1967

Conservatism in a Bayesian Probability Situation as a Function of the Sex of the Subject

Paul Frederick Miller
Central Washington University

178

CONSERVATISM IN A BAYESIAN PROBABILITY SITUATION AS A FUNCTION OF THE SEX OF THE SUBJECT

A Thesis

Presented to

the Graduate Faculty

Central Washington State College

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

Paul Fredrick Miller

August 1967

APPROVED FOR THE GRADUATE FACULTY

Eldon E. Jacobsen, COMMITTEE CHAIRMAN

Roger Stewart

Richard B. Morris

ACKNOWLEDGMENTS

Many people were instrumental in the designing, running, and reporting of this experiment. The efforts of Dr. Eldon Jacobsen with respect to the final presentation of this material was very helpful. The efforts of Mr. Richard Morris on the statistical analysis of the data were both helpful and necessary for accurate computation and meaningful presentation of the results. I also wish to express my appreciation to Dr. Lillian Lauer who served as a committee member until leaving in June. Her efforts and those of Dr. Roger Stewart, who agreed a week before oral examinations to sit on the committee, are represented in this paper.

I would also like to express my gratitude to Mr. Jerry Shulenbarger for his suggestions and the hours he voluntarily spent in aiding me with the statistical analysis.

Also, I wish to thank Mr. Richard Giroux for his suggestions pertaining to the design of the experiment and for acting as an assistant during the testing session.

Lastly, I want to extend my sincere thanks to my wife, Carol; through her own personal efforts and sacrifices she provided me with the opportunity to conduct this project. For this, I am most grateful.

ABSTRACT

The present study was designed to examine whether the conservatism present in a Bayesian probability situation could be partially attributable to the sex of the subjects performing the task. The experimental design required that the subjects estimate the probabilities of occurrence of two independent events. They were then given an opportunity to revise their estimates as additional information was experimentally introduced into the situation. These estimates were compared to estimates calculated from Bayes' theorem.

The results of this experiment failed to support the hypothesis that the female subjects would exhibit more conservatism than the male subjects.

Suggested explanations for these results centered around the nature of the task performed and the characteristics of the experimental situation.

Suggestions for further research were both theoretical and methodological in nature.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	Problem	1
II.	Method	19
	Subjects	19
	Apparatus	19
	Procedure	20
III.	Results	23
	Relation of Odds to Probabilities	24
	Measure of Estimate Revision	24
	Likelihood Ratio	26
	Calculation of Likelihood Ratios	26
	Relation of SLLR to BLLR	29
IV.	Discussion	39
V.	References	51
VI.	Appendices	54

LIST OF TABLES

TABLE PAGE

1. Bayesian Likelihood Ratios and Corresponding Probabilities per Trial 28
2. Mean Accuracy Ratio and Standard Deviation Per Trial 30
3. Frequency of Stimulus Cues and Correct Hypothesis per Trial 32
4. Analysis of Variance Summary Table 35
5. Observed Accuracy Ratios for Female Subjects . . 36
6. Observed Accuracy Ratios for Male Subjects . . . 37
7. Observed Accuracy Ratios for Female Subjects . . 38

LIST OF FIGURES

FIGURE PAGE

1. A Comparison between the Mean Observed Accuracy Ratios per Trial and a Theoretical Accuracy Ratio of 1.00 31

Problem

The purpose of the present study is to examine the phenomena of "conservatism" in probabilistic inference situations. Specifically, this study will investigate whether the amount of "conservatism" present in this type of situation differs according to the sex of the subject performing the task. The normative model to be used is Bayes' theorem. This model prescribes how people should behave in a situation which requires them to estimate the probability of occurrence of specific events. "Conservatism" will be defined as failure of the subjects to perform in a manner consistent with this model.

The importance of probability theory in statistics originally centered around the development of a "statistical method." The necessity for this type of method resulted from an increased interest in statistics during the middle part of the nineteenth century and consisted primarily of two aspects. The first aspect of this method was the development of a statistical method of investigation. Secondly, was the need for a statistical method of inference.

The statistical method of "investigation" consisted primarily of investigating a phenomena and measuring it by various sorts of averages. The emphasis was upon studying variability instead of eliminating it.

The second aspect of the statistical method, inference, was concerned with the relationship between theory and observation; theory suggesting what to observe and observation correcting theory. It was shown, however, that an interplay between observation and theory was needed which would make observations more relevant and meaningful with respect to particular theories. The development of probability theory was a natural consequence of this dilemma. Employing the characteristics of this theory, the scientist could make "mathematical guesses" with respect to observations made in the long run.

Probability theory was extended in the work of R. A. Fisher during the 1920's and culminated in the work of Neyman and Pearson during the 1930's. Through the careful development of appropriate inferential concepts and procedures, a stability or orthodoxy in the theory and practice of probability had developed by 1940. This orthodoxy is presently referred to as the classical interpretation of probability.

One of the primary constructs of the classical interpretation is that of <u>relative frequency</u>. Basically, this principle states that "For equally likely elementary events, the probability of an event <u>A</u> is its relative frequency in the sample space" (Hayes, 1963, p. 56). Fisher and Neyman both insisted that "all probabilities referred to should have a basis or interpretation in terms of frequencies" (Anscombe, 1964, p. 160). The probability of the data, therefore, denotes nothing other than its relative frequency to be expected in the long run. The primary purpose of statistical analysis is to summarize the data, collected by a scientist, relative to a particular phenomena, without considering how the particular method of analysis used might influence the beliefs and actions of the person employing them.

One problem with the "frequentist's" position is that it offers no direct theory of inference or of scientific procedure. It is concerned with how to make data meaningful, not how to meaningfully collect data.

A second problem with this position is related to the rules that should be used in relating mathematical properties to observations in the real world. Frequencies are not actual probabilities but only estimates of probabilities. The scientist is forced to work within the confines of finite occurrences; therefore, uncertainties are always present. These uncertainties cannot adequately be explained or described simply by means of probabilities.

Much of the criticism directed against the classical school has been a result of the development of statistical decision theory. The importance of this development centered around the work of A. Wald and L. J. Savage. Wald's efforts were primarily directed towards the scientist who was considered to be a rational decision maker. The work of Savage was prominent in the rehabilitation of subjective probability and utility.

The concept of utility was originally introduced by Danial Bernoulli (1738). He insisted that the value of an object could not be determined by its price but was rather a function of the utility it yielded. Values are therefore relative to the people assigning them and are inversely related to the quantity of the object the person already possesses. Thus, they are not constant for all people but vary depending upon the particular situation.

Recently the importance of this concept has been realized by psychologists. Miller (1964) has pointed out that there is a definite need for the development of a theory of normative application. This theory must be distinguished from descriptive theory and is primarily derived from logic instead of observation. Deductive logic tells us how we ought to think or behave, not how we do behave. The logician therefore specifies how people ought to think, not what they ought to think. Since utility is relative to the person assigning value to an object, there is a need for a normative model of utility.

One of the leading proponents of this idea has been a "school" which views men as probabilistic information processors. This movement, known as Bayesian statistics, operates from the principle that "probability is orderly opinion, and that inference from data is nothing other than the revision of such opinion in the light of relevant new information" (Edwards, Lindman, and Savage, 1963). Bayes' theorem is said to be the appropriate normative model for decision making situations. It prescribes how a person should behave if he is to be consistent with himself. As pointed out by Anscombe (1964, p. 161), "It is a theory of consistency of the person's body of beliefs or preferences; there is no consideration of ethics." His observed behaviors are characteristic of description only.

The major contribution of the Bayesian movement has been the concept of "personal probability." Personal probabilities are " . . . ideally consistent opinions, and conform to the axioms of probability theory . . ." (Peterson, Schneider & Miller, 1965). These probabilities are to be distinguished from "subjective probabilities" which are defined as " . . . a weight attached to an event by an individual that indicates the strength of his expectation that the event will occur" (Stilson, 1966, p. 79). Subjective probabilities may also be interpreted as the value that the occurrence of an event has for the individual; it is the utility of the event with respect to the person.

In order for personal probabilities to be both consistent and orderly opinion, it is necessary that they obey the axioms of probability theory. No such restriction is placed on subjective probabilities. The two primary axioms that personal probabilities must conform to are:

$$0 \leq p(A) \leq p(S) = 1.00 \qquad (1)$$

$$p(A \cup B) = p(A) + p(B) \qquad (2)$$

where S is the universal event; A and B are any two incompatible or nonintersecting events; and AUB is the event that A or B or both A and B are true.

Personal probabilities exist prior to the actual occurrence of an event. In an experimental situation, personal probabilities may reflect the degree of confidence the experimenter has with respect to his hypothesis. The Bayesian statistician insists that such probabilities are present prior to the collection of data. This is due to the fact that experimentors are usually not naive to the particular problem they are studying. Background literature and their own professional skills are available to them. Thus their opinions are different from the subjective opinions made by people in everyday situations. The insistence of these prior probabilities is the primary distinguishing feature between the Bayesian movement and the classical school.

One difficulty entailed in the use of subjective probabilities is that persons who are equally familiar with the situation may disagree on the probability that should be assigned to a given event. These results could be expected, however, if it is kept in mind that subjective probabilities involve more than the attitude of the person towards the event; they are also reflective of the gamble or risk he is willing to take with respect to the occurrence or non-occurrence of the event. This addition of "personal" into probability theory is a second contribution of the Bayesian movement.

One problem raised with respect to Bayes' theorem as a normative model is that when subjective and objective probabilities are combined, the assumption must be made that subjective probabilities combine according to the same rules of mathematical probability theory as the objective probabilities do. Savage (1962) has demonstrated that " . . . subjective probabilities have the same mathematical properties as objective probabilities; but there the resemblance ends." Upon examination of the degree to which subjective probability judgments conform to the Bayesian model, the present study assumes that subjective probabilities do obey the mathematical rules of probability theory. Two requirements for the application of Bayes' theorem and the rules this study assumes subjective probabilities obey are:

(1) The sum of all outcome probabilities in a given sample space must be unity; and (2) The sum of the probabilities of the outcomes combined in an event equals the probability of that event.

Bayes' theorem is contingent upon conditional probability theory. Simply stated, if an event <u>A</u> has a probability of occurrence assigned to it, p(A), and if the scientist assigning this probability is given some additional information relevant to <u>A</u>, such as the occurrence of event <u>B</u>, then knowledge of the occurrence of <u>B</u> must be considered when estimating p(A). The proper formula is:

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \qquad (3)$$

where p(A|B) is the posterior probability of <u>A</u> given that <u>B</u> has occurred; the p(A∩B) is the probability of <u>A</u> and <u>B</u> occurring jointly; and the p(B) is the prior probability of <u>B</u> occurring.

Simple algebraic transformation of equation 3 results in the basic form of Bayes' theorem:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \qquad (4)$$

where <u>A</u> is any hypothesis; the <u>B</u> is the data; the p(A) is the prior probability of the hypothesis under consideration being correct; the p(B|A) is the probability of the data being associated with the hypothesis; the p(B) is the prior probability of the data occurring; and the p(A|B) is the posterior probability that the hypothesis is correct, given the data associated with it.

The most widely used and appropriate technique for examining how people function as probabilistic information processors is to introduce them into an experimental situation in which they are required to choose between alternative events. By submitting relevant information into the situation, the amount of subjective revision can be compared to the normative model. Failure of subjects to behave in a manner consistent with this model is what has been defined as conservatism.

The most consistent result of this type of experiment has been the presence of the conservative effect (Phillips, Hays, and Edwards, 1965). These investigators asked subjects to decide whether detected aerial activity was indicative of enemy attack, friendly activity, meteor shower, or an enemy attempt to "spoof" the surveillance system. After recording their decision, the subjects were given additional information about more aerial activity. They were then asked to record another estimate with respect to the four events under consideration. This procedure was followed for thirty-two presentations of information. The results of their study demonstrated the small amount of subjective revision made by the subjects when Bayesian probabilities showed considerable change.

These findings indicate that human beings are fallible when it comes to processing information; that the 'average' person is not able to utilize all the information available to him. Also these findings indicate that subjects are not consistent in their use of the information they are processing.

Peterson and Miller (1965) attempted to compensate for the conservative effect by manipulating both the prior odds in favor of a hypothesis and the theoretical impact of the datum. A large number of dice, distinguished by different numbers of black versus white sides, were placed in an urn in varying proportions. For each trial the die with the greater number of black sides was termed H_a and the die with the greater number of white sides was termed H_b. The experimenter drew a single die from the urn and asked the subject to set prior probability estimates for H_a and H_b. He then rolled the die from a cup and informed the subject which side turned up. On the basis of this information the subject revised his estimates about which die had been rolled.

The theoretical impact of the data was manipulated by drawing either H_a or H_b from the urn. The prior probability of the hypothesis was manipulated by varying the proportion of white sides to black sides in the urn. These investigators failed to reduce the conservative effect; in fact, as the difference between Bayesian prior and posterior probabilities increased, conservatism increased.

Schum (1966) presented to subjects six classes of data describing the characteristics of a military deployment. The subjects were required to decide whether this information represented war, a forthcoming attack, or simply a maneuver. Conditional non-independency of data was controlled by the prior selection of the particular information to be presented. It was found that the use of non-independent data failed to reduce the amount of conservatism present.

In a series of three experiments, Phillips and Edwards (1966) examined the effect of the diagnostic value of the data, payoffs, and response modes on conservatism. In the first experiment the diagnostic value of the data was reflected by the likelihood ratio, a ratio describing the occurrence of the data dependent upon the separate probabilities of the two hypotheses under consideration, as a function of the numerical difference between the cues sampled. The subjects were told to imagine ten paper bags, each bag containing one hundred poker chips with red chips prodominant in _r_ bags and blue chips prodominant in _10-r_ bags. They were then shown one bag and told that it had randomly been drawn from the ten. After being asked to estimate whether an _r_ or _10-r_ bag had been chosen, the experimenter explained that with no other information given, the best estimates that could be made were r/10 and 10-r/10. If the subjects estimates differed from the suggested estimates, they were asked to change them. All subjects, therefore, started with the same probability estimates. The experimenter then drew twenty chips from the bag and displayed them, one at a time, to the subjects. After each chip was displayed, the subjects were asked to revise their previous estimates. This procedure was repeated twenty-four times.

Results of this experiment indicated that as data became more diagnostic (likelihood ratio approached one), subjects became more Bayesian, however conservatism was present for all the subjects.

In the second experiment the effect of motivation on conservatism was investigated. The design and procedure for this experiment was basically the same as the first experiment. The three primary differences between them were that the proportion of chips in each bag was a constant 70-30; all the sequences in experiment II started with probabilities of 50-50; and only twenty sequences were presented.

Three experimental groups and one control group were used for their second experiment. Experimental Group I received payoffs that were logarithmically related to their subjective estimates; Experimental Group II received payoffs quadratically related to their estimates; Experimental Group III received payoffs linearally related to their estimates; and the Control Group received no payoffs.

Following each sequence of twenty draws of chips, the subjects were told which hypothesis was correct. Each subject then spun a spinner which selected one of his twenty pairs of estimates. His estimate for the correct hypothesis on the chosen pair determined the amount of payoff. The subjects were aware throughout the experiment of the number of "points" they were accumulating. Following the experiment their points were converted into money.

The results of this experiment showed that payoffs had a positive effect on the amount of conservatism present. Experimental Group III (linear payoffs) exhibited less conservatism than the other three groups. The authors concluded that in probabilistic inference situations where small between subject variance is desired, payoffs should be used.

Experiment III was conducted to test the effect that the method of recording estimates had on conservatism. The design and procedure were basically the same as the first experiment. Four different methods of recording estimates were tested. Group I recorded their estimates by placing metal washers in two vertical troughs (discrete method). Group II verbalized their estimates while an experimenter recorded them. Groups III and IV were called the logarithmic continuous groups. Group III recorded estimates by setting a sliding pointer on a scale of odds spaced logarithmically. Group IV recorded their estimates on a scale where the size of the intervals was determined by converting the probabilities to odds and scaling the odds. None of the groups were told which hypothesis was correct and no payoffs were received.

Results of this experiment indicate that the method of recording estimates does not eliminate conservatism. The estimates for the verbal and logarithmic odds groups were less conservative than Group I; however, all subjects consistently failed to significantly approach the normative model.

As a result of these three experiments, Phillips and Edwards concluded that the failure of subjects to extract from the data all the certainty that is experimentally available is consistent and orderly and may reflect a general limitation on human ability to process information.

Due to the amount of experimental manipulation employed in these three experiments, the present investigater is hesitant to accept them as sufficient evidence for the above conclusion. It is possible that conservatism was encouraged in these experiments and that in the absence of the different manipulatory techniques the amount of conservatism present would have been of a smaller magnitude.

Motivation in a probabilistic inference situation may be influenced by variables other than payoffs. A large quantity of information may be overwhelming enough to prevent optimal performance by the subjects. Likewise, small quantities of information may be ineffective as motivators thus resulting in conservative behavior.

This problem was investigated by Peterson, Schneider, and Miller (1965). In an experiment involving the drawing of black and white marbles from one of two beakers, the effect of sample size, quantity of information presented, was investigated. For Group I the marbles were drawn one at a time; for Group II they were drawn four at a time; for Group III they were drawn twelve at a time; and for Group IV they were drawn forty-eight at a time. Following each draw the subjects were required to estimate which beaker was being used. They were subsequently told which hypothesis was correct after a total of forty-eight marbles had been drawn.

Employing this design results in only three of the groups making probability revisions. Group IV received all forty-eight marbles in one draw; therefore, revision was not required. Since Bayes' theorem is the normative model for estimate revision, it is questionable whether the results of this group can be analyzed with respect to it.

The results of this study indicate that sample size influences the conservative effect. The larger the sample size the more was the amount of revision; however, it occurred at the expense of revision accuracy, or increased conservatism. Size of the sample and revision accuracy were therefore found to be inversely related. These authors concluded that subjects are able to work with only a limited amount of information if they are to maintain revision accuracy.

Kogan and Wallach (1964) administered a five-hour battery of tests to 114 male subjects and 103 female subjects to investigate the various aspects of decision making and risk taking. The primary purpose of this research was " . . . an attempt to look at human thinking and problem solving from the point of view of the risks, potential costs, and potential gains that may face the individual as he proceeds in his efforts" (Kogan, et al., p. 1). Decision making involves alternatives and the avoidance or acceptance of various alternatives are likely to be important ingredients in the thinking process. The results of this study are to be accepted as descriptive and not explanatory of decision making processes; statements of explanation await future research.

The tests administered in the Kogan and Wallach study primarily involved measures of extremeness of judgment, extremeness of self rating, the personality factors of anxiety and defensiveness, and various measures of decision making strategies.

The findings of this research are relevant to the present paper for four reasons: (1) Probabilistic inference situations require decision making on the part of the subjects; (2) This type of situation also requires the processing of additional information and the opportunity for subjective estimate revision; (3) Subjective estimates are reflective of the motivational aspects operative in the situation; (4) The present study also involves a situation in which the subjects have no control over the problem outcome. Of particular importance to the present paper are the results which report differences found between the male and female subjects. Since the effect of sex as a moderator variable in decision making situations has received little attention in the past, Kogan and Wallach reported their results separately by sex throughout their paper.

In general, it was found that males exhibited greater confidence of judgment than females. This was evidenced by the fact that they were more sensitive with respect to when certain strategies ceased to payoff or be effective. The male subjects appeared more alert to the subtle distinctions between their specific task requirements and as a result adopted more risky but less extreme strategies than the female subjects. These results were especially evident in situations where the subjects had no personal control over the problem outcomes.

These investigators concluded that behavior in decision making situations is attributable primarily to failure-avoidance in females and image-maintenance in males. Also, manifest anxiety is related to conservatism of strategy preferences in males only, while rigidity yields a much more pervasive pattern of relationships with conservatism in females. In females, therefore, it was found that there was a direct relation between impulsiveness and risk taking under these hypothetical conditions.

The present study is concerned with whether the above described results are operative in an experimental situation employing two specific experimental conditions. The first condition is that subjects be required to choose between two specific alternative events and that by introducing additional information they be given the opportunity to revise their subjective estimates or opinions. The second condition is that the subjects have no control over the outcome of the problem.

The hypotheses of the present investigation are therefore twofold. First, it will be shown that conservatism is present when comparing the performance of human subjects to Bayes' theorem in this type of situation; secondly, it will be demonstrated that female subjects exhibit more conservatism than male subjects.

Method

Subjects

Forty-three volunteer students, 28 females and 15 males, from Psychology of Adjustment classes at Central Washington State College served as *S*s.

Apparatus

The stimulus cues consisted of 1000 cardboard squares (1"x1") and 1000 cardboard right triangles (1"x1"x1.4"). These cues were distributed in two wooden boxes (8"x8"x8") in the proportion of 600 squares and 400 triangles in one box, Hs, and 600 triangles and 400 squares in the other box, Ht. Large numbers of figures were used so that, with continual mixing, sampling without replacement could be considered a reasonable approximation of sampling with replacement.

Both boxes were painted with red enamel paint and had a yellow square or triangle painted on one side of them. The particular figure represented that the highest frequency of this geometric figure was contained in this box. Other apparatus consisted of standard overhead projector (see reference); a standard Radiant Wall movie screen; one

Wollensak tape recorder, Model T-1616 Electronic Control; one reel of Scotch Magnetic Tape, Tartan Series 141; two transparencies, one divided into 18 rectangles (Appendix A), and the other with an example of the Ss recording sheets drawn on it (Appendix B); a standard classroom table; two pair of eyebrow tweezers; a black wooden barrier (33"x12"x 15"); and Ss response sheets for each trial (Appendix C).

Procedure

The experimental design required the presentation of certain information and an evaluation of the effect upon revision of subjective probabilities as a result of this information. The two boxes with varying proportions of squares and triangles were the alternative events (hypotheses) available to the subjects (Ss). Each trial consisted of the random selection of 36 cues from one of the boxes. After each trial, the Ss were shown from which box the cues had been drawn.

Two assistants aided the experimenter (E) in conducting the experiment. One of the assistants, voluntarily chosen from among the Ss, was responsible for drawing the sample of stimulus cues from the box. This assistant was chosen in this manner to eliminate experimenter bias. The second assistant, pre-arranged by E, recorded the stimulus cues as they were drawn from the box. This information was necessary for computation of Bayesian probabilities.

There was a total of ten trials. Each trial consisted of drawing 18 pairs of stimulus cues from one of the two boxes. Following each draw of a pair of cues, the Ss were required to make two estimates. Their first estimate was with respect to Hs being the correct hypothesis. Their second estimate was with respect to Ht being correct. To insure that their subjective estimates obeyed the same mathematical rules as Bayesian probabilities, the Ss were told that their estimates for Hs and Ht on any one draw must sum to unity.

The experiment began by Ss receiving taped instructions (Appendix D) as to what was meant by probability estimates, the nature of the experimental procedure, and the nature of their specific task. Each S had a copy of the instructions. Since the instructions themselves could encourage conservatism, it was important that they were neither directive nor suggestive as to what estimates should be made. A pilot study of 27 Ss was run to insure that the instructions were completely clear as to the nature of the experimental situation.

Trial 1 began by E flipping a coin to determine which box would be used. The boxes had been placed behind a wooden barrier which was located on a table situated between the Ss and the projector. After determining which box would be used, E removed this box from behind the barrier and handed it to the subject assistant seated next to the projector. Since the box was visible to the Ss as it was handed to the assistant, it was necessary to conceal the geometric figure from their vision. After handing the box to the assistant, E sat on the opposite side of the projector.

With a pair of tweezers in each hand, the assistant randomly drew a pair of figures out of the box. These cues were handed to E who placed them on the first square of the transparency on the projector. This method of displaying the cues was decided on so that each S could clearly see which cues had been drawn. The cue in the assistant's right hand was placed on the right side of the square. All triangle cues faced the same direction.

Following this drawing of the first pair of cues, the Ss were required to make probability estimates with respect to Hs and Ht. These estimates were recorded on provided record sheets. After a period of five seconds, the subject assistant drew a second pair of cues from the box. These cues were displayed by E in the second square of the transparency. The Ss were then required to make two more estimates. This procedure continued until 36 cues (18 pairs) had been drawn from the box.

After the drawing of the last pair of cues, E picked up the box and placed it in front of the barrier with the geometric figure exposed to the Ss. E then returned the box behind the barrier. Following this the boxes were shuffled by E to prevent the Ss detection of the correct box by placement. E then flipped a coin to determine which box would be used for trial 2. All ten trials followed this same procedure.

Since the estimates made by the Ss offered no information with respect to the particular approach to the problem adopted by them, the Ss were asked to fill out a questionnaire following the testing session (Appendix E). This questionnaire primarily investigated their attitude towards the task and the particular approach they used to determine whether Hs or Ht was the correct hypothesis.

Peterson, Schneider, and Miller (1965) have demonstrated that Ss are able to process only limited amounts of information if they are to attain revision accuracy. To maintain accuracy in the present study, the cues were sampled in pairs. The total number of cues drawn, 36, was decided upon as a result of the functional area of the overhead projector. Using this procedure, Ss made a total of 180 estimates.

Results

The present experiment required the Ss to make subjective estimates about the occurrence of two independent events. These estimates were then compared, at the end of each trial, to estimates derived from Bayes' theorem. The appropriate statistic for making this comparison was the accuracy ratio. This ratio is a comparison between the subjective odds in favor of Hs and the corresponding Bayesian odds in favor of the same hypothesis.

<u>Relation of Odds to Probabilities</u>

The odds $\Omega_o(H_s)$ are related to the probability of H_s by the following equation:

$$\Omega_o\,(H_s)\;1.00\text{-}p(H_s) = p(H_s) \qquad (5)$$

Simply stated: if the probability of an event H_s is equal to $p(H_s)$, the odds in favor of the event are $p(H_s)$ to $1.00\text{-}p(H_s)$. Therefore odds and probability are related to each other as follows:

$$\Omega_o\,(H_s) = \frac{p(H_s)}{1.00\text{-}p(H_s)} = \frac{p(H_s)}{p(H_t)} \qquad (6)$$

Part of hypothesis testing in the Bayesian system is finding the posterior probability of H_s, $(H_s|D)$, or equivalently, finding the posterior odds $\Omega_1(H_s|D)$ in favor of H_s.

<u>Measure of Estimate Revision</u>

Bayes' theorem is a consequence of conditional probability theory and is therefore appropriate for calculating the probability of Hypothesis <u>H</u> as a result of information provided by the occurrence of Datum <u>D</u>. The basic form of Bayes' theorem presented in the introduction of this paper was:

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)} \qquad (4)$$

where $p(H|D)$ is equal to the posterior probability of H; $p(D|H)$ is the probability of the data given that H has occurred; $p(H)$ is the prior probability of H; and $p(D)$ is the prior probability of Datum D. An equation for two hypotheses which was relevant to this paper was obtained by dividing Equation 4 for Hs by Equation 4 for H_t:

$$\frac{p(H_s|D)}{p(H_t|D)} = \frac{p(D|H_s)p(H_s)}{p(D|H_t)p(H_t)} \qquad (7)$$

where $p(H_s|D)/p(H_t|D)$ is a ratio describing the posterior odds of H_s to H_t as a result of a single Datum D; the $p(D|H_s)/p(D|H_t)$ is a ratio describing the likelihood of D dependent upon the odds of H_s to H_t; and the $p(H_s)/p(H_t)$ is a ratio describing the prior odds of H_s to H_t. Equation 5 may be written more simply as:

$$\Omega_1 = L\,\Omega_0 \qquad (8)$$

where Ω_1 are the posterior odds in favor of H_s; L is the likelihood ratio; and Ω_0 are the prior odds in favor of H_s. From equation 6 it can be seen that the posterior odds are equal to the product of the likelihood ratio and the prior odds in favor of H_s. Since the prior odds of the present experiment were equal to 1:1 at the beginning of each trial, the posterior odds were equal to the likelihood ratio.

Likelihood Ratio

The likelihood ratio specifies the likelihood of the particular datum occurring and is defined as "the probability of the datum on one hypothesis divided by its probability on the other hypothesis" (Edwards, Lindman, & Phillips, 1965, p. 302). The likelihood ratio is therefore an odds ratio and was calculated in favor of H_s for the present experiment.

Calculation of Likelihood Ratios

Likelihood ratios for both the subjective estimates (SLLR) and the Bayesian estimates (BLLR) were calculated. The SLLRs were calculated by converting the posterior probability estimates of the subjects into posterior odds in favor of H_s. These odds were then transformed into Log posterior odds. Conversion of subjective estimates into Log odds necessitated that the most extreme estimates of certainty acceptable in this experiment be 99 and 1.

The BLLRs for the present experiment were calculated as suggested by Edwards, et. al. (1965, p. 300). Briefly, if the probability of drawing a square from the box is p and the probability of drawing a triangle is q, then the probability of drawing any particular sample is $p^r q^{n-r}$, where r is equal to the actual number of squares drawn and $n-r$ is equal to the actual number of triangles drawn.

The BLLRs for the present study are the odds-likelihood ratio form of Bayes' theorem. They are the ratios depicting the probability of the datum given the box being used containing 600 squares divided by the probability of the same datum given that the correct box contains 600 triangles. These BLLRs were calculated as follows:

$$(p/q)^{r-(n-r)} \qquad (9)$$

These BLLRs were transformed into Log posterior odds. They were therefore equal to .17609 times the difference between the number of squares and triangles sampled on each trial. A BLLR was indeterminate on any trial with equal cue sampling. The BLLRs and corresponding probabilities for the present experiment are presented in Table I.

On Trial 9 of this study the number of squares and triangles sampled was equal. To maintain between subject variance of *S*s on this trial, the accuracy ratio used to depict the *S*s performance was the mean accuracy ratio for each *S* on the other 9 trials. This procedure was preferred to computing a grand mean accuracy ratio so that the data would continue to represent individual performance. The degrees of freedom for the appropriate error term were reduced by 30.

TABLE 1

Bayesian Likelihood Ratios and Corresponding Probabilities Per Trial

Trial	BLLR	$p(H_s)$
1	2.465	.996
2	1.760	.982
3	1.408	.962
4	.352	.692
5	.704	.834
6	1.056	.919
7	2.113	.992
8	1.056	.919
9	*	.600
10	.352	.692

*A BLLR for trial 9 was indeterminate.

Relation of SLLR to BLLR

The dependent variable relating the SLLRs to the BLLRs in the present experiment was the accuracy ratio. This ratio has been defined by Peterson and Miller (1965) as:

$$\frac{SLLR}{BLLR} \qquad (10)$$

If the *Ss'* estimates were revised with the same magnitude as the Bayesian estimates, the accuracy ratio was equal to 1.00. An accuracy ratio of less than 1.00 was accepted as evidence of conservatism.

The accuracy ratios and standard deviations per trial are presented in Table 2. These results support the first hypothesis of this experiment. Figure 1 offers further evidence that conservatism was present in this situation. From this figure it can be seen that *Ss* most nearly approximated the Bayesian estimates on trials 4, 9, and 10. Examination of Table 3 shows that on trials 4 and 10 the frequency with which the cues were sampled from the box was 19 squares and 17 triangles while on trial 9 this ratio was 50:50. Since the Bayesian estimates were calculated in favor of H_S with respect to Datum *D*, these estimates would be lower on those trials which most nearly approximated equal sampling of cues. The fact that *Ss'* estimates on these trials were less conservative than their estimates on the remaining seven trials

Table 2

Mean Accuracy Ratio and Standard Deviation

Per Trial n = 43

Trial	$\frac{SLLR}{BLLR}$	SD
1	.242	.243
2	.123	.112
3	.107	.138
4	.534	2.557
5	.392	1.670
6	.259	1.241
7	.295	1.382
8	.212	1.114
9	.682	1.080
10	.696	3.489

* An accuracy ratio deviation from 1.00 indicates conservatism

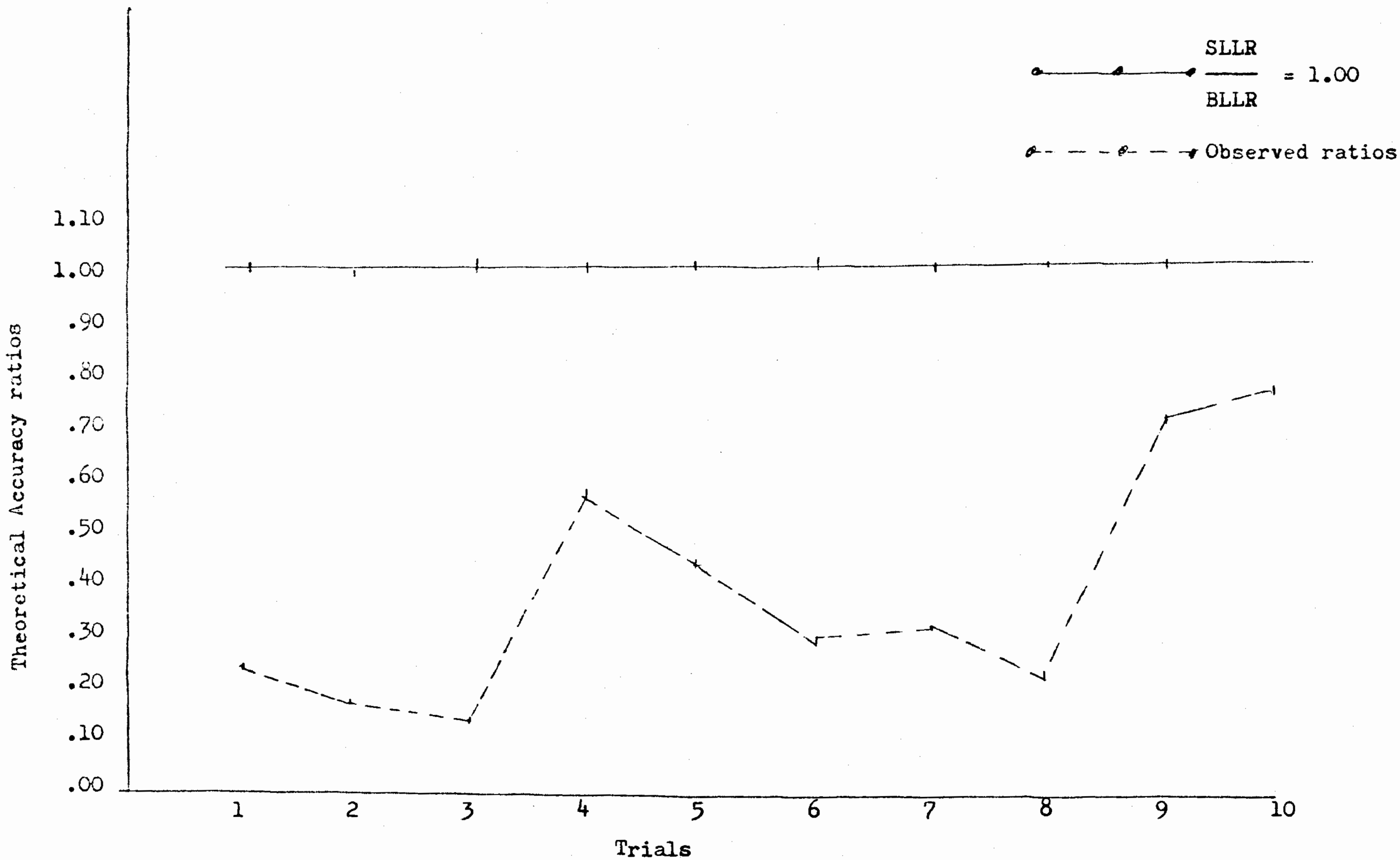


Figure 1 A comparison between the mean observed accuracy ratios per trial and a theoritical accuracy of ratio of 1.00(n=43). Deviation from 1.00 indicates conservatism.

Table 3

Frequency of Stimulus Cues and Correct Hypothesis Per Trial

Trial	Squares	Triangles	Correct Hypo.
1	25	11	H_s
2	23	13	H_t
3	22	14	H_t
4	19	17	H_s
5	20	16	H_s
6	21	15	H_t
7	24	12	H_s
8	21	15	H_t
9	18	18	H_s
10	19	17	H_t

suggests that they were operating or processing the information available. However, it is also possible that the Ss were continuing to function as consistently imperfect information processors. The higher accuracy ratios on trials 4, 9, and 10 may be only a matter of circumstance with respect to the lower BLLRs. No data is available to show which of the two above possibilities were operative on these trials.

To perform the analysis of variance to examine the differences in conservatism due to the particular sex of the subjects, the 28 female Ss were randomly divided into two groups of n=13 and n=15. Due to the specific nature of the experimental situation, random division of the female Ss was preferred to introducing 13 more male Ss to the task. Since a second group of males would have to be exposed to the same cue situation as the original group, this section would fail to meet the standards of randomization required for this experiment. A non-significant t-value of t=.019(df=26) between the differences of the two female groups allowed analysis of variance to be performed on the subsample of 15 females.

A summary of the analysis of variance appears in Table 4. The difference in the amount of conservatism due to the Ss sex was not significant. This analysis suggests that the amount of conservatism displayed by the female Ss with respect to H_s was of the same general magnitude as that displayed by the male Ss. Examination of the accuracy ratios for the female Ss and the male Ss on each trial support this analysis. These accuracy ratios are presented in Tables 5 and 6 respectively.

The trials effect of Table 4 reached significance beyond the $p < .001$ level. These results offer evidence that as the trials progressed both groups of Ss acquired some "skill" in their use of the cue information. However, it must be kept in mind that the possibility advanced earlier, that the Ss' estimates remained consistently imperfect but were less conservative towards the end of the experiment because the Bayesian estimates were lower on these later trials, is still tenable.

From the non-significant AB interaction of Table 4, it may be concluded that the female Ss adjusted to the situation with relatively the same magnitude as the Male Ss. Examinations of Tables 5 and 6 support this conclusion.

Table 4

Analysis of Variance Summary Table

n = 30

Source	ss	df	ms	F
Between Subjects	10.816	29		
A (sex)	.523	1	.534	1.425
Subjects within Groups	10.293	28	.367	
Within Subjects	67.580	270		
B (trials)	9.239	9	1.026	4.055**
AB (sex over trials)	2.091	9	.232	.917
B x Ss within Groups	56.250	222	.253	

** $p < .001$

Table 5

Observed Accuracy Ratios for n=15 Female Subjects

TRIALS

Subject	1	2	3	4	5	6	7	8	9	10
1	.341	.177	.058	.444	.197	.0001	.077	.869	.733	4.440
2	.019	.052	.058	.444	.296	.107	.011	.107	.181	.542
3	.003	.560	.039	.749	.012	.869	.077	.351	.378	.749
4	.341	.177	.117	.444	.592	.351	.770	.351	.398	.444
5	.341	.156	.039	.444	1.973	.869	.770	.351	.598	.444
6	.341	.117	.389	4.440	1.970	.860	.770	.0001	.995	.0001
7	.001	.001	.156	.444	.197	.132	.077	.132	.176	.444
8	.341	.177	.039	.012	.0001	.028	.770	.021	.260	.963
9	.341	.177	.156	.562	.197	.132	.069	.123	.310	1.030
10	.341	.040	.090	.444	.296	.263	.023	.869	.312	.444
11	.003	.177	.090	.444	.241	.204	.069	.351	.224	.444
12	.341	.040	.156	.249	.021	.204	.030	.132	.132	.018
13	.341	.0001	.0001	.0001	.0001	.869	.770	.869	.316	.0001
14	.013	.177	.090	.015	.197	.132	.011	.132	.134	.444
15	.341	.156	.039	.562	.790	.869	.770	.869	.981	4.440

Table 6

Observed Accuracy Ratios for n=15 Male Subjects

TRIALS

Subject	1	2	3	4	5	6	7	8	9	10
1	.341	.017	.058	.667	.296	.263	.043	.520	.194	.015
2	.341	.177	.039	.0001	.197	.087	.770	.052	.184	.0001
3	.341	.177	.741	.444	.241	.137	.077	.132	.303	.444
4	.431	.177	.156	.444	.592	.496	.770	.351	.461	.882
5	.341	.177	.039	.749	.597	.132	.770	.132	.399	.667
6	.341	.017	.001	.444	.241	.162	.030	.162	.215	.542
7	.019	.003	.001	.562	.460	.869	.043	.132	.292	.542
8	.082	.156	.085	.444	.365	.162	.030	.162	.218	.480
9	.341	.177	.0001	.249	.016	.869	.770	.869	.415	.444
10	.003	.017	.039	.444	.197	.087	.077	.087	.155	.444
11	.341	.052	.117	.440	.197	.132	.023	.132	.209	.444
12	.341	.069	.039	.749	.241	.204	.023	.132	.260	.542
13	.341	.177	.076	.562	.592	.132	.770	.869	.438	.426
14	.341	.177	.090	.444	.296	.162	.146	.162	.262	.542
15	.341	.177	.389	.444	.790	.351	.770	.204	.445	.542

Table 7

Observed Accuracy Ratios for n=13 Female Subjects

TRIALS

Subject	1	2	3	4	5	6	7	8	9	10
1	.341	.052	.117	1.332	.197	.869	.023	.869	.610	1.778
2	.341	.0001	.090	.444	.296	.204	.011	.052	.209	.444
3	.013	.040	.117	.444	.197	.263	.009	.107	.181	.444
4	.341	.177	.072	.444	.296	.132	.017	.132	.180	.015
5	.013	.052	.001	.444	.592	.087	.030	.037	.189	.444
6	.341	.177	.389	.444	.592	.790	.030	.263	.410	.667
7	.001	.032	.047	.444	.213	.107	.017	.162	.173	.542
8	.341	.177	.117	.562	.296	.132	.011	.132	.245	.444
9	.013	.040	.090	.444	.460	.162	.023	.204	.250	.822
10	.341	.177	.058	.444	.592	.263	.770	.263	.471	1.332
11	.341	.177	.117	.749	.592	.204	.770	.087	.420	.749
12	.341	.177	.0001	.0001	.0001	.0001	.770	.0001	.192	.444
13	.030	.040	.005	.015	.241	.132	.043	.204	.134	.500

Discussion

This study was designed to investigate the concept of conservatism in a Bayesian probability situation. The presence of conservatism was reflected by the accuracy ratio, presently defined as a ratio between the Log subjective odds in favor of the occurrence of one specific event and the corresponding Bayesian Log odds in favor of the same event. Evidence is offered to support the hypothesis that conservatism was present in a situation requiring Ss to make probability estimates with respect to the occurrence or non-occurrence of two mutually exclusive events.

Secondly, this study investigated the possibility that females would exhibit more conservatism than males in this situation. This hypothesis was not supported.

Bayes' theorem has been used as an appropriate normative model for experiments investigating the adequacy of subjective estimates (estimates made by people). This type of situation typically requires Ss to make probability estimates, given certain relevant information, about the occurrence of two or more independent events. These estimates are then compared with estimates calculated by Bayes' theorem for a measure of how adequately the Ss are utilizing the available information.

A consistent result of these studies has been that subjects fail to approximate the normative model prescribing their behavior. This phenomena, conservatism, has been offered as evidence that " . . . men are suboptimal processors of probability information" (Hays, 1963, p. 263). It has been suggested that subjects are unable to process all the information relevant to the situation and that they are inconsistent in their use of the information they are able to use. This may be one limitation on human behavior.

The controversy with respect to the Bayesian movement does not center around the statistical use of Bayes' theorem. Since Bayes' theorem is a direct consequence of conditional probability theory, it is not controversial with respect to its use in calculating the appropriate conditional probabilities in the above described situation. Instead, the present controversy is primarily directed towards the use of Bayes' theorem as a normative model prescribing how people should behave. It can be concluded that men are "suboptimal" processors of information with respect to Bayes' theorem only if Bayes' theorem is the appropriate model for prescribing "optimal" behavior. That this is the case has not been adequately demonstrated.

One difference between subjective and Bayesian estimates is that estimates derived from Bayes' theorem are required to obey the additivity principle, or sum to unity. No such requirement is required of subjective estimates. To the contrary, evidence is available to suggest that subjective estimates do not obey this principle.

Holmberg (1964) has conducted an experiment to investigate the additivity of subjective estimates. He found that the Ss estimates significantly failed to sum to unity. Those estimates which obeyed the additivity principle were primarily reflected in terms of "round" numbers (50:50, 60:40, 70:30, etc.). From these results he concluded that when unity was attained it was not on the basis of accurately estimating the separate probabilities. It was primarily attributable to a bias of estimates towards 50 per cent, or possibly convenience.

The present experiment, therefore, required the Ss estimates on any given presentation of cues to sum to unity; which was 100 in this study. Although it is possible that this manipulation placed a "psychological" restraint on the subjects' estimates, the present investigator felt that this control was necessary in order to insure that subjective estimates obeyed the same principles as Bayesian estimates. Since no indication was given by the Ss that they were aware they were only making one estimate, it is possible that this control failed to hamper or restrain their performance.

A second difference between subjective and Bayesian estimates is the distinction between psychological and mathematical probabilities. The information utilized by Bayes' theorem and that utilized by subjects are obviously different in nature. Bayes' theorem functions completely within the boundaries established by the separate prior probabilities of the alternatives under consideration and the cue proportions and sampled frequencies available. Such probabilities are "immune" to all other extraneous experimental variable.

Subjective estimates, however, are possibly reflective of more than these factors. Other variables influencing the performance of the subjects are his preferences for shapes, sizes, and colors. For example, in the present study three subjects expressed on the questionnaire filled out following the testing session that they had personal preferences for squares. One female subject reported, "I liked squares better than triangles although I often felt I was wrong." Another subject reported, "I determined my estimates by the number of squares and triangles drawn, but I had a bias towards squares for some unexplainable reason." These statements are suggestive that the geometric cues in this study had "meaning" to the subjects independent of the cue sampling on each trial.

Other variables which may affect subjective estimates are the time of day that the experiment is conducted, the nature of the specific task being performed, and the individual's previous attitudes and values. Out of the 30 subjects used for the analysis of variance of this study, 5 female and 10 male subjects expressed that they found the task interesting; however, only one male subject found the task uninteresting, while 4 female subjects expressed this attitude. Nine female subjects and 12 male subjects stated that they found it challenging, while only 2 females and 1 male expressed that it was boring. These results indicate that definite attitudes towards the experimental task were operative in the situation. These attitudes were possibly reflected in the subjects estimates. If the Bayesians insist that probability estimates are reflective of the persons making them, then the above factors must be considered in this type of situation. Since probabilities calculated from Bayes' theorem are not reflective of such factors, it is questionable if Bayes' theorem is the appropriate model for prescribing human behavior.

Stilson (1966, p. 79) points out that a second question related to the present controversy is the measurement of subjective estimates. He suggests that "The determination of subjective probabilities is a problem in psychological measurement." He further suggests that an adequate measure of subjective probabilities will require more precise measurement of the variables affecting them. The appropriate method of measurement and the identification of the variables to be measured are problems which can only be answered by further research.

In summary, it has been assumed in the past that because Bayes' theorem is derived from conditional probability theory, it is the appropriate or adequate normative model for prescribing behavior in a conditional probability situation. Two suggestions are presented by the present author. First, it is accepted that Bayes' theorem is the appropriate mathematical model for calculating the conditional probabilities in Bayesian probability situations, but it cannot be assumed nor has it been demonstrated that it is the appropriate normative model prescribing how people should behave in these situations.

Secondly, it is suggested that psychologists should possibly concern themselves more with the identification and measurement of the variables which affect subjective estimates. One such variable, the sex of the subject performing the task, was investigated in the present study. Another important variable would be utility, or how important it was to the subject to be correct in his choice. It seems likely that both the personality and the preferences of the subject are operative in a non-payoff situation as well as a situation offering monetary rewards. With the proper identification of these variables, it may then be possible to derive a model which would combine them in a manner that would more adequately prescribe human behavior. The possibility of such a model primarily depends on " . . . our ability to observe the right things in the right ways under the right circumstances" (Hays, 1963, p. 46).

The present experiment failed to offer evidence that the results of the Kogan and Wallach research would be operative in this situation. A number of possible explanations may account for these results. In their study the subjects physically manipulated the apparatus or were allowed to interrupt the testing session when they had reached a decision. No such procedures were operative in the present study. The Kogan et al. results therefore may be primarily due to the "physical" manipulation of the testing procedure by the subjects. The absence of the opportunity for this type of involvement in the present study may have affected the approaches adopted by the subjects in such a manner as to eliminate the differences in conservatism that Kogan and his colleagues found between male and female subjects. Further research is needed to investigate the possibility that the degree to which the subject is allowed to participate in the experimental session markedly affects the amount of conservatism present.

Kogan and Wallach also found that females were more conservative in a payoff situation. These results were not evident in the present non-payoff experiment. It is suggested, therefore, that payoffs may not only be necessary to enhance the subject's potential in making probability estimates, but that the effects of payoffs are not the same for both sexes. Further research should concentrate not only on the types of reinforcement and the amounts necessary to maximize human judgment, but should investigate the relationship between the subject's sex and reinforcement.

In the Kogan and Wallach study it was reported that female subjects tended to adopt more extreme but less risky strategies than male subjects. The present study does offer some evidence in support of this conclusion. The term "strategy" is presently used the same as in game theory and is defined as " . . . the selection of a probability distribution over events and the subsequent use of this distribution at each trial in a series to determine the particular succession of choices to be made." (Siegel, Siegel & Andrews, 1964, p. 6).

The male subjects in the present study adopted more risky strategies in approaching a solution to the problem than the female subjects. One male reported, "If the figures were equal I estimated 50:50. Each time the cues were unequal I added 5 per cent for each figure with the greatest frequency." A second male subject reported that, "The cue with the lowest frequency in the box used on the previous trial would, by law of average, be less likely to be drawn. If this minority symbol approached 50 per cent on the present trial, I would choose that box." A third male reported, "I used a score method, dividing the boxes and figures as teams and when they matched they gained points."

The strategies reported by the female subjects were more extreme than those of the males. One female reported, "It was the box with the most squares until I was proven otherwise." A second subject reported, "Usually I estimated for the first 3 or 4 draws and then I decided on a guess." A third female reported, "When 2 cues of the same kind were drawn I thought it came from that box."

Of the 28 female subjects in the present experiment, four expressed that their estimates were based on guessing. None of the male subjects reported adopting this approach. Six of the female subjects and 9 of the male subjects stated that their estimates were determined by the frequency of the cues drawn. It is apparent, therefore, that females may adopt a more extreme strategy than males.

A number of problems were evident in the methodology of the present experiment. The first is concerned with biased cue sampling as a result of the nature of the figures themselves. Since a square occupies twice as much area as a right triangle, the length of the legs of the triangle being equal to the length of the sides of the square, the box with the proportion of 600 squares and 400 triangles in it actually contained an area ratio of 600 square inches to 200 square inches. Even though there were 3 squares in the box for every 2 triangles, the squares actually occupy three times as much area as the triangles. In the box with 600 triangles and 400 squares the area ratio was 300 square inches to 400 square inches in favor of the squares. When sampling these figures from a wooden box with a pair of eyebrow tweezers it seems plausible to assume that the area inside the box occupied by each figure will influence which figures are drawn out. It could therefore be expected that more squares than triangles would be sampled regardless of which box was being used. Table 3 presents a breakdown of the cues actually sampled on each trial in the present experiment. It is evident from this table that more squares were sampled on each of the 10 trials.

The present writer is therefore suggesting that the lack of differences found between the amount of conservatism exhibited between the male and female subjects of the present experiment might be partially attributable to the lack of ambiguity associated with the sampled cues. It would not be expected that the subjects would record high probability estimates for the box with the greater proportion of triangles in it if twice as many squares had actually been sampled. An experimental situation where more ambiguity was present with respect to the available alternatives might result in the differences found between males and females as in the Kogan and Wallach research. Careful design of the particular cues used would be an important factor in future research.

Squares and right-triangles were used in the present study to permit easy discrimination on the part of the subjects. A second methodological problem evident in this study was with the displaying of these cues. Since each displayed square occupied twice as much area on the screen as a displayed triangle, projected draws of 50:50 may not have been perceived as such by the subjects. Such a draw may appear to contain more squares than triangles. An example of this situation in the present study was on trial 9. Even though the cues were sampled with equal frequency, the projection area occupied by the squares was twice that of the triangles. The possibility that this type of draw would be misrepresented to the subjects is reflected in the results that only 17 of the 43 subjects in this experiment recorded estimates of 50:50 on this trial. Research is therefore needed to investigate the method of presentation of information to the subjects in this type of situation.

In summary, the present experiment offers evidence to support the conclusion that in a Bayesian probability situation, estimates made by the subjects would fail to be equivalent to estimates computed by Bayes' theorem. The present author, however, is hesitant to accept these results as evidence that human beings are "suboptimal" processors of information. Of the 430 estimates used to examine the present results, thirteen were above the corresponding Bayesian estimates. If Bayesian estimates depict "optimal" behavior, then these thirteen estimates represent "optimal-plus" behavior. In other words, these estimates represent behavior which is superior to optimal behavior. This contradiction is therefore suggestive of further research to investigate the adequacy of Bayes' theorem as a normative model in this type of situation.

Secondly, this study failed to offer evidence that females would be more conservative than males. The present data does depict that males were more conservative than females on 7 of the 10 trials; however, these differences failed to reach significance.

This study does offer evidence that males adopt more risky but less extreme strategies in problem solving than females. Further research is needed to investigate the specific types of strategies employed by males and females and the effect that these strategies have on subjective estimates. Siegel, Siegel, and Andrews (1964) report that men tend to adopt strategies which maximize utility, in decision making situation; however, they stress that " . . . now research is needed to examine the generality of the findings (of their research) for both sexes."

REFERENCES

REFERENCES

Anscombe, F. J. Some remarks on Bayesian statistics. In M. W. Shelly and G. L. Bryan (Ed.), _Human judgment and optimality_. New York: John Wiley and Sons, 1964.

Bernoulli, D. Exposition of a new theory on the measurement of risk, transl. by L. Sommer, _Econometrics_., 1954, 22, 23-36. In G. A. Miller, _Mathematics and psychology_. New York: John Wiley and Sons, Inc., 1964, 36-52.

Edwards, W., Lindman, H., and Phillips, L. D. Actual versus optimal revision of opinion in the light of information. In J. Olds and Marianne Olds (Ed.), _New Directions in psychology II_. New York: Holt, Rinehart and Winston, 1965, p. 303.

Edwards, W., Lindman, H., and Savage, L. J. Bayesian statistical inference for psychological research. _Psychological Review_, 1963, 70, 193-242.

Hayes, W. L. _Statistics for psychologists_. New York: Holt, Rinehart, and Winston, 1963.

Holmberg, B. _Application of the additivity theorem to subjective probability_. Unpublished Master's thesis, Central Washington State College, 1964.

Kogan, N., and Wallach, M. A. _Risk taking: A study in cognition and personality_. New York: Holt, Rinehart, and Winston, 1964.

Miller, G. A. Mathematics and psychology. New York: John Wiley and Sons, Inc., 1964, 32-35.

Peterson, C. R., and Miller, A. J. Sensitivity of subjective probability revision. Journal of Experimental Psychology, 1965, 70, 117-121.

Peterson, C. R., Schneider, R. J., and Miller, A. J. Sample size and the revision of subjective probabilities. Journal of Experimental Psychology, 1965, 59, 522-527.

Phillips, L. D., and Edwards, W. Conservatism in a simple probability inference task. Journal of Experimental Psychology., 1966, 72(3), 345-354.

Phillips, L. D., Hays, W. L., and Edwards, W. Conservatism in complex probabilistic inference. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7, 7-18.

Savage, L. J. The foundations of statistics. New York: John Wiley and Sons. 1954. (Annual Review of Psychology, 1961, 12, 473-498)

Schum, D. A. Inferences on the basis of conditionally non-independent data. Journal of Experimental Psychology, 1966, 72(3), 401-409.

Siegel, S., Siegel, Alberta, and Andrews, Julia. Choice, strategy and utility. New York: McGraw Hill Book Co., 1964.

Stilson, D. W. Probability and statistics in psychological research and theory. San Francisco: Holden-Day, Inc. 1966.

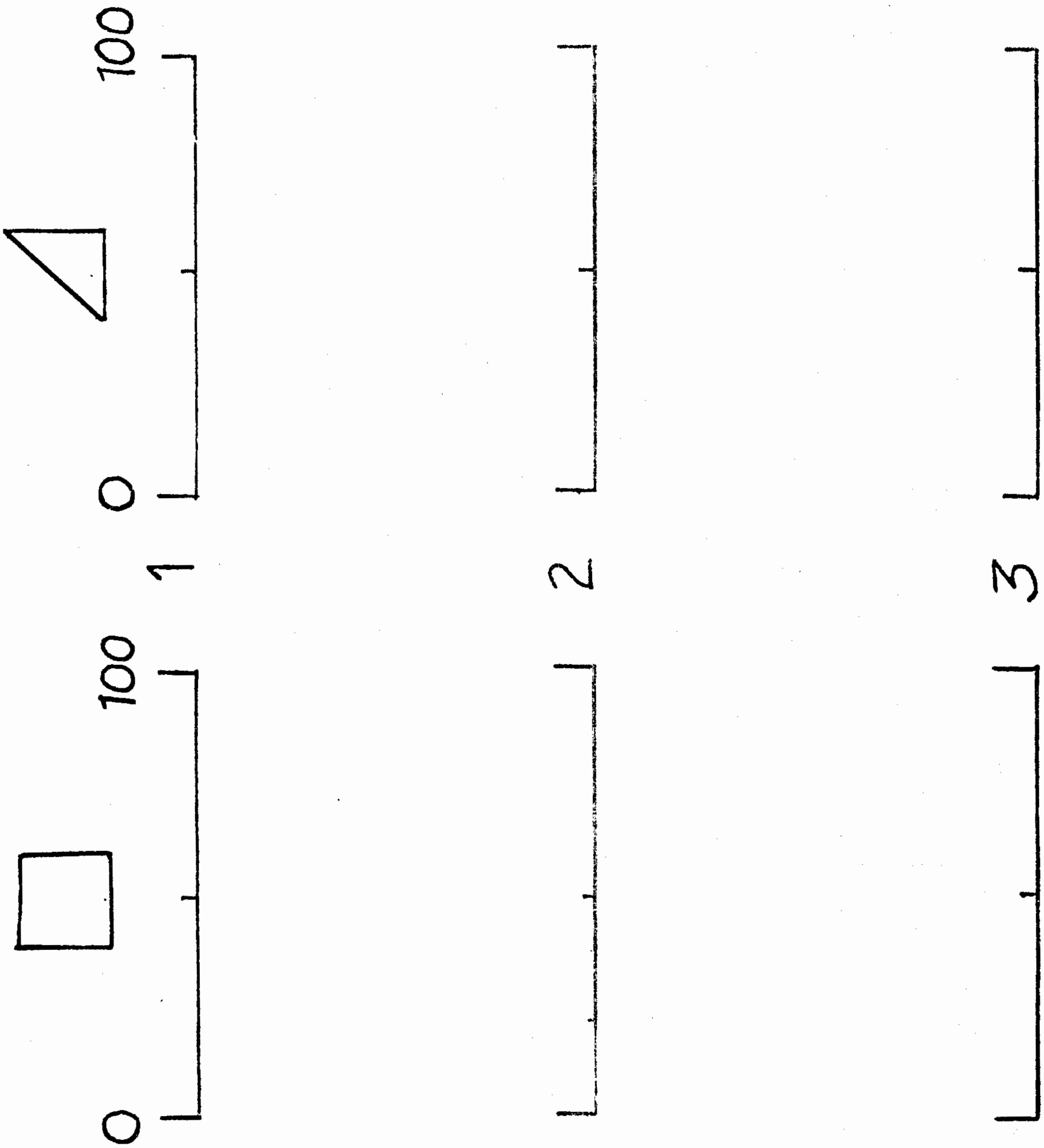
Master Vu-Graph overhead projector. Cat. No. 6600; 115 volts, 60 chcles, A.C. only: Charles Beseler Co., East Orange, New Jersey.

APPENDIX A

TRANSPARENCY USED ON PROJECTOR

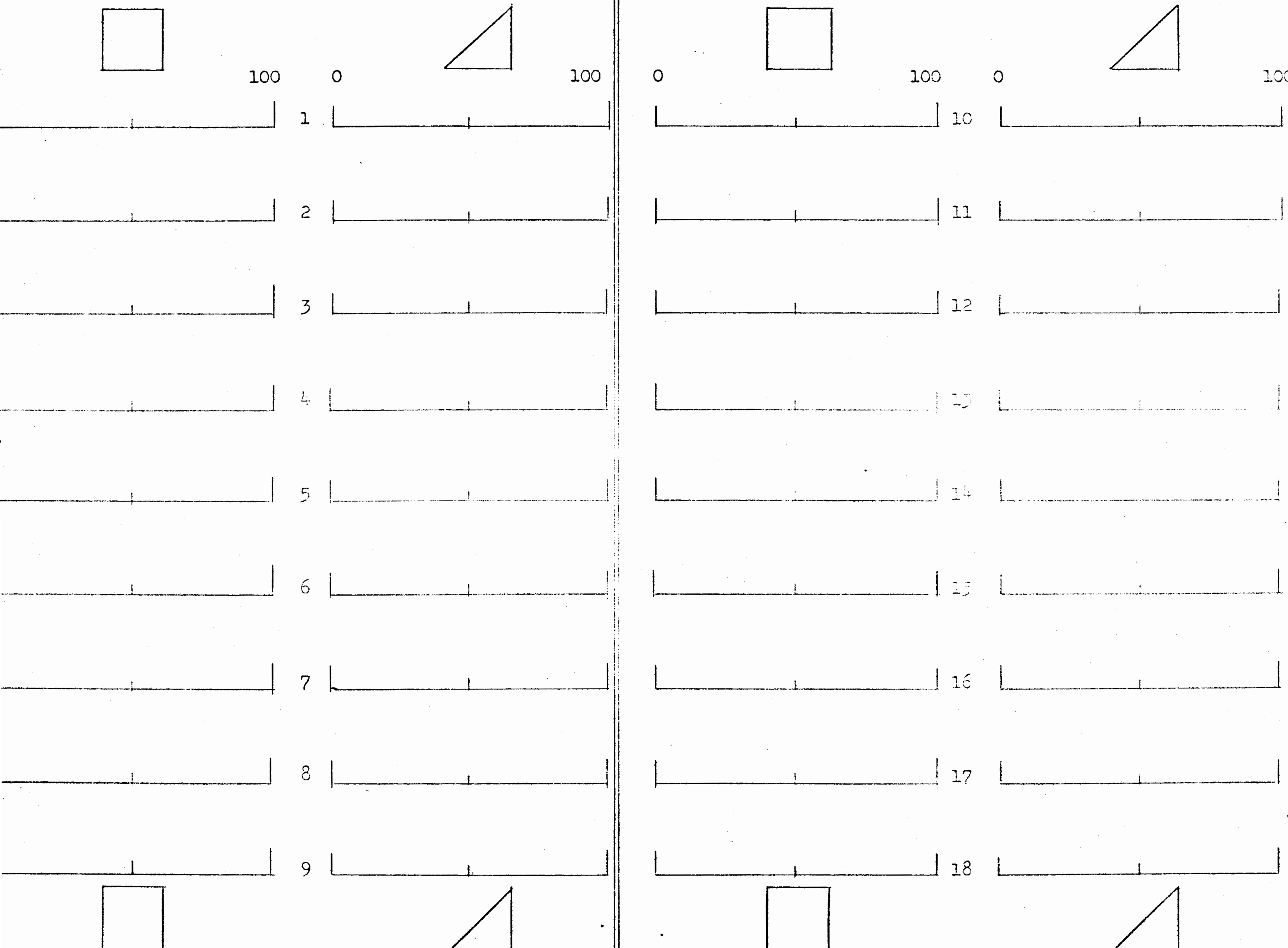
APPENDIX B

TRANSPARENCY OF EXAMPLE RESPONSE SHEET


0
100
1
2
3

APPENDIX C

SAMPLE SUBJECT RECORD SHEETS


0
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

APPENDIX D

INSTRUCTIONS

INSTRUCTIONS

You have volunteered to perform tonight a task in a psychological experiment. We appreciate your interest in our project and wish to thank all of you for donating your time this evening to our efforts.

The two red boxes you see setting on the table in the front of the room will be used. Each box contains a mixture of small cardboard geometric figures. The chart you see underneath the boxes summarizes the contents of each box. The box with the yellow square painted on it contains 600 square figures and 400 right triangle figures. The box with the yellow right triangle painted on it contains 600 right triangles and 400 squares.

For each trial we are going to draw 18 pairs of figures out of <u>one</u> of the boxes. As each pair is drawn out we will ask you to report which box you think we are using and to numerically record "how sure" you are that we are using that box. You will be told at the end of the "trial" which box we are using. There will be a total of 10 trials.

I will now summarize what has been said so far. Each time we draw a pair of figures out of the box you will be asked to do two things. First, you are to form opinions as to which box you think we are using. Your first opinion will refer to the box with the 600 squares in it. Your second opinion will refer to the box with the 600 triangles in it. Secondly, you are to assign a numerical estimate to each opinion. These estimates will indicate "how sure" you are as to which box we are using. Both your opinions and how sure you are of them are to be recorded on the record sheets provided.

The projection that is now on the screen in the front of the room shows you what the record sheet looks like. On the left side of the sheet there are three lines with a square above the top line. On this side of the sheet you will record how sure you are that we are using the box with 600 squares in it. On the right side of the sheet are three more lines with a right triangle above the top line. On these lines you will record how sure you are that we are using the box with 600 right triangles in it. The zero on the left side of the lines means that there is no chance that we could be using that box. The 100 on the right side of the lines means that you are absolutely sure that this is the box we are using. Anytime that you aren't able to record a 0 or a 100, in other words, anytime you are not absolutely sure as to which box we are using, you will record "how sure" you are somewhere on the lines between the 0 and the 100. Where on the line for each box you record "how sure" you are depends upon how strongly you feel about your opinions concerning each box. The only requirement is that each time we draw out a pair of figures, the two numbers you record, one on each line, must sum or add up to 100.

You will notice that between each pair of lines are the numbers 1, 2, and 3. This means that your first estimates will be made on the pair of lines numbered one, your second estimates (after we have drawn out the second pair of figures) will be made on the pair of lines numbered two, your third pair of estimates (after we have drawn out a third pair of figures) will be made on the lines numbered three.

I will now summarize what you are to do. To begin a "trial" we will draw, from the box to be used on the trial, one pair of figures. At this time you will be asked to form opinions and record "how sure" you are of them. You might not be too sure as to which box we are using. However, it is important that you make the best guess or estimates that you can. After you record your estimates we will draw out two more figures and show these to you. We will then ask you to record two more estimates. This procedure will be followed for all 18 pairs of figures.

To record how sure you are of your opinions place a mark on each of the lines and write in above each mark the number that the mark stands for. It is necessary that you write the numbers above the marks since the lines are <u>not</u>

marked off in equal units (10, 15, 20, 25, 30, 35, etc.). A brief demonstration of the task will now be given.

Suppose that the first two figures drawn out of the box are a square and a triangle. You might feel that with only this information given, one square and one triangle, that the chances are equal that we could be using either box. Therefore, you might want to mark a 50 on each line. Since there is already a mark in the middle of each line to indicate where 50 is, it would be necessary only to write in 50 above this mark on each of the lines. Notice that 50 + 50 = 100, which satisfies the requirement stated earlier. After you have written in your numbers we would proceed by drawing out two more figures. Now with the information of 4 figures to go by, you would record two more estimates on the lines with the 2 between them. This would be done the same as for the first pair of figures, by placing a mark on each of the lines and writing the number the mark stands for above each mark. The same procedure would be followed for the third pair of figures.

The record sheet for the first trial follows this page. You may now turn this page. Remember, even though you may not be sure as to which box we are using, please make the best estimates that you can. After the 18th pair of figures is drawn we will show you which box we are using.

ARE THERE ANY QUESTIONS?

APPENDIX E

QUESTIONNAIRE

QUESTIONNAIRE

It would be appreciated by the experimentors if you would take a few minutes to fill out this brief questionnaire pertaining to tonight's task. We are interested primarily in what you thought of the task and what approach you felt worked best for you.

1. Did you find the task interesting? (Briefly explain your answer.)

2. Did you feel that it was challenging? (Explain)

3. What was your particular approach to deciding which box we were using?

4. What phenomena do you think the experimenters are studying? (Your opinion, please)

5. Would you like to receive a classroom explanation as to what we are studying and why this particular task was employed?