Psychological Bulletin
1967, Vol. 68, No. 1, 29–46

MAN AS AN INTUITIVE STATISTICIAN [1]

CAMERON R. PETERSON AND LEE ROY BEACH
University of Michigan *University of Washington*

This review considers experimental research that has used probability theory and statistics as a framework within which to study human statistical inference. The experiments have investigated estimates of proportions, means, variances, and correlations, both of samples and of populations. In some experiments, parameters of populations were stationary; in others, the parameters changed over time. The experiments also investigated the determination of sample size and trial-by-trial predictions of events to be sampled from a population. In general, the results indicate that probability theory and statistics can be used as the basis for psychological models that integrate and account for human performance in a wide range of inferential tasks.

"Given . . . an intelligence which could comprehend all the forces of which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis . . . nothing would be uncertain and the future, as the past, would be present to its eyes [Laplace, 1814]." In lieu of such omniscience, man must cope with an environment about which he has only fallible information, "while God may not gamble, animals and humans do, . . . they cannot help but to gamble in an ecology that is of essence only partly accessible to their foresight [Brunswik, 1955]." And man gambles well. He survives and prospers while using the fallible information to infer the states of his uncertain environment and to predict future events.

Man's problems with his uncertain environment are similar to those faced by social enterprises such as science, industry, and agriculture. Satisfactory decisions require sound inferences about prevailing and future states of the environments in which these enterprises operate. Consequently, a great deal of effort has been invested in the development of coherent, formal procedures for dealing with fallible information in making inferences. These procedures, complex and sophisticated enough to have become a discipline, are called probability theory and statistics.

Because of the parallels between many of the inference tasks faced by man and by social enterprises, a number of investigators have used formal statistical theory as a point of reference for the study of human inference. For many uncertain situations, statistical theory provides models for making optimal inferences. The psychological research consists of examining the relation between inferences made by man and corresponding optimal inferences as would be made by "statistical man." [2]

The procedure is to use a normative model in order to identify variables relevant to the inference process. In this sense, probability theory and statistics fulfill a role similar to that of optics and acoustics in the study of vision and hearing. Just as optics and acoustics are theories of the environments in which eyes and ears operate, statistics is a theory of the uncertain environment in which man must make inferences. Sense organs do not merely mirror their physical environments, so their behavior cannot be described solely by a description of the environment. Instead, op-

[1] This research was undertaken in the Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan, under United States Public Health Service Fellowships MF-12,744 and MH-12,012 from the National Institute of Mental Health and was also supported by the Air Force Office of Scientific Research under Contract AF 49(638)-1731. Many of the ideas expressed here have their origins in discussions and arguments with Ward Edwards, to whom we are deeply indebted. We also would like to thank Kenneth R. Hammond who introduced us to the probabilistic functionalism of the late Egon Brunswik, an approach that has influenced much of the research in this review.

[2] Our use of "statistical man" as a model is analogous to the normative use of the "ideal observer" in signal detectability theory and "economic man" in economics. We mean the statistical logic and procedures appropriate to the task subjects must perform.

tical and acoustical theories have provided a basis for building descriptive theories that link vision and hearing to the physical dimensions of their environments. In the same manner, the theory of statistical inference can provide a basis for a descriptive theory of imperfect human inference.

A primary reason for selecting the strategy of evolving a theory of human inference from statistics is that the descriptive theory remains couched in the language of, and is structurally related to, the broad framework of the theory of statistical inference. This means that experimental findings from otherwise diverse areas may be logically integrated through reference to that theoretical framework.

The ultimate goal of this research is to develop a theory about human behavior in an uncertain environment, but the scope of this paper is necessarily more limited. First, it includes only behavior interpretable within the framework of statistical decision theory. Within this realm, the complete normative theory includes both statistical inference, as a model about how to gain knowledge of the environment, and decision theory, as a model for selecting courses of action in that environment. The psychological counterparts of these two components are intuitive statistics and psychological decision theory. This review explores only the predecisional process of intuitive statistics; reviews of the psychological decision literature are available elsewhere (Becker & McClintock, 1967; Edwards, 1954, 1961a).

This literature is organized in the familiar outline of an introductory statistics book. First, we examine intuitive descriptive statistics, the process of describing samples of data. We then consider research on intuitive inferential statistics, the process of using samples of data as a basis for making inferences about parent populations. Finally, we review studies of intuitive prediction, the process of using inferences about populations as the basis for predicting future samples to be drawn from those populations.

Intuitive Descriptive Statistics

By and large, psychologists have devoted less attention to studying intuitive descriptive statistics than they have to studying inferences. Perhaps this is because inference is inherently more interesting. Still, inferences about populations require prior summarization of sample data, and it can be argued that intuitive descriptive statistics underlie subsequent inferences.

Typically, experiments on descriptive statistics display a sample of data and ask the subjects for estimates of the proportion, mean, variance, correlation, or some other descriptive statistic. The correspondence between the estimates and the calculated statistics serves as the measure of accuracy.

Judgments of Proportion

Subjects have estimated proportions of both sequential and simultaneous displays of binary events (lights, horizontal and vertical lines, letters, numbers, etc.). The most striking aspect of the results is that the relation between mean estimates and sample proportions is described well by an identity function. The deviations from this function are small; the maximum deviation of the mean estimate from the sample proportion is usually only .03–.05, and the average deviations are very close to zero. Within the constraint of these small discrepancies, experiments have reported two different shapes for the slightly biased function, overestimation of low and underestimation of high proportions (Erlick, 1964; Stevens & Galanter, 1957), and underestimation of low and overestimation of high proportions (Nash, 1964; Pitz, 1965, 1966; Shuford, 1961; Simpson & Voss, 1961). The conflict in these results is particularly difficult to understand because similar procedures were used by Stevens and Galanter (1957) and Shuford (1961) on the one hand, and by Erlick (1964) and Pitz (1965, 1966) on the other.

We view the task of judging a proportion as one of statistical description. The subject never actually counts the elements in a display, however, so the task may also be viewed as one of inference. The displayed stimuli make up the population, and whatever information the subject can glean from observing the display is the sample. Results support this view. Accuracy of estimating proportions increases both with longer presentation times

(Erlick, 1961; Robinson, 1964; Shuford & Wiesen, 1959) and with the length of a sequence of elements (Erlick, 1964). Assuming that subjects gather larger samples during longer times or longer sequences, inferences based on those larger samples should have a smaller standard error of estimation and thus greater expected accuracy. Furthermore, with the exception of the .5 position (Nash, 1964), errors are smaller (Robinson, 1964) and fewer (Stevens & Galanter, 1957) and response variance is less (Shuford, 1961) for extreme proportions than for estimates in the middle of the scale. The variance of a sample and therefore the standard error of estimation is theoretically smaller for samples with more extreme proportions, so accuracy should be greater.

Judgments of Means and Variances

The central tendency and variability of samples of binary data are tied to a single statistic, the proportion. By contrast, separate statistics must be used to describe these properties in samples of interval or ratio scaled data. A number of statistics describe central tendency and naïve subjects reflect this variety by giving responses that sometimes correspond to the mean, sometimes to the median, and sometimes to the midrange (Spencer, 1963). When instructions specify the mean as the average to be estimated, the resulting estimates are nearly accurate (Beach & Swensson, 1966). Though there are no apparent biases, the variance among estimates increases with the variance of the sample, the sample size, and the speed of presentation (Beach & Swensson, 1966; Spencer, 1961). Since these variables would influence the standard error of estimate, which would in turn control the variability among estimates from different samples, these results provide further support for the hypothesis that subjects in a descriptive task are actually making inferences.

Just as judgments of means are influenced by the variance of the sample, judgments of variability are influenced by the mean, but in a different way. Hofstatter (1939) obtained judgments of the variability in the lengths of sticks tied in bundles. The judgments increased appropriately as the sample variance increased. However, as the means increased, the judgments decreased, much as though the subjects were estimating the coefficient of variation (standard deviation/mean) rather than the variance. Put another way, it is as though they were judging discrepancies from the mean in relation to the magnitude of the mean, an interpretation related to the Weber fraction, $\Delta I/I$, in psychophysics.

Lathrop (1967) has replicated this aspect of Hofstatter's results. It is as if subjects regard variance as relative to the general magnitude of the stimuli. This is intuitively compelling. Think of the top of a forest. The tree tops seem to form a fairly smooth surface, considering that the tree may be 60 or 70 feet tall. Now, look at your desk top. In all probability it is littered with many objects and if a cloth were thrown over it the surface would seem very bumpy and variable. The forest top is far more variable than the surface of your desk, but not relative to the sizes of the objects being considered. Perhaps this is a place where intuition and typical statistical usage are divergent; statisticians are seldom interested in variances relevant to means, but people may be.

Even when means are taken into consideration, there are still systematic discrepancies between intuitive judgments and objective values of sample variance. These discrepancies can be accounted for in part by the way in which subjects weight deviations of individual data from the sample mean. The mathematical variance is the average of the *squared* deviations. The power to which they are raised dictates the relative weighting of large and small deviations. An increase in the power increases the relative weight of large deviations; a decrease in the power increases the relative weight of small ones. In order to investigate the relative weights assigned by subjects, experimenters have calculated that power that permits the best prediction of intuitive estimates of variability. Hofstatter (1939) found large values, ranging up to 6, when experimental conditions emphasized large deviations. He found small values, ranging down to 0.5, with an emphasis on small deviations. Beach and Scopp (1967) used normally distributed samples and found that a small power, .39, best simulated the judgments of their subjects. In normally distributed samples, most of the data lie relatively near the mean; the resulting prevalence of small deviations may emphasize them. It seems likely that distributions that emphasize extreme scores, such as saddle-shaped distributions, would result in large powers. At any rate, this modification of the normative exponent, and the accompanying psychological interpretation, illustrates a way of modifying a normative statistical model in order to arrive at a model more descriptive of intuitive statistics.

Intuitive Inferential Statistics

Although many psychological studies of descriptive statistics may have investigated inference inadvertently, it is the explicit topic of the research discussed next. Experiments on intuitive inference explore how man uses samples of data to reach conclusions about characteristics of his environment. The data provide the basis for his judgments about the covert, underlying statistical structure of events. The theory of statistical inference specifies what kind of inferences should be made from the samples, and the experiments compare inferences made by men with optimal inferences.

Inferences about Population Parameters

Experiments on inference have used the optimal inferences specified by statistics as a basis for evaluating the optimality of human inferences. Note the difference in orientation between this approach and that of studies of intuitive descriptive statistics. The latter use accuracy as the criterion for good performance, that is, they ask "To what degree do estimates agree with the experimenter's measurements of the stimulus being estimated?" Optimality, on the other hand, is the degree to which intuitive inferences agree with optimal inferences given by the statistical model. The distinction is between using God or using statistical man as a criterion for performance. Even in an uncertain and probabilistic environment, an omniscient being would know the actual population parameters. But statistical man must be content to work with only the data in a sample and to make the best possible inference. When a sample is the only information provided to the subject, it is reasonable to use optimality rather than accuracy as the primary criterion for intuitive inference.

Inferences about proportions. In most investigations of inferences about proportions, subjects observe samples of binary data, and, after each datum in a sequence, they revise their probability estimates of each proportion being the population parameter. These revisions are compared with optimal revisions as calculated by using Bayes' theorem (see Edwards, Lindman, and Savage, 1963, for an extensive discussion of Bayesian statistical inference).

Imagine yourself in the following experiment. Two urns are filled with a large number of poker chips. The first urn contains 70% red chips and 30% blue. The second contains 70% blue chips and 30% red. The experimenter flips a fair coin to select one of the two urns, so the prior probability for each urn is .50. He then draws a succession of chips from the selected urn. Suppose that the sample contains eight red and four blue chips. What is your revised probability that the selected urn is the predominantly red one? If your answer is greater than .50, you favor the same urn that is favored by most subjects and by statistical man. If your probability for the red urn is about .75, your revision agrees with that given by most subjects. However, that revised estimate is very *conservative* when compared to the statistical man's revised probability of .97. That is, when statistical man and subjects start with the same prior probabilities for two population proportions, subjects revise their probabilities in the same direction but not as much as statistical man does (Edwards, Lindman, & Phillips, 1965).

Conservatism is suboptimal, but it is systematic, so research has looked for reasons for it. A number of studies have attempted to find out if conservatism is due merely to procedural variables. Earlier investigations had used probability estimates as the response, and it seemed possible that subjects avoided approaching the bounds of the scale. To check this, probability estimates were compared to unbounded odds estimates (Phillips & Edwards, 1966); odds estimates were only slightly less conservative than the probability estimates. Another hypothesis was that subjects had no incentive to perform well. However, while payoffs decreased response variance, they decreased conservatism only slightly (Phillips & Edwards, 1966). Other variables, such as sample size (Peterson, Schneider, & Miller, 1965) and sequential order of the data (Peterson & DuCharme, 1967; Phillips, Hays, & Edwards, 1966) affect conservatism, but instructions have virtually no influence. In short, while procedural variables influence the degree of conservatism, they do not eliminate it.

The persistence of conservatism in spite of variations in procedure suggests that it has roots in the fundamental aspects of subjects' understanding and use of information. One possibility is that peoples' intuitions about the relation between population and sample differ from the relations specified by statistical theory; or, in more formal terms, subjects have an inaccurate understanding of sampling distributions. In agreement with this hypothesis, when subjects make estimates about sampling distributions, the distributions are too flat (Peterson, DuCharme, & Edwards, in press). Moreover, probability revisions of individual subjects were predicted more accurately by substituting their flat distributions in the appropriate Bayesian equations than by using the theoretical sampling distributions.

In addition to a failure to understand the relation of samples to populations, there is also evidence that subjects have difficulty in aggregating evidence over trials (Edwards, 1966; Phillips, 1966). When they make datum-by-datum revisions throughout a sequence of data, the final subjective probability is far more conservative than when the experimenter optimally combines a series of single estimates made by subjects for each datum in the sequence. The former task requires retention of the previous inference and augmenting it in light of the succeeding datum, while the single estimates require only that subjects assess the meaning of each datum separately. At present, then, conservatism appears to be due in some small part to procedural variables, and in large part both to subjects' misunderstanding of sampling distributions and to their nonoptimal sequential revision of their subjective probabilities.

Inferences about means and variances. The experimental paradigm used to study inferences about means and variances is analogous to that used in studies of inferences about proportions. Data that vary along a dimension are sampled from one of two populations, and subjects decide from which of the two populations the data have been drawn. Some experiments using numerical samples had the subject infer which of two hypotheses about the parameter value was correct and state his confidence in the accuracy of that inference.[3] Irwin, Smith, and Mayfield (1956) used populations consisting of decks of cards upon which numbers were written. On the basis of each sample, subjects inferred whether the mean of the population was greater or less than zero. In a second experiment, the cards were sampled from two decks and the task was to infer which of the decks had the larger mean. In both experiments, confidence increased with the size of the sample, with either the difference between the population mean and zero or the difference between the two population means, and as the population variance decreased. Little and Lintz (1965) performed a similar experiment and found that on a trial-by-trial basis, confidence increased with sample size.

These experimenters used the t test as a method of summarizing their independent variables, but they used no normative model in the sense that the term has been used here. That is, they did not use a statistical model to prescribe the optimal confidence statement. The t test would not be the normative model because it yields the probability of the sample of data if the null hypothesis were true. This was not the question the subjects were asked (and it is claimed in some quarters

[3] We treat the confidence estimates and probability estimates as interchangeable measures of subjective probability when both have been measured on a 0–1.0 scale. For confidence estimates, the subject usually states which event he thinks is most likely to occur and then states his confidence that the choice is correct. For probability estimates, the subject merely states how certain he is that a given event will occur. These estimates are formally equivalent, but it is yet to be demonstrated that they are psychologically so.

that this is not a question that anybody should be asked; Edwards, Lindman, & Savage, 1963). Rather, they were asked for the probabilities of the alternative hypotheses on the basis of the data, a question answered by Bayesian statistics. Probabilities based on the normative model would be influenced by the three independent variables in the directions found in these experiments, but it is not clear whether subjects were conservative in arriving at their confidence statements.

A Bayesian model has, however, been applied to another experiment that used the paradigm just discussed (Beach & Scopp, 1967). The subjects inferred which of two decks of cards had the larger variance and stated their confidence. Confidence increased as the ratio of the judged sample variances increased, but not as much as prescribed by the model. These results are similar to the conservatism found with population proportions.

When subjects directly infer the central tendency of a population by specifying a value on a continuum of possible values, the inference must in some way represent all the values in the population. Various measures of central tendency represent the population values in different ways; the mode is equal to the most frequently occurring value, the median minimizes the sum of the absolute deviations between itself and the individual values, and the mean minimizes the sum of the squared deviations. For a skewed population distribution, the values of these measures are all different. When subjects base inferences on a sample that is displayed as a J shaped frequency distribution, intuitive inferences of the mode and median are accurate, but inferences of the mean are biased toward the median (Peterson & Miller, 1964). It would be possible to simulate this bias with the approach used to simulate judgments of sample variances (Beach & Scopp, 1967), that is, by modifying the power to which deviations are raised, away from 2 in the direction of 1. This means that subjects were unwilling to weight large deviations heavily. The deviant events were also rare events, so subjects may have regarded them as unrepresentative and thus not more important than the most frequently occurring events.

Much of the research using nonnumerical samples has been conducted within the framework of the theory of signal detectability (Swets, 1964). While we have no intention of reviewing this entire literature, the model of signal detection is a statistical model and several experiments are particularly relevant to intuitive statistics. As in the research discussed above, the formal problem for the subject is one of making an inference about the population from which the observation has been sampled. One population is that of normally distributed random noise. The second population is one of signal plus noise, with the same variance but a different mean than that of the noise population. From the subjects' point of view, the task is one of deciding whether or not a signal was present in the observation.

The majority of signal detection experiments have used auditory, visual, or other sensory stimuli, but the model has also been applied outside the realm of sensory psychophysics. For example, in perceptual defense experiments, the task is to decide whether the observation is a clean word or a taboo word (Dorfman, Grossberg, & Kroeker, 1965); in recognition memory experiments, the task is to decide whether the observation is an old word or a new word (Parks, 1966); in the perception of tilt, the task is to decide whether a line is tipped to the left or the right (Ulehla, 1966); and in one series of experiments, the task was to decide whether a dot was sampled from one spatial distribution or another (Lee, 1963; Lee & Janke, 1964, 1965). The model has even been extended to the judgment of the source of short phrases from a man's magazine or a woman's magazine (Ulehla, Canges, & Dowda, in press) and to reaction time experiments where the subjects' task is to react to a left or a right stimulus light (Edwards, 1965; Stone, 1960). These experiments show that it is possible to interpret a wide range of psychological phenomena within the framework of statistical decision theory. The results are in general accord with the predictions; many deviations from optimal performance are similar to those found in other areas of intuitive statistics. For example, it is possible to manipulate the subjective decision criterion by changing the probability of sampling from a signal distribution or by varying payoffs, but the amount of change in the subjective criterion is less than optimal (Green, 1960; Ulehla, 1966). The subjects also have difficulty in aggregating information across a sequence of trials (Swets & Green, 1961; Swets, Shipley, McKey, & Green, 1959), a result that bears a strong resemblance to the finding of conservatism in the probability-revision experiments discussed above.

Inferences about correlations. Thus far the tasks discussed have involved populations of events that vary along a single dimension. Nonlaboratory tasks, however, often involve a number of dimensions. Frequently these dimensions are not independent, and therefore it is important to examine intuitive inferences about correlations in multivariate populations.

Experiments using populations that contain two binary dimensions show that subjects do not attend to all cells of the 2×2 contingency table when inferring correlation. In some cases, judgments about the relatedness of the two dimensions depend solely upon one cell of the table, the cell in which the two favorable outcomes occur together (Jenkins & Ward, 1965; Smedslund, 1963; Ward & Jenkins, 1965); in other cases, judgments depend upon both cells of the positive diagonal (Inhelder & Piaget, 1958; Ward & Jenkins, 1965). The reason for the conflict is unclear, but even when subjects use the diagonal it appears that they do not fully appreciate the negative evidence represented in the remaining two cells of the 2×2 table.

It may be that failure to use all cells of the matrix is restricted to the special case of the 2×2 contingency table. Erlick (1966) presented samples from two 5-valued dimensions and had the subjects estimate the degree of positive or negative relatedness. The mean estimates were nearly linear with the objective correlations, except for a tendency to underestimate the magnitude of negative correlations. Beach and Scopp (1966) displayed samples from two 10-valued dimensions; the subjects inferred the sign of the population correlations and stated their confidence in the inferences. For both positive and negative correlations the proportion of optimal inferences and average confidence increased with the magnitude of the sample correlations, although confidence was conservative by comparison with the optimal values. In a more complex multiple regression experiment (Peterson, Hammond, & Summers, 1965b), subjects' estimates of cue weights ranked in the same order as optimal weights, further evidence that subjects do not restrict their attention to only a few cells of a data matrix. "Statistical man" appears to provide a better match to behavior when the stimulus situation becomes more complex, that is, when one moves beyond the special case of a 2×2 matrix.

Consistency among Inferences

We have discussed two criteria, accuracy and optimality, for evaluating performance in a statistical task. A third criterion is consistency, the degree to which relations among subjects' inferences correspond to the constraints required of statistical theory.

Optimality implies consistency, and thus optimality is the more stringent of the two criteria. Yet, consistency is an important criterion from a psychological point of view. If one's inferences are suboptimal but they fit together in a consistent manner, then the research problem is to learn why the consistent inferences are suboptimal and to modify the statistical model in order to develop a descriptive psychological theory. If, on the other hand, inferences are also inconsistent, then behavior is far less congruent with statistical theory and the outlook is dim for providing an orderly account of human inference within the framework of statistical theory.

The criterion of consistency requires that relations among sets of inferences be similar to those prescribed by statistical theory, even though the inferences themselves may be inaccurate. Experimenters have obtained inferences about two or more aspects of a population, often two probabilities, and then evaluated how well these inferences fit together when substituted into equations from the appropriate statistical model. Since accurate inferences about probabilities are consistent by definition, investigators usually take steps to insure inaccuracy.

One of the simplest relations to be examined is that the probabilities of an exhaustive set of mutually exclusive events should sum to 1.0. Because most experiments use response devices that automatically normalize, insuring that probability estimates sum to 1.0, few data are available. What data there are come from subsidiary parts of larger studies in which sums were not constrained. The results are conflicting. Phillips et al. (1966) measured the revision of probability estimates for four hypotheses in the light of sequentially presented data. One subject constrained his estimates to equal 1.0, but four other subjects revised their estimates for the most likely hypothesis upward without making corresponding decreases in the probabilities of the less likely hypotheses. In the latter case, of course, the sum of the estimates increased above 1.0 as evidence accumulated over trials. Alberoni (1962) had subjects estimate various binomial sampling distributions for samples of Size 4. The sums of the estimated probabilities for the different outcomes consistently totalled about .85, considerably less than the 1.0 required by probability theory.

When experimenters infer subjective probabilities from choices among bets, the subjective probabilities sometimes sum to approximately 1.0 (Lindman, 1965) and sometimes do not (Leibermann, 1958), and in one case they summed to 1.0 only with certain assumptions about utility for gambling (Tversky, 1964). The unresolved problem of whether or not subjective probabilities inferred from decisions sum to 1.0 has important implications for psychological decision theory, but is too complex to be discussed here. The interested reader is referred to Edwards (1962), Lindman (1965), and Tversky (1964).

Related to the question of whether exhaustive sets of probability estimates sum to 1.0 is the question of whether estimates for unions of events are equal to the sums of estimates for the component events. Beach and Peterson (1966) found that this correspondence held with high reliability when probability distributions were estimated for three different classes of events: a binomial sampling distribution, seven different events of a probability learning task, and the probabilities of each of seven well-known Republicans obtaining the Presidential nomination for the next election.

Experiments have also tested the consistency of probability estimates for the joint occurrence of two independent events. The estimates of the joint event should equal the product of the estimates of the component events. For adult subjects, estimates were roughly similar to the product when they were made for various combinations of skill and chance; but the relation did not hold for children (Cohen, Dearnaley, & Hansel, 1958). Shuford (1959) inferred subjective probabilities from the amount subjects were willing to pay in order to play various bets. Such inferred subjective probabilities for joint events were very nearly equal to the product of the inferred subjective probabilities of the component events.

When events are dependent, it is necessary to deal with conditional probabilities. Subjects perform as consistently as they do in the simpler case of independent events (Peterson, Ulehla, Miller, Bourne, & Stilson, 1965).

So far, we have been discussing structural consistency, the degree to which relations among probability estimates for a specific set of events correspond to the relations demanded by statistical theory. The introduction of change into a static system of probabilities necessitates the evaluation of a second kind of consistency, process consistency. This is the degree to which changes in the system corresponded to the changes demanded by probability theory.

Three experiments investigated consistency among changing probability estimates. In one, subjects observed a sequence of data sampled from one of two populations. After the presentation of each datum in the sequence, they revised probability estimates about which population was being sampled and about which datum would occur on the next trial. The relation between the two revisions was almost identical to that specified by probability theory (Peterson, Ulehla, Miller, Bourne, & Stilson, 1965).

In a more complex situation, subjects were faced with two different tasks. In the first, they revised probability estimates on the basis of a single datum. In the second, they revised probability estimates on the basis of combinations of those data. Consistency demands that revisions based upon the combinations be equal to products of revisions based upon the single datum. The revisions were highly correlated with this demand in all but extremely complicated situations (Beach, 1966).

The third experiment on process consistency was discussed earlier in conjunction with conservatism. Recall that conservative probability revisions were predicted more accurately by using each subject's own conservative estimates of the sampling distribution in the appropriate equations than by using theoretical sampling distributions. Although the subjects had conservative opinions about the sampling distributions, they apparently used revision rules that were nearly the same as those prescribed by probability theory (Peterson, DuCharme, & Edwards, in press).

Consistency need not be restricted to the relation among probability estimates. In the previously discussed investigation of inferences about population variances (Beach & Scopp, 1967), subjects also judged the relative magnitudes of the sample variances. The inferences and the judgments were both inaccurate. Inferences were not systematically related to the ratios of the objective sample variances, as demanded by the normative model, but both the accuracy of the inferences and the subjects' confidence in them increased monotonically with the ratios of the *judged* variances. That is, the subjects' inferences were constrained to be consistent with their inaccurate judgments of the sample variances. If a statistician observed sample variances equal to subjects' judgments, his inferences also would have been monotonically related to the ratios of those variances.

These last experiments, showing consistency between structure and process, illustrate that suboptimal performance may result from appropriate use of erroneous assumptions about the statistical structure of the task. In all of these studies of consistency, incorporation of subjective assumptions into the statistical models leads to improved predictions of performance, a modification that transforms the normative models into descriptive models.

Determining the Size of the Sample

In the experiments discussed so far, the subject has been a passive recipient of the samples of data upon which he based his inferences. In nonlaboratory situations, however, one seldom has such a passive role; an important ingredient of many inference tasks is active control of the amount of data in the sample. Larger samples tend to permit more accurate inferences, but they also cost more in terms of time, effort, and money. The essence of the sampling task, then, is to balance the value of making more accurate inferences against the cost of larger samples.

Formally, there are two ways that the subject can designate the size of the sample. The first is to specify size in advance, observe the data, and then make an inference. The second, called optional stopping, consists of sampling sequentially; after each datum the subject has the option of continuing to sample or of stopping and making his inference. Most research has focused on the latter case. Formal models for optional stopping (Edwards, 1965; Raiffa & Schlaifer, 1961; Schlaifer, 1959, 1961; Wald, 1947) can be summarized by a simple, intuitively appealing rule: Sample another datum if its cost is less than the expected increase in payoff from the information it will provide. In other words, purchase another datum only if it is worth more than it costs. In addition to costs and payoffs, probability variables play roles in determining when sampling should cease. Examples include the probability of each hypothesis prior to a sample, and the expected diagnostic value of the next datum. The models themselves are complex in that all of these variables seem to interact with each other (see, e.g., Edwards, 1965). Our goal here, however, is not to explore these formal models, but to consider the ways in which intuitive sampling processes relate to them.

Experiments have manipulated cost of data, payoff for accurate inferences, or both, and measured the consequent effect upon the number of data purchased. The task in these experiments was to decide from which population data were being sampled. The dependent variable was the number of data purchased prior to making that decision. The manipulations influenced the selected sample sizes, but the magnitude of influence was less than that prescribed by the models (Edwards & Kramer, 1963; Irwin & Smith, 1957; Lanzetta & Kanareff, 1962; Swets & Green, 1961). An exception to this generalization is an experiment in which subjects predetermined the size of the sample they wanted. The size of the payoff had no systematic influence on the number of data purchased (Green, Halbert, & Minas, 1964). Perhaps this is because the optimal stopping procedure more closely resembles nonlaboratory information purchasing tasks than does predetermining the sample size.

Manipulation of the prior probabilities of the alternative hypotheses also influences sample size. When the prior probabilities are reduced by increasing the number of alternative hypotheses, subjects select larger samples before making a decision (Becker, 1958; Messick, 1964). With just two hypotheses, the average amount of data purchased decreases as the prior probabilities become more extreme, that is, depart from .50-.50, but the rate of decrease is somewhat less than that called for by the optimal model (Green et al., 1964). Here, too, there is one exception in which the amount of data purchased was insensitive to the independent variable. Messick (1964) found no effect when he contrasted rectangular with peaked prior distributions.

The story is the same for the expected diagnostic value of data. When diagnostic value is increased by separating proportions for two alternative populations, subjects purchase less data (Becker, 1958; Edwards & Kramer, 1963), but the amount of change is not quite as much as that prescribed by the optimal strategy (Edwards & Kramer, 1963). Once again, there is an exception: Green et al. (1964) found no systematic effect. With normal rather than binomial populations, the diagnostic value is increased by separation of the population means or by decreasing the population variance; fewer data are purchased when they are more diagnostic (Irwin & Smith, 1956, 1957).

The results of these experiments on controlling the size of samples are similar to those obtained in experiments on other inference tasks. Variables that would influence the behavior of statistical man also influence subjects' behavior, but to a smaller degree. This effect may be summarized by the statement that subjects are only partially sensitive to the relevant variables. Recall that the same kind of effect characterizes conservatism (e.g., Peterson, DuCharme, & Edwards, in press; Peterson & Miller, 1965). These two sets of results may be consistent: If subjects are only partially sensitive to variables in probability revision tasks, the hypothesis of consistency requires that they also be only partially sensitive to the same variables in information purchasing tasks.

Intuitive Predictions of Samples

The first section of this paper considered the intuitive description of statistical characteristics of samples of data; the second section discussed the use of samples as a basis for intuitive inferences about populations. This section examines intuitive predictions about events that are to be sampled from populations.

Samples from Unidimensional Populations

The conceptually simplest prediction task requires trial-by-trial predictions of events that are randomly drawn from a unidimensional population with a stationary probability. When feedback is provided, this is the familiar paradigm of the probability learning experiment. Faced with this task, statistical man would always predict the most frequent event, but subjects do not; over trials the distribution of responses tends to match the distribution of stimuli.

Probability learning experiments constitute the majority of investigations of behavior in the face of uncertainty. Therefore, it is important to interpret this apparently nonrational behavior within the framework of intuitive statistics. One possibility, if we were merely attempting to describe the data, would be to follow the lead of the very successful stochastic learning models and postulate a dice thrower in the subject's head. That is, we would not only assume that the stimuli occur with a degree of randomness, but also that behavior is typified by randomness. This alternative is antagonistic to our point of view that man is an intuitive statistician who seeks to behave optimally. Behavior should be random only when attempting to befuddle a hostile environment and perhaps not even then; otherwise, it should be deterministic. Even in a probabilistic environment one response is usually more profitable than others. That is the response statistical man would choose and that is the response the subject should select.

The behavior to account for in a probability learning task is not matching; demonstration of matching requires that data be summarized across subjects and across blocks of trials. Closer analysis shows that neither the group nor the individual responds randomly with a probability equal to the stimulus probability. Group response proportions change drastically from trial to trial within a block (Overall & Brown, 1959; Toda, 1963), and different subjects yield grossly different response proportions over a block of trials (Peterson & Beach, 1967). Rather than matching, what must be explained is the fact that individual subjects systematically vary their responses from trial to trial instead of always predicting the most frequent event.

The reason that statistical man would always predict the most frequent event is that he understands the implications of drawing events at random from a population with a stationary probability. There is evidence that intuitive theories of randomness do not coincide with the mathematical theory (Brown, 1964; Tune, 1964). When subjects produce "random" sequences of events, they produce too few long runs and too many alternations. Subjective theories about random sequences apparently do not contain the property of independence through trials. The subjective probability of an event on a trial depends upon which events precede it.

Trial-by-trial variations in probability learning experiments also show sequential dependencies; there are too few long runs and too many alternations (Anderson, 1960; Beach & Swensson, 1967; Edwards, 1961b; Jarvik, 1951; Lindman & Edwards, 1961; Tune, 1964). The similarity between the sequential dependencies in these two situations suggests that the subjects' responses in a probability learning task may be determined by their assumptions about random sequences. That is, perhaps each subject has his own theory of randomness, a theory that differs from the mathematical theory in that it admits sequential dependencies. Statistical man using the subjects' theory of randomness in a probability learning experiment might well produce similar response sequences.

Samples from Multidimensional Populations

Though unidimensional sampling is theoretically the simplest case, nonlaboratory tasks are seldom so informationally impoverished. If, for example, you wish to predict the intelligence of a potential employee, you do not rely only on the proportion of previous interviewees who have been intelligent. You rely on the additional information provided by test scores, recommendations, appearance, mannerisms, and so on.

Simulation of such information-rich environments has used multidimensional populations. In relevant experiments, each trial is a random sample from a population with correlated dimensions. The use of cue information is investigated by permitting subjects to observe the outcome of all but one of the dimensions in the sample. These observations, the cues, are used to predict the value of the observation on the remaining dimension, the criterion. Then the sampled value of the criterion is revealed to provide feedback and to permit learning of the relations between the various cue-dimensions and the criterion dimensions.

In the unidimensional experiment, the optimal strategy is to learn which event has the highest probability of occurrence and to predict that event on all trials. The multidimensional case is more complex. Here it is necessary to learn the validities of the different cues, to rely on each cue dimension according to its validity, and to predict the criterion value that has the highest probability on the basis of the evidence provided by all of the cues.

Weighting of cues. Most empirical research on the problem of cue weighting has used multiple regression as the statistical model. Statistical man, faced with the task of using continuous cue and criterion dimensions, would calculate regression weights for each cue dimension and then use the weights to predict the criterion. The research question is, to what degree are responses the result of appropriate weighting of the cues? (See Hursch, Hammond, and Hursch, 1964, or Peterson, Hammond, and Summers, 1965b.)

Subjective cue-weighting is inferred from a variety of measures: by the correlation between each cue dimension and the responses, by the regression weights of the responses upon the cue dimensions, or by the subjects' direct estimates of the relative importance of each cue dimension in predicting the criterion. Early experimenters were interested in concept formation and the subjects' ability to differentiate relevant cues from complex stimuli (Smedslund, 1955, 1961b; Summers, 1962). They generally obtained poor performance, a result that was probably due to the difficulty in discriminating the cues and criterion rather than to an inability to use the cues correctly after they were discriminated.

More recent studies have used simpler stimuli. The magnitudes of the subjective cue-weights achieve the same rank order as the objective cue-weights and do so in relatively few trials, but the amount of separation among the subjective weights is sometimes less than the separation in statistical man's multiple regression equation. As in the experiments on conservatism and on information purchase, subjects are only partially sensitive to differences in relevant variables; they treat the cues as more equal in predictive value than they actually are (Azuma & Cronbach, 1966; Dudycha & Naylor, 1966; Hammond, Hursch, & Todd, 1964; Peterson, Hammond, & Summers, 1965b, Schenck & Naylor, 1966; Uhl, 1963).

Maximizing versus distributing responses. Some experiments report that response distributions approximately match the conditional probability distributions of the criteria (Binder & Feldman, 1960; Estes, 1959). Others find that the response distribution is more peaked than the conditional probability distribution, indicating a deviation from matching in the direction of optimality (Azuma, 1960; Beach, 1964; Goodnow, 1954; Peterson & Ulehla, 1964). Although these results are in conflict about the degree of optimality, they agree that subjects distribute responses rather than maximize. In this respect these results are similar to those obtained in the unidimensional experiments. The explanation in the unidimensional case, misunderstanding of random sequences, is less tenable in the multidimensional case. Until more is known about the microstructure of behavior in this situation, these results remain unexplained within the framework of intuitive statistics.

When the assumptions of regression models are met, the criterion with the highest conditional probability is the value that lies on the regression plane. As in a conditional probability learning experiment, this value of the criterion changes from trial to trial because the sampled cues change. The degree to which responses lie on any linear regression plane is measured by calculating the multiple correlation between cue dimensions and responses. Cue weights reflect the slope of the regression plane; the experiments discussed two paragraphs above show that response regression planes are close to the experimental regression planes, as they should be. The results are conflicting, however, with respect to the degree to which all responses lie on or near that regression plane. When only a single cue dimension is available, all responses do not lie on the plane (line)—they are distributed around it. The variance of the response distributions around the regression line increases as cues become less predictive of the criterion (Gray, Barnes, & Wilkinson, 1965; Schenck & Naylor, 1965).

The nonoptimal behavior found in single-cue experiments does not extend to multiple-cue studies. In the latter, responses are almost completely dependent upon the cues. The very high multiple correlations between responses and cues indicate that almost all responses fall directly on the response regression plane. This result holds not only where criteria are perfectly predictable from cues (Azuma & Cronbach, 1966; McHale & Stolurow, 1962, 1964; Uhl, 1963), but also when they are not (Grebstein, 1963; Peterson, Hammond, & Summers, 1965b; Todd, 1954).

In summary, subjects in conditional probability learning experiments scatter their responses. They do the same thing in single-cue regression experiments. In the seemingly more complex multiple-cue regression experiments, however, almost all responses fall on the response regression plane. It is not clear why the results conflict, but the evidence is abundant. Once again, greater task complexity appears to lead to more nearly optimal performance.

In addition to results on cue weighting and maximizing, there is other evidence that subjects are able to deal with functions relating criteria to cues. They can learn and use functions with both positive and negative slopes (Bjorkman, 1965); they can handle nonlinear as well as linear functions (Summers & Hammond, 1966); and perhaps most impressive of all, when confronted with a cue that they have never seen before, predictions fall on the regression line derived from previous observations (Bjorkman, 1965; Gray, Barnes, & Wilkinson, 1965).

Nonstationary Parameter Values

We have examined the ability of the intuitive statistician to perform in uncertain but stationary situations. Although the relation between population and sample was a noisy one, the population remained the same over time. The subjects were aware that changes in the sequential sample of data from one time to another were due to random fluctuations.

The nonlaboratory environment, however, is characterized by *nonstationary* situations as well, situations in which values of parameters change over time. This complicates matters considerably, because temporal fluctuations in the sequential sample can be due either to random perturbations or to real changes in the population. It is therefore necessary to penetrate through random variations, not only to detect population parameters, but also to detect changes in those parameters.

Statistical Models

The statistical procedures used in nonstationary situations are themselves models that assume stationarity. Adapting such a model to a changing situation "consists of finding ways of looking at a changing world so that it seems to be unchanging [Edwards et al., 1965, p. 310]."

Attempts to describe nonstationary situations with stationary statistical models fall into two general classes, which we will call deletion models and detection models. The essence of the deletion model is the analysis of data in small blocks, small enough so that the assumption of stationarity during the block is not too unreasonable, and the deletion of all other data. Another version is to take running averages, a process that slides the blocks through trials by deleting the oldest trial as it adds each new trial. A variation of the running average attributes more weight to recent data than to older data (Dodson, 1961). The deletion models suffer from an arbitrariness in the choice of the number of data included in a given block. The choice requires a compromise between the need for a sample large enough to yield a stable estimate of the population parameter and one small enough to make the assumption of stationarity during the block a reasonable one.

The detection model is less arbitrary. The idea is to compare incoming data with current estimates of the population parameter, until the new data become so divergent that the no-change hypothesis must be rejected in favor of the hypothesis that there has been a change. While these are hypotheses about change, they themselves do not change, thereby permitting the use of conventional statistical models that assume a stationary situation. Thus, detection models yield a hierarchy of inferences; inferences about the population are controlled by inferences about whether or not a change has occurred.

Since conventional statistical models generally assume stationary population parameters, their application to nonstationary situations depends to a considerable degree upon the ingenuity of the user. Few of the experiments have compared the performance of subjects with theoretically optimal performance in nonstationary statistical tasks. The typical procedure is to indicate the trials on which the parameter changed and to display the effect of the change on estimates made by subjects.

Experiments. Experimenters have used nonstationary situations to study the three classes of tasks discussed so far: description, inference, and prediction. In a description task, Robinson (1964) presented sequences of two rapidly flashing lights; the proportion of trials on which each light flashed changed with discrete steps at various points in the sequence. The task was to track the sample proportion by continuously adjusting a pointer on a proportion scale. The behavior could be described better by a detection than by a deletion model. Estimates changed abruptly following changes in the stimulus proportion; deletion models call for more gradual changes in response. Robinson pointed out, however, that the step changes in the proportions being estimated were especially compatible with detection models. It may be that the class of model that will describe behavior more accurately depends upon the characteristic of change in the experimental situation. Whatever the eventual status of these kinds of models, Robinson's results demonstrate that subjects can accurately estimate a time-varying binary probability.

A similar conclusion can be drawn when subjects infer nonstationary values of a population parameter. Rapoport (1964a, 1964b) selected values of the population proportion by a process that changed over time. The subjects used samples of data drawn from these populations to infer the value of the parameter. They made direct estimates and also estimated the interval within which they expected the parameter to fall. Intuitive inferences about the parameter changed in the direction of the shifts in the nonstationary process.

Responses are also sensitive to changes in parameter values when the task is to make sequential predictions about samples to be drawn from nonstationary populations. In probability learning experiments, when the stimulus probability changes over trials, the group response proportion tracks that change (e.g., Estes, 1959; Friedman, Burke, Cole, Keller, Millward, & Estes, 1964). The same thing happens when stimulus probabilities change, not simply as a function of trials, but as a function of the stimulus event of the preceding trial (e.g., Anderson, 1960). Finally, subjective cue-weights track changes in corresponding objective cue-weights when the task is to predict events to be sampled from multidimensional populations (Peterson, Hammond, & Summers, 1965a).

Additional support for the principle that subjects are sensitive to change comes from decision research, particularly research on multistage decisions. Rapoport (1965a, 1965b) developed tasks in which the state of the experiment changed over trials; the change depended on the state of the previous trial, on the decision made by the subject, and on some random process. Costs and payoffs were related both to decisions and to states, and the subject's goal was to make decisions that would maximize his net payoff. Rapoport found that intuitive decisions were remarkably near optimal decisions as prescribed by dynamic programming models (see Rapoport, 1965a, for references on dynamic programming). Although the task was primarily one of decision making and no model of statistical inference was tested, the nearly optimal decisions required sensitive inferences about a complex nonstationary process.

Application of conventional statistical models to changing situations is a complicated process, but the results of these experiments suggest that subjects are very sensitive to change; they are adaptable to nonstationary aspects of probabilistic situations.

Summary and Conclusions

The point of view underlying the research reviewed in this paper is that man must come to terms with his uncertain environment; he is not aware of all present conditions and he does not always know what will occur in the future. The formal theory of probability with its statistical applications describes the structure of that uncertain environment and the processes governing the occurrence of events within it. In addition, probability theory is normative; it provides optimal models for making inferences under conditions of uncertainty. This normative characteristic is the basis of the concept of "statistical man," a set of procedures for making optimal statistical inferences.

Experiments that have compared human inferences with those of statistical man show that the normative model provides a good first approximation for a psychological theory of inference. Inferences made by subjects are influenced by appropriate variables and in appropriate directions. But there are systematic discrepancies between normative and intuitive inferences. For example, the latter are usually too conservative; subjects apparently fail to extract all the information latent in samples of data. In addition, while intuitive inferences are sensitive to variables relevant to the normative model, the degree of sensitivity is often less than optimal.

A recurrent theme of the research reviewed is that some discrepancies are due to the fact that subjects in an inference task make assumptions different from those of statistical man. If statistical man were to incorporate subjects' assumptions, his inferences would be more descriptive of those made by subjects. Current research integrating subjective assumptions with the concept of statistical man may be a major step toward a psychological theory of intuitive statistical inference.

Such a theory would encompass only a restricted subset of human behavior, but there are some obvious possibilities for expansion. The subset increases considerably when the related normative models of probability theory and decision theory are joined as a basis for a broader psychological model including choice behavior as well as inference processes in an uncertain environment. Research by Piaget and his collaborators suggests another direction for expanding this normative approach to developing psychological models. For example, they have studied children's acquisition of principles such as the conservatism of substance and weight (e.g., see Smedslund, 1961a). Once the principle of conservation has been acquired, the child knows that the amount of substance and the weight of the object must remain unchanged if nothing is added or taken away, even though the form of the object may change. Principles such as the law of conservation are normative in that they lead to correct predictions of future events where alternative notions would lead to error. Thus, research on man as an intuitive statistician and as an intuitive decision maker could be extended to other disciplines offering normative models. The research could consider man as an intuitive scientist, logician, mathematician, and so on, and the resulting psychological theory would indeed apply to a large segment of human behavior.

REFERENCES

ALBERONI, F. Contribution to the study of subjective probability. I. *Journal of General Psychology,* 1962, **66**, 241–264.

ANDERSON, N. H. Effects of first-order conditional probability in a two-choice learning situation. *Journal of Experimental Psychology,* 1960, **59**, 73–93.

AZUMA, H. Comparison of a correlation with a probabilistic approach to concept learning. Unpublished doctoral dissertation, University of Illinois, 1960.

AZUMA, H., & CRONBACH, L. J. Cue-response correlations in the attainment of a scaler concept. *American Journal of Psychology,* 1966, **79**, 38–49.

BEACH, L. R. Cue probabilism and inference behavior. *Psychological Monographs,* 1964, **78**(5, Whole No. 582).

BEACH, L. R. Accuracy and consistency in the revision of subjective probabilities. *IEEE Transactions on Human Factors in Electronics,* 1966, HFE-7, 29–37.

BEACH, L. R., & PETERSON, C. R. Subjective probabilities for unions of events. *Psychonomic Science,* 1966, **5**, 307–308.

BEACH, L. R., & SCOPP, T. S. Inferences about correlations. *Psychonomic Science,* 1966, **6**, 253–254.

BEACH, L. R., & SCOPP, T. S. Intuitive statistical inferences about variances. Seattle: L. R. Beach, 1967. (Mimeo)

BEACH, L. R., & SWENSSON, R. G. Intuitive estimation of means. *Psychonomic Science,* 1966, **5**, 161–162.

BEACH, L. R., & SWENSSON, R. G. Instructions about randomness and run-dependency in two-choice learning. *Journal of Experimental Psychology,* 1967, in press.

BECKER, G. M. Sequential decision-making: Wald's model and estimates of parameters. *Journal of Experimental Psychology,* 1958, **55**, 628–636.

BECKER, G. M., & MCCLINTOCK, C. G. Value: Behavioral decision theory. *Annual Review of Psychology,* 1967, in press.

BINDER, A., & FELDMAN, S. E. The effects of experimentally controlled experience upon recognition responses. *Psychological Monographs,* 1960, **74**(9, Whole No. 496).

BJORKMAN, M. Learning of linear functions; comparison between a positive and a negative slope. *Reports from the Psychological Laboratories* (University of Stockholm), 1965, No. 183.

BROWN, D. L. Non-independence in subjectively random binary sequences. (Research Bulletin No. 27) Princeton, N. J.: Educational Testing Service, 1964.

BRUNSWIK, E. In defense of probabilistic functionalism: A reply. *Psychological Review,* 1955, **62**, 236–242.

COHEN, J., DEARNALEY, E. J., & HANSEL, C. E. M. Skill and chance: Variation in estimates of skill with an increasing element of chance. *British Journal of Psychology,* 1958, **49**, 319–323.

Dodson, J. D. Simulation system design for a TEAS simulation research facility. (Report R-194) Los Angeles: Planning Research Corporation, 1961.

Dorfman, D. D., Grossberg, J. M., & Kroeker, L. Recognition of taboo stimuli as a function of exposure time. *Journal of Personality and Social Psychology,* 1965, **2**, 552–562.

Dudycha, L. W., & Naylor, J. C. Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Performance,* 1966, **1**, 110–128.

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

Edwards, W. Behavioral decision theory. *Annual Review of Psychology,* 1961, **12**, 473–498. (a)

Edwards, W. Probability learning in 1000 trials. *Journal of Experimental Psychology,* 1961, **62**, 381–390. (b)

Edwards, W. Subjective probabilities inferred from decisions. *Psychological Review,* 1962, **69**, 109–135.

Edwards, W. Optimal strategies for seeking information: Models for statistics, choice reaction time, and human information processing. *Journal of Mathematical Psychology,* 1965, **2**, 312–329.

Edwards, W. Non-conservative probabilistic information processing systems. ESD-TR-66-404, Report 05893-22-F, August 1966, University of Michigan, Institute of Science and Technology, Electronic Systems Division.

Edwards, W., & Kramer, E. F. Boodle from Bayes' at a penny a peek. Ann Arbor: W. Edwards, 1963. (Mimeo)

Edwards, W., Lindman, H., & Phillips, L. D. Emerging technologies for making decisions. In, *New directions in psychology. II.* New York: Holt, Rinehart & Winston, 1965. Pp. 265–325.

Edwards, W., Lindman, H., & Savage, L. J. Bayesian statistical inference for psychological research. *Psychological Review,* 1963, **70**, 193–242.

Erlick, D. E. Judgments of the relative frequency of a sequential series of two events. *Journal of Experimental Psychology,* 1961, **62**, 105–112.

Erlick, D. E. Absolute judgments of discrete quantities randomly distributed over time. *Journal of Experimental Psychology,* 1964, **67**, 475–482.

Erlick, D. E. Human estimates of statistical relatedness. *Psychonomic Science,* 1966, **5**, 365–366.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 380–491.

Friedman, M. P., Burke, C. J., Cole, M., Keller, L., Millward, R. B., & Estes, W. K. Two choice behavior under extended training with shifting probabilities of reinforcement. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Palo Alto: Stanford University Press, 1964.

Goodnow, R. Utilization of partially valid cues in perceptual identification. Unpublished doctoral dissertation, Harvard University, 1954.

Gray, C. W., Barnes, C. B., & Wilkinson, E. F. The process of prediction as a function of the correlation between two scaled variables. *Psychonomic Science,* 1965, **3**, 231–232.

Grebstein, L. Relative accuracy of actuarial predictions, experienced clinicians, and graduate students in a clinical judgment task. *Journal of Consulting Psychology,* 1963, **37**, 127–132.

Green, D. M. Psychoacoustics and detection theory. *Journal of the Acoustical Society of America,* 1960, **32**, 1189–1203.

Green, P. E., Halbert, M. H., & Minas, J. S. An experiment in information buying. *Journal of Advertising Research,* 1964, **4**, 17–23.

Hammond, K. R., Hursch, C. J., & Todd, F. J. Analyzing of the components of clinical inference. *Psychological Review,* 1964, **71**, 438–456.

Hofstatter, P. R. Über die Schätzung von gruppeneigenschaften. *Zeitschrift für Psychologie,* 1939, **145**, 1–44.

Hursch, C. J., Hammond, K. R., & Hursch, J. L. Some methodological considerations in multiple-cue probability studies. *Psychological Review,* 1964, **71**, 41–60.

Inhelder, B., & Piaget, J. *The growth of logical thinking from childhood to adolescence.* New York: Basic Books, 1958.

Irwin, F. W., & Smith, W. A. S. Further tests of theories of decision in an "expanded judgment" situation. *Journal of Experimental Psychology,* 1956, **52**, 345–348.

Irwin, F. W., & Smith, W. A. S. Value, cost and information as determiners of a decision. *Journal of Experimental Psychology,* 1957, **54**, 229–231.

Irwin, F. W., Smith, W. A. S., & Mayfield, J. F. Tests of two theories of decision in an "expanded judgments" situation. *Journal of Experimental Psychology,* 1956, **51**, 261–268.

Jarvik, M. E. Probability learning in a negative recency effect in a serial anticipation of alternative symbols. *Journal of Experimental Psychology,* 1951, **41**, 291–297.

Jenkins, H. M., & Ward, W. C. The judgment of contingency between responses and outcomes. *Psychological Monographs,* 1965, **79**(1, Whole No. 594).

Lanzetta, J. T., & Kanareff, Z. T. Information cost, amount of payoff and level of aspiration as determinants of information seeking in decision making. *Behavioral Science,* 1962, **7**, 459–473.

Laplace, P. S. de *Essai Philosophique sur les Probabilitiés.* [*Concerning probability.*] (Orig. publ. 1814) In J. R. Neumann (Ed.), *The world of mathematics.* Vol. 2. New York: Simon and Schuster, 1956. P. 1325.

Lathrop, R. G. Perceived variability. *Journal of Experimental Psychology,* 1967, **23**, 498–502.

Lee, W. Choosing among confusably distributed stimuli with specified likelihood ratios. *Perceptual and Motor Skills,* 1963, **16**, 445–467.

Lee, W., Janke, M. Categorizing externally distributed stimulus samples for three continua. *Journal of Experimental Psychology,* 1964, **68**, 376–382.

LEE, W., & JANKE, M. Categorizing externally distributed stimulus samples for unequal molar probabilities. *Psychological Reports,* 1965, **17**, 79-90.

LEIBERMANN, B. The auction values of uncertain outcomes in win and loss-type situations. Unpublished doctoral dissertation, Boston University, 1958.

LINDMAN, H. R. The simultaneous measurement of utilities and subjective probabilities. Unpublished doctorial dissertation, University of Michigan, 1965.

LINDMAN, H., & EDWARDS, W. Unlearning the gambler's fallacy. *Journal of Experimental Psychology,* 1961, **62**, 630.

LITTLE, K. B., & LINTZ, L. M. Information and certainty. *Journal of Experimental Psychology,* 1965, **70**, 428-432.

McHALE, T. J., & STOLUROW, L. M. Concept formation with metrically multivalued cues and response categories. Technical Report No. 3, 1962, University of Illinois, Training Research Laboratory.

McHALE, T. J., & STOLUROW, L. M. More information—Cues or principle? Technical Report No. 5, 1964, University of Illinois, Training Research Laboratory.

MESSICK, D. N. Sequential information seeking: Effects of the number of terminal acts in prior information. Technical Documentary Report No. ESD-TDR-64-606, 1964, University of North Carolina, Electronic System Division.

NASH, H. The judgement of linear proportions. *American Journal of Psychology,* 1964, **77**, 480-484.

OVERALL, J. E., & BROWN, W. L. The comparison of the decision-behavior of rats and human subjects. *American Journal of Psychology,* 1959, **72**, 258-261.

PARKS, T. E. Signal-detectability theory of recognition-memory performance. *Psychological Review,* 1966, **73**, 44-58.

PETERSON, C. R., & BEACH, L. R. Probability learning? Ann Arbor: C. R. Peterson, 1967. (Mimeo)

PETERSON, C. R., & DUCHARME, W. M. A primacy effect in subjective probability revision. *Journal of Experimental Psychology,* 1967, **73**, 61-65.

PETERSON, C. R., DUCHARME, W. M., & EDWARDS, W. Estimated binomial sampling distributions in the revision of probability estimates. *Journal of Experimental Psychology,* in press.

PETERSON, C. R., HAMMOND, K. R., & SUMMERS, D. A. Multiple probability learning with shifting cue weights. *American Journal of Psychology,* 1965, **78**, 660-663. (a)

PETERSON, C. R., HAMMOND, K. R., & SUMMERS, D. A. Optimal responding in multiple-cue probability learning. *Journal of Experimental Psychology,* 1965, **70**, 270-276. (b)

PETERSON, C. R., & MILLER, A. Mode, median and mean as optimal strategies. *Journal of Experimental Psychology,* 1964, **68**, 363-367.

PETERSON, C. R., & MILLER, A. J. Sensitivity of subjective probability revision. *Journal of Experimental Psychology,* 1965, 70, 117-121.

PETERSON, C. R., SCHNEIDER, R. J., & MILLER, A. J. Sample size in the revision of subjective probabilities. *Journal of Experimental Psychology,* 1965, **69**, 522-527.

PETERSON, C. R., & ULEHLA, Z. J. Uncertainty, inference difficulty, and probability learning. *Journal of Experimental Psychology,* 1964, **67**, 523-530.

PETERSON, C. R., ULEHLA, Z. J., MILLER, A. J., BOURNE, L. E., JR., & STILSON, D. W. Internal consistency of subjective probabilities. *Journal of Experimental Psychology,* 1965, **70**, 526-533.

PHILLIPS, L. D. Some components of probabilistic inference. Unpublished doctoral dissertation, University of Michigan, 1966.

PHILLIPS, L. D., & EDWARDS, W. Conservatism in a simple probability inference task. *Journal of Experimental Psychology,* 1966, **72**, 346-354.

PHILLIPS, L. D., HAYS, W. L., & EDWARDS, W. Conservatism in complex probability inference. *IEEE Transactions on Human Factors in Electronics,* 1966, HFE-7, 7-18.

PITZ, G. F. Response variables in the estimation of relative frequency. *Perceptual and Motor Skills,* 1965, **21**, 867-873.

PITZ, G. F. The sequential judgment of proportion. *Psychonomic Science,* 1966, **4**, 397-398.

RAIFFA, H., & SCHLAIFFER, R. *Applied statistical decision theory.* Boston: Harvard Business School, Division of Research, 1961.

RAPOPORT, A. Sequential decision-making in a computer-controlled task. *Journal of Mathematical Psychology,* 1964 **1**, 351-374. (a)

RAPOPORT, A. A study of human control in a stochastic multistage decision task. Report No. 43, 1964, University of North Carolina, Psychometric Laboratory. (b)

RAPOPORT, A. Dynamic programming models for multistage decision making tasks. Reprint 150, 1965, University of Michigan, Mental Health Research Institute. (a)

RAPOPORT, A. A study of a multistage decision making task with an unknown duration. Reprint 152, 1965, University of Michigan, Mental Health Research Institute. (b)

ROBINSON, G. H. Continuous estimation of a time-varying probability. *Ergonomics,* 1964, **7**, 7-21.

SCHENCK, A., & NAYLOR, J. D. The effect of cue intercorrelation on performance in a multiple-cue choice situation. Paper presented at the meeting of Midwestern Psychological Association, Chicago, May 1965.

SCHLAIFFER, R. *Probability in statistics for business decisions.* New York: McGraw-Hill, 1959.

SCHLAIFFER, R. *Introduction to statistics for business decisions.* New York: McGraw-Hill, 1961.

SHUFORD, E. H. A comparison of subjective probabilities for elementary and compound events. Report No. 20, 1959, University of North Carolina, Psychometric Laboratory.

SHUFORD, E. H. Percentage estimation of proportion as a function of element type, exposure type, and task. *Journal of Experimental Psychology,* 1961, **61**, 430-436.

SHUFORD, E. H., & WIESEN, R. A. Bayes' estimation of proportions: The effect of stimulus distribution in exposure time. Report No. 23, 1959, University of North Carolina, Psychometric Laboratory.

SIMPSON, W., & VOSS, J. F. Psychophysical judgments of probabilistic stimulus sequences. *Journal of Experimental Psychology,* 1961, **62**, 416–422.

SMEDSLUND, J. *Multiple-probability learning: An inquiry into the origins of perception.* Oslo: Akademisk, 1955.

SMEDSLUND, J. The acquisition of conservation of substance and weight in children. *Scandanavian Journal of Psychology,* 1961, **2**, 11–20. (a)

SMEDSLUND, J. The utilization of probabilistic cues after 1100 and 4800 stimulus presentations. *Acta Psychologica,* 1961, **18**, 383–386. (b)

SMEDSLUND, J. The concept of correlation in adults. *Scandinavian Journal of Psychology,* 1963, **4**, 165–173.

SPENCER, J. Estimating averages. *Ergonomics,* 1961, **4**, 317–328.

SPENCER, J. A further study of estimating averages. *Ergonomics,* 1963, **6**, 255–265.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology,* 1957, **54**, 377–411.

STONE, M. Models for choice-reaction time. *Psychometrika,* 1960, **25**, 251–260.

SUMMERS, S. A. The learning of responses to multiple weighted cues. *Journal of Experimental Psychology,* 1962, **64**, 29–34.

SUMMERS, D. A., & HAMMOND, K. R. Inference behavior in multiple-cue tasks involving both linear and nonlinear relations. *Journal of Experimental Psychology,* 1966, **71**, 751–757.

SWETS, J. A. *Signal detection and recognition by human observers.* New York: Wiley, 1964.

SWETS, J. A., & GREEN, D. M. Sequential observations by human observers of signals and noise. In C. Cherry (Ed.), *Information theory.* London: Butterworth, 1961. Pp. 177–195.

SWETS, J. A., SHIPLEY, E. F., MCKEY, M. J., & GREEN, D. M. Multiple observations of signals and noise. *Journal of the Acoustical Society of America,* 1959, **31**, 514–521.

TODA, M. Micro-structure of guess processes: Part C. Report No. 7, June 1963, Pennsylvania State University, Institute of Research, Division of Mathematical Psychology.

TODD, F. J. A methodological study of clinical judgment. Unpublished doctoral dissertation, University of Colorado, 1954.

TUNE, G. S. Response preferences: A review of some relevant literature. *Psychological Bulletin,* 1964, **61**, 286–302.

TVERSKY, A. N. Additive choice structures. Unpublished doctoral dissertation, University of Michigan, 1964.

UHL, C. N. Learning interval concepts: I. Effects of differences in stimulus weights. *Journal of Experimental Psychology,* 1963, **66**, 264–273.

ULEHLA, Z. J. Optimality of perceptual decision criteria. *Journal of Experimental Psychology,* 1966, **71**, 564–569.

ULEHLA, Z. J., CANGES, L., & DOWDA, F. Signal detectability theory applied to conceptual discrimination. *Psychonomic Science,* in press.

WALD, A. *Sequential analysis.* New York: Wiley, 1947.

WARD, W. C., & JENKINS, H. M. The display of information and the judgment of contingency. *Canadian Journal of Psychology,* 1965, **19**, 231–241.

(Received September 27, 1966)