5195	18.00 **	1.4324	56.9006 to 66.5438

* significant $< .05$

** significant $< .01$

TABLE IV

GROUP PERFORMANCE VS. THE OBJECTIVE PROPORTION OF EACH % LEVEL
(2 SETS OF STIMULI)

		N	Mean	Standard Deviation	Chi square	t-test	Confidence Intervals
GROUP I	Est. vs. 83% obj. prop.	33	78.0303	16.5330	33 **	1.7267	72.3895 to 83.6711
	Est. vs. 17% obj. prop.	33	17.3636	7.6000	33 **	.8690	14.7705 to 19.9567
GROUP II	Est. vs. 64% obj. prop.	35	58.1142	17.6362	35 **	1.9744	52.2715 to 63.9569
	Est. vs. 36% obj. prop.	35	41.5428	10.4417	35 **	* 3.14039	38.0836 to 45.0020*
GROUP III	Est. vs. 50% obj. prop.	44	48.7727	11.8115	.00009	.7660	45.2828 to 52.2626
	Est. vs. 50% obj. prop.	44	51.3636	11.3628	0.0000	.7164	48.0062 to 54.7210

* significant $< .05$

** significant $< .01$

TABLE V

SECTION A. GROUP PERFORMANCE VS. OBJECTIVE TOTAL OF 100% (1 SET OF STIMULI)

		N	Mean	Standard Deviation	Chi square	t-test	Confidence Intervals
GROUP IV	36%-64% B C	23	99.3478	10.8460	1.0868	.2883	94.9153 to 103.7803
GROUP V	17%-83% B C	38	95.2368	13.7592	.8772	2.134*	* 89.8621 to 99.6115

SECTION B. GROUP PERFORMANCE VS. OBJECTIVE PROPORTION OF EACH % LEVEL (1 SET OF STIMULI)

		N	Mean	Standard Deviation	Chi square	t-test	Confidence Intervals
GROUP IV	36%	23	40.260	12.3638	23 **	1.6525	35.207 to 45.313
	64%	23	59.0869	15.3267	23 **	1.5377	52.8232 to 65.3506
GROUP V	17%	38	11.8157	6.4304	38 **	** 4.9705	* 9.7712 to 13.8602
	83%	38	83.421	14.0210	38 **	.1846	78.963 to 87.879

* significant < .05

** significant < .01

CHAPTER IV

DISCUSSION AND CONCLUSIONS

The main purpose of these experiments was to determine the validity of the application of the additivity principle to subjective estimates of proportion. As a first step in clarifying the results, the methodological structure of the study will be examined.

This study was constructed to assess the problem of additivity, defined as: the process of adding the probability of the occurrence of an event and the probability of the non-occurrence of the same event and arriving at a total of <u>one</u> or in this case 100%. The problem was approached by two relatively different methods. The first method was designed to determine whether additivity is a functional principle, or not, by having groups of subjects make estimates of expected proportions of one category (from a set of stimuli containing two categories), <u>plus</u> a similar estimate of a comparable category (from a different but analogous set of stimuli) which was the numerical complement of the first category. Both estimates were then tallied and compared to the total relative frequency of 100% for experimental groups I, II, and III. This method will be referred to as the "two sets of stimuli" method of stimulus presentation.

A control group was included in this method to determine the consistency of subjectively estimating the same relative frequency level from two different (but comparable) sets of stimuli. The control group subjects were asked to make a subjective estimate of one category, plus a subjective estimate of a comparable category which was numerically equivalent to the first category (the two sets of stimuli used were the same sets as those used above). Again both estimates were tallied and compared to the total relative frequencies of 72% and 128% for control groups A and B respectively.

An analysis of the data from the experimental groups tested under the "two sets of stimuli" method (experimental groups I, II, and III) indicated that there were no significant differences, between the total mean of the subjective estimates of each group, and the total relative frequencies to which they were compared. In other words, the summed subjective estimates did not differ significantly from the chance sampling of a population of summed estimates whose mean was equal to 100%. In this sense, then, additivity was confirmed.

On the other hand, the chi square tests (which assumed that 50% of the subjects would estimate the relative frequency total of 100% and 50% would not) show significant evidence that the majority of the subjects were unable to accurately estimate the summed relative frequencies of the objective proportion presented. Thus, if the additivity theorem is strictly interpreted to mean that individual subjects will make summed estimates totaling 100%, this evidence contradicts it.

Furthermore, the performance of the control groups indicated that subjects were significantly unable to consistently estimate a specific proportion from two different sets of stimuli. Table III, Section A, illustrates the extent and magnitude of the significance of the findings related to this phenomenon.

The performance of the control group is of crucial importance to the study, due to the design of the "two sets of stimuli" procedure. It had been assumed beforehand that if subjects were consistent when estimating the same relative frequencies from two similar sets of stimuli, the method of summing a subjective estimate of the first event from one set plus a subjective estimate from a comparable and empirically complementary event from a second set (estimate one plus estimate two) would be a valid test of the additivity theorem. However, due to the failure of the control group to function as expected, the "two sets of stimuli" method appears to be an inadequate measure.

However, although the "two sets of stimuli" method is inadequate, the data does indicate that the additivity principle should not imply a strict one-to-one relationship between subjective estimates and objective proportions as some of its proponents claim must be the case (Ramsey and Hunt). A great deal of variation was shown in relation to individual total estimates and the group mean estimates.

At least half of the subjects tested who estimated a total of 100% did so by estimating their categories in terms of round figures, i.e., 50%-50% or 60%-40% or 80%-20% and so on. The result of such estimates generally tended to be meaningfully different from the actual relative frequency presented. For example, when asked about the 17% category from a set of two categories of stimuli (17%-83%), 13 out of 26 subjects underestimated at 10% and vice versa. That is, the same subjects who underestimated the 17% category at 10% overestimated the 83% category at 90%. Thus by combining the two estimates these subjects ultimately reached the objective total relative frequency of 100%. However, the specific objective proportion of each stimulus category was not accurately estimated.

This same phenomenon was apparent in the 36% and 64% categories. At these particular levels 11 out of 21 subjects overestimated the 36% level at 50%, and underestimated the 64% level at 50% and thus reached the relative frequency total of 100%. This type of estimation is further supported by the performance of the control group who were tested twice on the same percentage level. That is, the control group subjects were asked to make an estimate of a 36% category from the first set of stimuli; then they were asked to make an estimate of a 36% category from a second set of stimuli. The procedure was the same for the 64% control group subjects. About one-fourth, 9 out of 39, of the control group subjects tested estimated

100% (estimate one plus estimate two) and of these 9 subjects, 8 of them did so by estimating 50% for both categories. It is evident, therefore, in this study that although the total relative frequency of 100% was attained, it was not attained on the basis of accurately estimating the separate relative frequencies.

The results from the "one set of stimuli" method, which was the second type of approach used in assessing the problem of additivity, were highly similar to those of the "two sets of stimuli" method. What has been said in relation to individual and the group performances for the "two sets of stimuli" method is also true for the "one set of stimuli" method, except, of course, for the control group. The "one set of stimuli" method was designed to determine whether, or not, additivity is a functional principle, by asking groups of subjects to make a subjective estimate of one category of stimuli (from a set of stimuli containing two categories), plus a subjective estimate of the second category of stimuli (which was the direct complement of the first category--both categories of the stimuli being contained in the same set). The purpose of the method was to determine whether a significant variation from a total frequency of 100% would occur (though anticipated that no variation would occur) when the initial estimate of a specific category and the estimate of that specific category's complement were drawn from the same set of stimuli. Implications derived from the results of this method will be discussed later in this chapter.

In order to study further the nature of the variations which appeared in relation to the total relative frequency measures made, the data from all groups was analyzed for significant variations between subjective estimates and relative frequencies at each percentage level presented. The results of this type of analysis yielded a rather startling fact regarding the chi square tests which were applied. No one estimated the actual relative frequency of any category except at the 50% level presented. The peculiarity of this finding can only be guessed at on a purely speculative basis at this point. One presumption underlying this behavior is that subjects may have personal biases regarding the percentage estimates they subjectively select. On the other hand, the subjects may have simply been inaccurate in their estimations. Whatever the case may be, the factors which seem to influence subjective estimates of probability are particularly worthy of further investigation if psychology is to attain a comprehensive analysis of the applicability of the additivity theorem and related notions of probability.

Further significant findings regarding the analysis conducted at each percentage level presented were derived from t-tests and confidence interval calculations which were applied. Of the fourteen t-test and confidence interval calculations conducted to analyze subjective estimates versus objective frequencies (these two tests were used concomitantly), only four proved to be significant. Two of the four significant findings were directly related to the control group's results. The significance of these control group findings tend to further support the fact that a great deal of variability exists regarding a subject's ability to accurately and consistently estimate objective proportions. This in turn gives further support to the proposal that a strict adherence to the additivity principle is untenable.

The significance of the 36% experimental group II t-test and confidence interval calculation is minimized by virtue of the fact that it was the only significant finding resulting from the "two sets of stimuli" method. However, the significance of the 17% experimental group V finding merits consideration in that it was an estimate which was the direct complement of the 83% category. That is, this estimate was asked for almost immediately (approximately four minutes) after the 83% category had been asked for, and both estimates were drawn from the same set of stimuli. It would seem safe to say at this point that if an inaccuracy in estimation of this degree exists regarding this type of an approach to additivity, then it certainly gives added weight to the argument cited above against a strict adherence to additivity.

To carry the evidence compiled thus far, regarding the preceding statement, a step further, it should be pointed out that in relation to the confidence interval calculations used in this study, the subjectivity estimated population means computed for both the total relative frequencies and the relative frequencies at each percentage level, could have fallen anywhere between the two limits calculated. Although the computations show that the subjectively estimated mean and the 100% total fell within the same interval, a cursory survey of the subjective means for each group will indicate that the two were by no means the same.

Before the discussion moves too far afield from the "two sets of stimuli" method described above, the results of a questionnaire designed to introspectively assess the concepts subjects formulated during the viewing of each of the two sets of stimuli will be presented. This questionnaire was introduced as an attempt to gain some understanding of the sets, hypotheses, or situational cues which may have influenced the observations of the subjects. The questionnaire, (Refer to Appendix VI for an example of the questionnaire) asked each subject to list what he, or she, thought were the important things to look for and remember in each set of slides. The following list is a compilation of the range and the frequency of concepts given by all subjects tested under the "two sets of stimuli" method.

Important things looked for in the first set of stimuli were:

		Frequence of Occurrence
1.	percentage of each category of stimuli presented	14
2.	colors of the stimuli	25
3.	sequence of stimuli	36
4.	time each picture appeared on the screen	16
5.	backgrounds in each picture	6
6.	number of animal pictures	11
7.	identify the picture to appear by the picture preceding it	8
8.	sequence of colors	3
9.	size and shapes of stimuli	9
10.	whether or not pictures were black and white or in color	1
11.	stimulus repetitions	2
12.	characteristics of the animals shown	2
13.	associate each stimulus object with its name	6
14.	whether each stimulus was a painting or a photograph	1
15.	the different species of animals	4

Things looked for in the second set of stimuli (after first set was presented) were:

		Frequency of Occurrence
1.	percentage of each category of stimuli presented	16
2.	whether or not a woman was present in any picture	6
3.	sequence of stimuli	28
4.	types of stimuli presented (i.e., waterfowl, bowls, saber-tooth lion, daisies, etc.)	18
5.	color of stimuli	15
6.	setting of the dominant figure	7
7.	number of each type of picture (i.e., number of birds, number of flowers, etc.)	13
8.	identify the picture to appear by the picture preceding it	3
9.	number of items in each picture (i.e., two birds, three cups, two prehistoric animals, etc.)	4
10.	number of pictures showing a body of water	10

11. proportions of the various stimuli within the same category (i.e., the number of waterfowl to other types of birds, number of lilies to number of daisies, etc.) 11

12. look for the same things asked for on the first questionnaire 24

While compiling the above lists it appeared that many of the subjects professed to be looking specifically for the percentage of each category in the sets of stimuli they were observing. Also many of the subjects indicated, either by inference or by direct admission, that the first questionnaire structured what they were to look for in the second set of stimuli. Apart from these factors, and yet relative to the study, is the fact that many of the concepts in the ranges listed above would seem to act as confounding variables to the purpose of the "two sets of stimuli" method. Although the validity of the items listed above is questionable, the nature of these responses indicate that there is a great deal of complexity involved in the selection of stimulus categories appropriate to the specific method being utilized.

A follow-up observation was made regarding the fact that many of the subjects professed to be looking specifically for the percentage of the categories they were observing. The inquiry sheets were separated into a group of 14 Ss who specifically stated that they were looking for the percentage of each category, and a group of 14 Ss who made no mention of the fact that they were looking for a ratio or percentage. The purpose of this grouping was to determine whether the subjects who stated they were specifically oriented toward this task were more accurate in their objective proportion and relative frequency total estimations than those who did not profess such orientation. The results indicated that neither group proved to be more accurate than the other. In fact, they were about equal in the randomness of their variations around the specific proportions at either level. A chi square of .0109 probability $> .90$ was derived from this analysis.

In summing the evidence of the study to this point, the necessary conclusion that must be drawn is that the additivity principle seems to be functional in aiding people to estimate total situations on the basis of repeated sets of like circumstances. However, as has been pointed out on the basis of individual variations around specific objective proportions and total relative frequencies, the additivity principle does not seem to be functional through an adherence to strict objective proportions. The differential implications from the two major types of statistics used in the study, namely the t-test and the chi square test, pose the need for a flexible interpretation of additivity. Consider first the t-test. A t-test is based on the notion that a total population adheres to the concept of a normal distribution and that the mean or average of the population is a specified value. Evidence to confirm or disconfirm this theoretical notion stems from whether the sample observations deviate from the chance sampling expected by the theory. A hypothesis confirmed by a t-test is confirmed only to the extent that the notion of a population average at a particular value is not inconsistent with the sample evidence. Deviations of theoretical population values about the theoretical population mean are assumed in a t-test rationale. Such spreads are estimated by the sample variance. Furthermore, the population mean so confirmed may well be an artifact of averaging and no particular individual observation in the population sample is necessarily expected at such a value. A cursory examination of the group means for total subjective estimates in this study will reveal that a broad distribution of summed estimates did occur which realistically confirms the proposition that a relatively flexible interpretation of additivity is needed.

Along this same line of reasoning a consideration of the chi square test of the type used in this study (by chance 50% of the subjects will estimate 100% and 50% will not) which invokes a stricter sense of adherence to additivity, revealed that in almost every case the chi square levels obtained were significant in that the majority of the subjects tested failed to estimate either the objective proportion at each level, or the total relative frequency presented. Therefore, on the basis of these implications, the data of this study tends to indicate that an applicable additivity principle must incorporate a considerable degree of variability. The degree of variation suggested by this study may be as great as fifteen percent above or below the suggested mean.

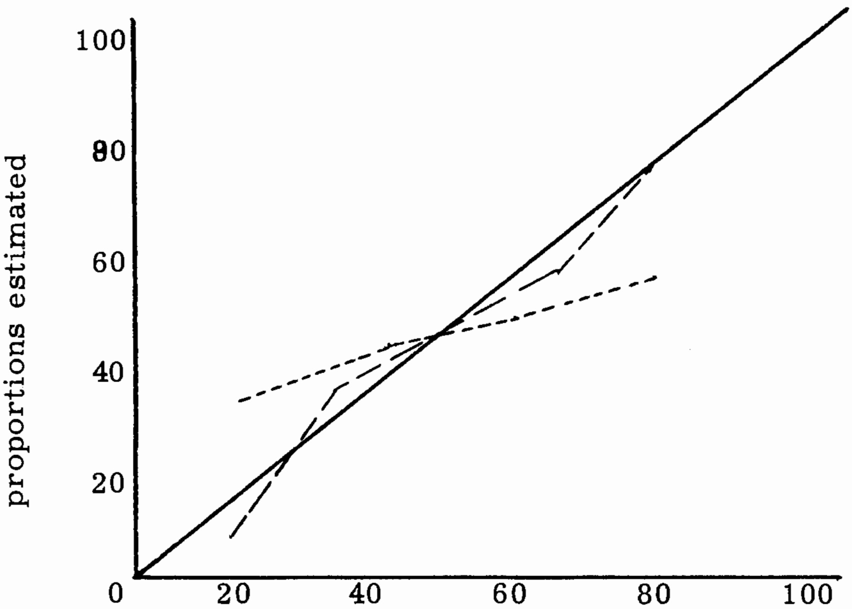
The discussion thus far has tended to de-emphasize the "two sets of stimuli: method used to determine the applicability of the additivity theorem and instead has concentrated on the significance of the findings regarding variability around the objective proportions of each category, and the significance of the findings from the "one set of stimuli" method. The purpose for such a de-emphasis stems from the two specific factors. First, the control group failed to respond as expected. Second, the "two sets of stimuli" method apparently invoked complexities which confounded the basic simplicity of the problem. Results observationally suggest that the subjects involved in the "two sets of stimuli" method interpreted each set of stimuli as a separate universe. And, if this was the case, the effect of unknown stimulus characteristics in each set could have increased the complexity due to the increased variety of stimuli shown.

The "one set of stimuli" method, though it had been assumed to be a relatively poor measure at the outset of this study due to the fact that it might encourage artificially summating the objective proportions, resulted in a significant variation from additivity for group V (17% - 83%) and thus merits a considerable amount of attention in further research. It should also be pointed out that the other "one set of stimuli" group (Group IV) tended to distort (though not significantly) the objective proportions in their probability estimates in the direction of a 50 - 50 split. This latter group thus conformed to additivity by a bias toward a 50% estimate. A similar effect was found in the "two sets of stimuli" method under the 36% - 64% proportions. This suggests that perhaps the additivity principle holds within limits, if objective proportions are close to 50%. However, if an event and its complements, vary markedly from 50% the additivity notion breaks down.

The relative uniqueness of the present study presents a difficulty in making comparisons with past studies dealing with the additivity principle. As was pointed out in the review of the literature all studies to this point have failed to ask for direct estimations of the relative frequency of an event and of the relative frequency of the complement of that event. Furthermore, most of the studies cited incorporated the confounding variables of risk, gain, utility, etc. However, there is one area in which plausible analogies can be drawn and that is in the area of overestimation and underestimation of objective proportions. The data from this experiment shows that a significant number of subjects underestimated the 17% level, overestimated the 36% level, and underestimated the 64% level.

At the 83% and 50% levels, subjects were relatively equal in their overestimations and underestimations, although at the 50% level many subjects estimated accurately (the first citing of the 50% level refers to those subjects who did not accurately estimate 50%). This evidence is not in total agreement with the findings of Howard and Komorita. Both experimenters state unequivocally that their studies indicate low probabilities are overestimated and high probabilities are underestimated. The evidence of the present study indicates that a more complicated relationship seems to exist.

The following graph illustrates the results of a typical overestimation, underestimation study and the results of the present study.



objective proportions

_____ objective frequency
---- findings from Komorita's study
— — findings from the present study

The considerations for further research, apart from those previously mentioned, derived from this study, are somewhat general due to the aforementioned uniqueness. One of the first considerations is concerned with the aspect of personal identification of the subjects involved in this study. Each subject was asked to write his name, the name of his instructor, and the date at the top of each questionnaire he was asked to fill out (Refer to Appendixes III, IV, V, and VI for examples of this particular point). The principal reason for doing this in the present study pertains to the ease of handling and ease of identification between the first and second questionnaires filled out by the same subject. The question does arise, however, as to the need and functional applicability of such a technique. For instance, some subjects expressed a great deal of concern over the fact that they were required to personally identify themselves. Feedback after several groups had been run was that they would have felt more at ease if they had been identified by some impersonal means. Thus, it is believed that most of the subjects tested would have responded with a greater degree of freedom and spontaneity had they not been asked to identify themselves in this manner. This factor is important in that every effort should be made toward making the conditions of a study of this nature analogous to environmental experiences and expectations.

A second area of consideration relates to individual versus group application of the methods described above. The question here being whether the variations from objectivity which were so apparent in this study would also occur under an individually applied method. This question is in keeping with the idea of providing as near normal experiences as possible.

A third area of concern relates to the particular percentage levels used and their relationship to commonly perceived levels of relative frequency. In other words, do individuals commonly perceive and interpret relative frequencies they are confronted with in terms of round figures such as 20%, 30%, etc.? Throughout the experiment subjects displayed a pronounced bias toward a 50% estimate. In fact, the results of such a bias was a major factor in their conformance to the additivity principle. There was evidence of a less marked bias toward estimates in round figures such as 10%, 20%, 40%, 60%, 80%, and so on. Confirmation of this question could be assessed by relatively straightforward experiments.

Finally, a good deal of consideration should be given to the area of confounding attributes of any stimulus to be utilized, in a study of this nature. Crawford (4:46) in a study dealing with the effect of certain stimulus characteristics upon subjective estimates of proportion found that the variables of heterogeneity, order, and atmosphere tended to produce estimates further from objective proportions rather than close to objective proportions. Stimulus materials used in the present study were highly complex. It was assumed, however, that such variables would be as operative in one category as in another and on the basis of this assumption no assessment of the subject's perception of the degree of heterogeneity, for example, was made.

In conclusion, this experimenter believes that the present study has given rise to a considerable degree of doubt regarding a strict application of the additivity theorem. The considerable variance displayed by all groups indicates the need for parameters of dispersion to be included in any theory of additivity. Furthermore, summed estimates of proportion appeared to be confounded by guessing biases, e. g. 50% - 50%, etc. Removal of or change of such biases may well effect changes in conformance to additivity.

CHAPTER V

SUMMARY OF STUDY

Six groups of college undergraduates, a total of 212 subjects, made subjective estimates of proportion in a series of experiments designed to explore the validity of the additivity principle. The stimulus series consisted of 2 x 2 slides of birds, modern ceramic art, pre-historic animals, and flowers, or simply birds and modern ceramic art. Each stimulus series consisted of seventy-two instances of two categories of stimuli, e.g. birds and modern ceramic art.

The specific hypothesis tested was: a significant proportion of subjective estimates of expected proportions of a stimulus class and its complement, from a universe composed of two stimulus classes, will total to some figure different from 100%.

The problem was approached by two relatively different methods. The first method used two sets of stimuli and estimates from each set were tallied and compared to a total relative frequency of 100%. The second method used only one set of stimuli and two estimates from the same set were tallied and compared to a total relative frequency of 100%.

The over-all results of this study failed to give support to the proposed hypothesis. However, due to gross deviations from the objective proportions and the degree of dispersion from additivity predictions, the results indicate that parameters of dispersion must be incorporated in any principle applying additivity to estimates of proportion.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Blommers, P., and E. F. Lindquist. Elementary Statistical Methods. Cambridge: The Riverside Press, 1960.

2. Cohen, J., and M. Hansel. Risk and Gambling: The Study of Subjective Probability. London: Longmans, Green and Co., 1956.

3. Coombs, C. H., and S. S. Komorita. Measuring Utility of Money Through Decisions. American Journal of Psychology, 1958, 71, 383-389.

4. Crawford, J. J. The Effect of Certain Stimulus Characteristics upon Quantitative Inductive Judgments. Unpublished Doctoral thesis, University of Washington, 1963.

5. Davidson, D., P. Suppes, and S. Siegel. Decision-making: An experimental approach. Stanford, Calif.: Stanford University Press, 1957.

6. Edwards, W. Behavioral Decision Theory. Annual Review of Psychology, 1961, 12, 473-498.

7. Good, I. J. Kinds of Probability. Science, 1959, 129, 443-447.

8. Howard, T. C. The Relation Between Psychological and Mathematical Probability. American Journal of Psychology, 1963, 76, 335.

9. Hunt, J. McV. Personality and the Behavior Disorders. New York: The Ronald Co., 1944.

10. Komorita, S. S. Factors which Influence Subjective Probability. Journal of Experimental Psychology, 1959, 58, 386-389.

11. Peatman, J. G. Introduction to Applied Statistics. New York: Harper & Row, 1963.

12. Ramsey, F. P. The Foundations of Mathematics and Other Logical Essays. London: Routledge & Kegan Paul, 1931.

13. Savage, L. J. The Foundations of Statistics. New York: Wiley, 1954.

14. Siegel, S. Level of Aspiration and Decision Making. Psychological Review, 1957, 64, 253-262.

15. Thrall, R. M., C. H. Coombs, and R. L. Davis (eds.). Decision Processes. New York: Wiley, 1956.

16. Von Neumann, J., and O. Morgenstern. Theory of Games and Economic Behavior. Princeton, New Jersey, 1947.

APPENDIX I

APPENDIX I

STIMULUS DETERMINATION PROCEDURE AND DATA

(PILOT STUDY)

(Instructions for Relatedness, Unrelatedness Determinant Group)

Today I want you to help me make some judgments on the relatedness of certain categories of paired items. I am going to ask you to judge the relationship between such items as shrubs and paintings, ceramics and birds, trees and diagrams of airplanes, flowers and names of birds, and so on. To give you an idea of what I mean I will show you some sets of slides which I want you to observe carefully. This is a trial run to acquaint you with the sort of pictures and names that you are to make judgments on. Watch the screen now as I show you the slides. There will be a blank between each pair of slides so you will know which categories I am referring to.

(Show Slides)

(Pass out rating charts)

You have before you now a rating chart upon which you are to judge the pairs of items which I will show you on the screen. We will follow the same procedure as we did in the trial run, only now when the blank appears on the screen I want you to judge the pair of slides you have just seen. You are to judge them according to the 7 point scale on the chart before you. As you can see the chart runs from

Highly Unrelated to Highly Related. Please put a check mark in the box most representative of the relationship as you see it. The slides will be shown according to the sequence of Roman Numerals listed in the left hand column of the chart. In other words, I will show a set of slides for Roman Numeral #1 and then show a blank. When the blank appears record your judgment of the relationship on the chart in one of the boxes following Roman Numeral #1. I will then show another set of slides for Roman Numeral #2 and follow it with a blank. You again record your judgment of the relationship in one of the boxes following Roman Numeral #2 and so on. Are there any questions?

The following items are to be judged on a relationship-non-relationship basis. You are to judge them according to the 7 point scale indicated on the chart below. Please put a checkmark in the box most representative of the relationship as you see it. Please give careful consideration to these comparisons.

Pairs of Items	Highly Un-related	Un-related	Slightly Un-related	Not sure of relation	Slightly related	Related	Highly Related
I							
II							
III							
IV							
V							
VI							
VII							
VIII							
IX							
X							
XI							
XII							

RESULT SHEET (29 Subjects)

The following items are to be judged on a relationship-non-relationship basis. You are to judge them according to the 7 point scale indicated on the chart below. Please put a checkmark in the box most representative of the relationship as you see it. Please give careful consideration to these comparisons.

Pairs of Items	Highly Unrelated	Unrelated	Slightly Unrelated	Not sure of relation	Slightly Related	Related	Highly Related
Flowers I Trees					5	21	3
Ceramics II Paintings	4	1	2	2	7	11	2
Trees III Airplanes	3	12		4	9	1	
Flowers IV Ceramics	6	11	6	3	3		
Names of Flowers V Flowers	1	1	1	1	4	11	10
Birds VI Names of Flowers	7	11	1	3	5	2	
Airplanes VII Diagrams of Airplanes					1	5	23
Birds VIII Names of Birds				1	2	15	11
Paintings IX Diagrams of Airplanes	7	8	2	1	7	4	
Names of Birds X Flowers	6	9	3	1	8	2	
Shrubs XI Trees					6	18	4
Birds XII Trees		3	3	2	13	7	1

APPENDIX II

APPENDIX II

INDIVIDUAL SUBJECTS' RECORDS (PILOT STUDY)

Name: ____________________ Instructor's name: ______________

Date: ______________

1. The first picture shown on the screen was a:
   flower ceramic

2. If the experiment were continued I would expect that the next picture to appear would be a:
   flower ceramic

3. The majority of ceramics seen in this experiment were:
   bowls cups plates figurines

4. The length of time each picture appeared on the screen was approximately:
   4 sec. 6 sec. 3 sec. 1 sec. 5 sec. 2 sec.

5. If this experiment were to continue, I would expect the number of pictures of flowers to be ______% of the total number of slides presented.

6. The last slide presented was a;
   flower picture ceramic picture

7. How many times did the picture of a Daisy appear in the experiment?
   1 4 3 7 5 6 2 0

8. A ceramic bowl was always followed by a ceramic plate:
   true false

9. The flower shown most often in this experiment was a:
   Rose Orchid Daisy Snapdragon

10. A flower not shown in this experiment was a:
    Lily Violet Pansy Sunflower

11. There were approximately how many ceramic bowls shown in this experiment?
    2 4 6 8 10 12 14 16 18 20

Name:________________________ Instructor's name_______________

Date: _______________

1. The first bird name to appear on the screen was:
   Bluejay, Pheasant, Eagle, Owl, Pigeon, Duck

2. The majority of bird names seen in this experiment began with the letter:
   F D P C S

3. The first bird picture presented on the screen was a:
   Robin, Finch, Sparrow, Hawk, Duck, Bluejay

4. If the experiment continued I would expect that the next slide to appear would be a:
   bird name bird picture

5. How many times did the picture of a robin appear in this experiment?
   1, 4, 3, 7, 5, 6, 2, 0

6. The length of time each slide was presented on the screen was approximately:
   4 sec. 6 sec. 3 sec. 1 sec. 5 sec. 2 sec.

7. The name <u>Eagle</u> appeared how many times in this experiment?
   0, 4, 6, 3, 1, 5, 2, 7

8. If this experiment were to continue, I would expect the number of pictures of birds to be _______% of the total number of slides presented.

9. The last slide presented was a:
   Bird picture Bird name

10. A bird neither shown nor mentioned in this experiment was a:
    Bluebird, Hawk, Sparrow, Bobolink, Chicken

11. The ratio of water birds (ex. seagulls, ducks, geese, etc.) to land birds (ex. robin, sparrow, finch, etc.) was:
    1-1 3-1 5-1 2-1 4-1

Name: ______________________ Instructor's name: ____________
Date: ____________

1. The first airplane picture to appear on the screen had how many engines?
   1 2 3 4

2. The first slide shown on the screen was an:
   airplane diagram airplane picture

3. If the experiment were continued I would expect that the next slide to appear would be an:
   airplane diagram airplane picture

4. How many times did the picture of a four-engine airplane appear in this experiment?
   1 2 3 4 5 6 7 8

5. The length of time each slide was presented on the screen was approximately:
   4 sec. 6 sec. 3 sec. 1 sec. 5 sec. 2 sec.

6. If this experiment were to continue, I would expect the number of pictures of airplanes to be ________% of the total number of slides presented.

7. The last slide presented was an:
   airplane diagram airplane picture

8. A twin engine plane was always followed by a single engine plane:
   true false

9. There were approximately how many slides showing more than 1 airplane?
   1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

10. The ratio of fighter planes to transport planes was approximately:
    1:1 2:1 3:1 4:1 5:1 6:1 7:1 8:1 9:1 10:1

11. How many airplanes were shown parked on the ground?
    1 2 3 4 5 6 7 8 9 10

Name: ______________________ Instructor's name: ____________

Date: ____________

1. The first flower name to appear on the screen was:
   Petunia, Daffodil, Rose, Tulip, Daisy

2. The majority of flower names seen in this experiment began with the letter:
   T R D P V

3. The first bird picture presented on the screen was a:
   Robin, Finch, Sparrow, Hawk, Duck, Bluejay

4. If the experiment continued I would expect that the next slide to appear would be:
   a bird picture a flower name

5. How many times did the picture of a robin appear in this experiment?
   1 4 3 7 5 6 2 0

6. The length of time each slide was presented on the screen was approximately:
   4 sec. 6 sec. 3 sec. 1 sec. 5 sec. 2 sec.

7. The name Tulip appeared how many times in this experiment?
   0 4 6 3 1 5 2 7

8. If this experiment were to continue, I would expect the number of flower names to be _____% of the total number of slides presented.

9. The last slide presented was a:
   bird picture flower name

10. A bird not shown in this experiment was a:
    Bluebird, Hawk, Sparrow, Seagull, Chicken

11. The ratio of water birds (ex. seagulls, ducks, geese, etc.) to land birds (ex. robin, sparrow, finch, etc.) was:
    1-1 3-1 5-1 2-1 4-1

APPENDIX III

APPENDIX III

STIMULUS DETERMINATION PROCEDURE AND DATA INSTRUCTIONS

Today I want you to help me make some judgments on the relatedness (or unrelatedness) of certain categories of paired items. This is not a test of intelligence; it is merely a preliminary study to help me determine what sorts of materials I might be able to utilize in a further study. I would sincerely appreciate it if you would simply look at each pair of slides and judge the relationship as you see it. To give you an idea of what I mean by paired items, I will show you some pairs of slides. This is a trial run to acquaint you with the sorts of picture pairs you are to make judgments on. Watch the screen now as I show you the slides. There will be a blank between each pair of slides so you will know which pair you are to make a judgment on.

Show slides

Pass out rating charts

You have before you now a rating chart upon which you are to judge the pairs of items which I will show you on the screen. We will follow the same procedure as we did in the trial run, only now when the blank appears on the screen I want you to place a check mark in the appropriate box on the chart to indicate the relationship as you see it. You are to judge each pair of slides according to the 7 point scale on the chart before you. As you can see, the chart runs from Highly Unrelated to Highly Related. I repeat, please put a check mark in the box most representative of the relationship, as you see it. The slides will be shown according to the sequence of numbers listed in the left hand column of the chart. In other words, I will show you a pair of slides for number 1 and then show a blank. When the blank appears, record your judgment of the relationship on the chart in one of the boxes following number one. I will then show another pair of slides for number two and follow it with a blank. You again record your judgment of the relationship in one of the boxes following number two and so on.

Are there any questions?

The following items are to be judged on a relationship, non-relationship basis. You are to judge them according to the 7 point scale indicated on the chart below. Please put a checkmark in the box most representative of the relationship as you see it.

Pairs of Items	Highly Un-related	Un-related	Slight-ly Un-related	Not sure of relation	Slight-ly Related	Related	Highly Related
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							

20							
21							
22							
23							
24							

RESULT SHEET

Pairs of Items	Highly Un-related	Un-related	Slightly Un-related	Not sure of relation	Slightly Related	Related	Highly Related
Modern Ceramics 1 Flowers	8	9	1	2	2	1	
Flowers 2 Seashore Animals	5	10	1	2	3	2	
Modern Ceramics 3 Pre-Historic Animals	7	9	2		3	2	
Flowers 4 Flower Names	2	2	1	2	5	4	5
Airplane Diagrams 5 Architecture					1	8	16
Birds 6 Modern Ceramics	10	4	1	3	3	2	
Prehistoric Animals 7 Architecture	4	4	3	1	4	7	
Flowers 8 Airplane Diagrams	19	4					
Prehistoric Animals 9 Seashore Animals	1	5	3	1	9	4	
Birds 10 Airplanes		1		1	13	7	1
Airplane Diagrams 11 Seashore Animals	9	9	1		2		
Modern Ceramics 12 Airplane Diagrams	13	4	2		3	1	
Birds 13 Bird Names	1	3			9	8	2

Birds	14	Architecture	3	6	2	2	4	4	2
Flowers	15	Flowers		1			1	10	11
Modern Ceramics	16	Seashore Animals	6	9	1	2	4	1	
	17								
Birds	18	Prehistoric Animals	6	6	1	1	6	2	1
Modern Ceramics	19	Architecture	2	4	6	1	9		
Airplane Diagrams	20	Birds		2	1	1	13	5	1
Birds	21	Flowers	5	10	1		6	2	
Birds	22	Birds	2					8	12
Bird Names	23	Airplane Diagrams	2	2		1	13	3	1
Flowers	24	Prehistoric Animals	7	7	2	1	4		

APPENDIX IV

APPENDIX IV

INDIVIDUAL SUBJECTS' RECORDS--TWO SETS OF STIMULI

Name:______________________ Instructor's Name:________________

Date:________________

1. A picture of a __________________ was the first picture shown on the screen.

2. The length of time each picture appeared on the screen was approximately ______________ seconds.

3. The picture of a body of water appeared in ______________% of the total number of pictures shown.

4. A bowl picture was <u>never</u> followed by a plate picture.

   TRUE FALSE

5. I would expect the next picture to be a __________________ if this series of slides were continued.

6. A picture of a __________________ was not shown in this series of pictures.

7. If this series of pictures were to continue, I would expect the number of pictures of birds to be __________________% of the total number of pictures shown.

8. How many times did the picture of a woman appear in this experiment? ________________

9. The last picture presented was the picture of a ______________.

10. During the instructions, the experimenter mentioned the word __________________ at least three times.

Name: ______________________ Instructor's Name: ______________
Date: ______________

1. A picture of a ____________________ was the first picture shown on the screen.

2. The length of time each picture appeared on the screen was approximately ______________ seconds.

3. The picture of a body of water appeared in _______% of the total number of pictures shown.

4. A Robin picture was <u>never</u> followed by a Bluejay picture.

   TRUE FALSE

5. I would expect the next picture to be a ______________ if this series of slides were continued.

6. A picture of a ________________ was not shown in this series of pictures.

7. If this series of pictures were to continue I would expect the number of pictures of ceramics to be _____________% of the total number of pictures presented.

8. How many times did the picture of a woman appear in this experiment? _______________

9. The last picture presented was the picture of a ____________.

10. During the instructions, the experimenter mentioned the word ___________________ at least three times.

Name:__________________________ Instructor's Name:_______________
Date:_______________

1. A picture of a ____________________ was the first picture shown on the screen.

2. The length of time each picture appeared on the screen was approximately ___________ seconds.

3. The picture of a body of water appeared in ____________% of the total number of pictures shown.

4. A Lily picture was <u>never</u> followed by a Daisy picture.

   TRUE FALSE

5. I would expect the next picture to be a ______________if this series of slides were continued.

6. A picture of a ____________________ was not shown in this series of pictures.

7. If this series of pictures were to continue I would expect the number of pictures of prehistoric animals to be __________% of the total number of pictures presented.

8. How many times did the picture of a woman appear in this experiment?_______________

9. The last picture presented was the picture of a______________.

10. During the instructions, the experimenter mentioned the word __________________ at least three times.

Name:________________________ Instructor's Name:______________

Date:______________

1. A picture of a ________________ was the first picture shown on the screen.

2. The length of time each picture appeared on the screen was approximately ___________ seconds.

3. The picture of a body of water appeared in ____________% of the total number of pictures shown.

4. A Dinosaur picture was <u>never</u> followed by a Saber Tooth Tiger picture.

TRUE FALSE

5. I would expect the next picture to be a ___________ if this series of slides were continued.

6. A picture of a ____________ was not shown in this series of pictures.

7. If this series of pictures were to continue I would expect the number of pictures of flowers to be ___________% of the total number of pictures presented.

8. How many times did the picture of a woman appear in this experiment?_______________

9. The last picture presented was the picture of a _____________.

10. During the instructions, the experimenter mentioned the word ____________________at least three times.

APPENDIX V

APPENDIX V

INDIVIDUAL SUBJECTS' RECORDS--ONE SET OF STIMULI

Name:______________________ Instructor's Name:_____________

Date:_____________

1. A picture of a ____________________ was the first picture shown on the screen.

2. The length of time each picture appeared on the screen was approximately _____________ seconds.

3. The picture of a body of water appeared in _____________% of the total number of pictures shown.

4. A bowl picture was <u>never</u> followed by a plate picture.

   TRUE FALSE

5. I would expect the next picture to be a ___________________ if this series of slides were continued.

6. A picture of a __________________ was not shown in this series of pictures.

7. If this series of pictures were to continue, I would expect the number of pictures of birds to be ____________% of the total number of pictures shown.

8. How many times did the picture of a woman appear in this experiment?

9. The last picture presented was the picture of a _____________.

10. During the instructions, the experimenter mentioned the word ___________________ at least three times.

Name:______________________ Instructor's Name:____________
Date:____________

1. The length of time between each picture was approximately _______________ seconds.

2. A picture of a ___________________ was the second picture shown on the screen.

3. I would expect the last picture to be a _______________ if ten more pictures were to be shown.

4. A picture of a _______________ was shown four times in this series of pictures.

5. How many times did the picture of a boy appear in this experiment? ___________________

6. During the instructions, the experimenter had his __________ hand in his hip pocket.

7. A Robin picture was never followed by a Bluejay picture.

TRUE FALSE

8. The picture of a bowl appeared in ____________% of the total number of pictures shown.

9. The next to the last picture presented was the picture of a ____________________.

10. If this series of pictures were to continue, I would expect the number of pictures of ceramics to be ____________% of the total number of pictures presented.

APPENDIX VI

APPENDIX VI

INDIVIDUAL SUBJECTS' RECORDS -- INQUIRY SHEET

Name:______________________ Instructor's Name:______________

Date:______________

The following questions are related to the experiment in visual perception conducted earlier this week by Mr. Holmberg. As you will recall, the experiment involved showing two sets of pictures with a questionnaire following each set. Will you please answer the questions below as accurately as you can, as they will be of significant value in determining the results of the experiment.

1. The <u>first</u> set of pictures shown were pictures of:

2. As I watched the <u>first</u> set of pictures, the following things seemed important to look for and remember.

3. To me it seemed, after watching the first set of pictures, the main types of pictures shown were:

4. The <u>second</u> set of pictures shown were pictures of:

5. As I watched the second set of pictures, the following things seemed important to look for and remember.

6. To me it seemed, after watching the second set of pictures, the main types of pictures shown were: