BELIEF STATES: A PRELIMINARY EMPIRICAL STUDY

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-238

March 1964

Thornton B. Roby

DECISION SCIENCES LABORATORY

DEPUTY FOR ENGINEERING AND TECHNOLOGY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

(Prepared under contract No. AF 19 (628)2450 by Tufts University, Medford, Mass.)
Project No. -- 4690
Task No. -- 469003

ESD-TDR-64-238

BELIEF STATES: A PRELIMINARY EMPIRICAL STUDY

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-238

March 1964

Thornton B. Roby

DECISION SCIENCES LABORATORY

DEPUTY FOR ENGINEERING AND TECHNOLOGY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

(Prepared under contract No. AF 19 (628)2450 by Tufts University, Medford, Mass.)
Project No. -- 4690
Task No. -- 469003

FOREWORD

This research was supported in part by a research grant, G-10947, from the National Science Foundation, and in part by Contract No. 494(15) with the Office of Naval Research, and in part by Contract AF19(628)2450 with the Decision Sciences Laboratory, Electronic Systems Division, Air Force Systems Command, United States Air Force, L. G. Hanscom Field, Bedford, Massachusetts.

BELIEF STATES: A PRELIMINARY EMPIRICAL STUDY

Abstract

The 'belief state,' as a technique for describing beliefs, attitudes and judgments, is proposed here as an important adjunct to psychological research, especially in the areas of decision making by individuals and by groups. The 'belief state' is defined as a device for representing in probabilistic form an exact quantitative description of the information or beliefs an individual has about possible alternative conditions of the external world.

The present study investigates the feasibility of measuring belief states in a simple laboratory situation, and of ascertaining by various statistical tests the degree to which these empirical measures conform to the normative standard of a perfect Bayesian calculation. It is shown that Ss in general depart from the standard in certain properties; that there are reliable individual differences among Ss' belief state measures as an apparent result of increasing experience with the task. Several suggestions are offered for future investigations of these measures under more closely controlled conditions.

PUBLICATION REVIEW AND APPROVAL

This Technical Documentary Report has been reviewed and is approved.

Herbert Rubenstein
HERBERT RUBENSTEIN
Chief, Decision Techniques Division
Decision Sciences Laboratory

Roy Morgan
ROY MORGAN
Colonel, USAF
Director, Decision Sciences Laboratory

KEYWORD LIST

PSYCHOLOGY
DECISION MAKING
ATTITUDES
PROBABILITY
PSYCHOMETRICS

TABLE OF CONTENTS

Tables

Belief States: A Preliminary Empirical Study

1. Introduction

The typical adult human cranium appears to contain an almost limitless variety of cognitive entities which in one way or another reflect the individual's experience of, and reaction to, the external world. These may take the cerebral form of axioms and hypotheses, or they may take the more visceral form of attitudes, sentiments, and prejudices. They may be clearly articulated and precise, or amorphous and inchoate. They may be based on specific sources of evidence or they may be the product of experiences or lessons which have since faded into the background.

The present empirical study is based on a thesis which has been more fully elaborated in an earlier theoretical paper (Roby, 1962). The suggestion is that it may be both feasible and fruitful to represent all such cognitive entities by a uniform mathematical device which permits an exact quantitative description of any given attitude or belief, and which lends itself to certain important conceptual operations.

The suggested mathematical construct will be referred to as a belief state. The belief state is a vectorial unit whose components describe a certain probability distribution, namely, the set of probabilities that each one of a specified list of alternative conditions is true. The latter are referred to as E-states (environmental states) and are herein assumed to be mutually exclusive and exhaustive.

To illustrate, let the set of alternatives be:

E_1) Shakespeare wrote Shakespeare's works

E_2) Bacon wrote Shakespeare's works

E_3) DeVere wrote Shakespeare's works

E_4) Some individual not named wrote Shakespeare's works

E_5) Some committee wrote Shakespeare's works

Then the typical B-states might be

B_o = (.2, .2, .2, .2, .2) ;

B_s = (1.0, .0, .0, .0, .0) ;

B_u = (.1, .2, .2, .5, .0) ; and so forth. The first, B_o, indicates complete uncertainty among the five alternatives; B_s indicates a total faith in the Historical Bard; B_u indicates a more diffuse belief that any one of several individuals might have penned the works.

This mode of representation is proposed instead of such familiar alternatives as "who do you believe wrote Shakespeare's works?" with a single answer, or "how sure are you that (x) wrote Shakespeare's works?". Of course, the intended form of representation also suggests that the beliefs, attitudes, etc., should be measured in this form if possible.

This position does not imply any conviction that the internal correlate of a B-state is adequately described by this mathematical symbolism. There may be nothing in the S's introspective experience that corresponds to the segmentation of the belief into probability components. Moreover, we are not advancing the extreme operationist claim that the belief is definitionally what is measured.

The assumed value of the proposed method of representing B-states is based entirely on the usefulness of such a representation in theory construction if the corresponding empirical quantities can be shown to follow certain functional rules.

Perhaps the most pressing need for a well-behaved mathematical construct for describing beliefs and attitudes arises in the area of social psychology and group processes. Several potential applications will illustrate this point: first, to provide a measure of agreement or consensus among individuals; second, to measure the influence of one person's announced opinion on the opinions of others; and, third, to compare the composite information or judgement of a group of persons with that of the individual group members. It can be shown that the representation of beliefs here suggested lends itself very well to each of these applications.

Before building any theoretical superstructure on this concept, however, it appears prudent to examine the salient properties of belief states generated and measured under controlled conditions. This is the purpose of the present methodological investigation. If belief states, as here defined, exhibit the desirable conceptual properties that are considered, their use in more ambitious theoretical systems will be partially justified. If they fail to meet these conditions, it may still be possible to use them, but with specific qualifications and restrictions.

2. Experimental Procedure

The data discussed here were obtained in a single experimental session with a group of 24 college students of either sex who were participating in a series of studies of various aspects of decision-making behavior. The task was to estimate the proportion of blue chips in an urn that contained a total of seven blue and white chips. The evidence on which Ss based their estimates was a succession of 20 honest draws with replacement; that is, at each of 20 trials a chip was drawn at random, displayed, and returned to the urn.

Ss indicated their estimates after each draw by distributing 12 hypothetical betting points over the four proportions that were announced as possible and equiprobable, vis 2/7, 3/7, 4/7, 5/7. The payoff for a given run, or sequence of 20 trials, depended on an Ss pattern of bets on a particular trial that was also randomly selected and was not known to Ss in advance. The actual payoff schedule, and the written instructions given Ss, are attached as the Appendix.

On any single run the S's payoff increases monotonically with the number of points he happens to have bet on the E-state (i.e. proportion of blue chips) that turn out to be correct. However, the magnitude of the payoff is not directly proportional to the number of chips on the correct outcome, so that S's best strategy, unless he is quite certain of the correct answer, is to distribute his betting points rather uniformly across all four possible outcomes. In fact, this payoff schedule has the mathematical property that the best strategy, on each trial, is to distribute the betting points in exact proportion to the objective probability of each of the four E-states as determined by the standard Bayes formula for inverse probability, applied to existing evidence. This differs from the conventional betting task in which S's best strategy (from an external viewpoint) is to go "all out" for the most likely alternatives.

The above property of the payoff schedule is obtained by dividing all the possible payoffs for any distribution of bets by the square root of the sum of squares of the components, a quantity that increases directly with the standard deviation of the distribution. Thus, if the bets are equal on each of the four proportions, the adjusting denominator is minimum: if all 12 points are bet on one proportion, the adjusting denominator is maximum. The former yields the highest <u>average</u> payoff over all four states but the lowest maximum payoff on any single state, whereas the converse is true of the single proportion bet.

As an illustrative numerical example, the payoff schedule for the distribution (6, 3, 2, 1) is obtained as follows:

a) $\sum x^2 = 6^2 + 3^2 + 2^2 + 1^2 = 50.$

b) $\sqrt{\sum x^2} \cong 7.07$

c) $1/\sqrt{\sum x^2} \cong .141$

d) $6 \times .141 \cong .85$; $3 \times .141 \cong .42$; $2 \times .141 \cong .28$; $1 \times .141 \cong .14$

The tabled values correspond to two figures multiplied by 100. It will be seen that the sum of these payoffs, over all four outcomes, decreases toward the bottom of the table, even though the maximum payoff increases.

Five drawing sequences, or runs, were used in order to obtain several measures of each S's behavior. For each of these runs Ss were told that their actual payoff would be based upon a single trial, the trial number to be announced only after all 20 draws. Thus, if the critical trial was trial 13, and the correct proportion was 4/7, S's payoff would be determined by the number of betting points he had on the component 4/7 on the 13th trial, (adjusted as shown above). The Ss were told in advance that all proportions and critical trial numbers were equally likely. The actual proportions were 2/7, 4/7, 3/7, 5/7, and 3/7 blue chips, and the critical trials were 5, 13, 10, 16 and 6. Ss were asked to determine their payoffs from the tabled values and write them on their answer sheets.

3. Results [2]

In the theoretical paper cited above a number of properties are discussed

---

2. This work was done in part at the Computation Center at the Massachusetts Institute of Technology, Cambridge, Massachusetts. The assistance of Miss Florence Gray and Mrs. Susan Goldberg is gratefully acknowledged.

that may characterize epistemic systems. Some of these properties are descriptive and some are expressly normative, defined as rational reactions to the available information.

Although a direct examination of the present data for each of these theoretical properties would be highly desirable, such an examination is not warranted either by the data themselves or by the analytical techniques presently at hand. Instead the data will be examined for general characteristics that are suggested by the more abstract theoretical properties, without any attempt to establish a precise correspondence. The actual properties investigated begin with rather static features and proceed to the more kinetic aspects of change and reaction to evidence.

Several types of questions are considered. First, what properties do the belief-states of these Ss exhibit in general and on the average? Second, how do these properties depart from randomicity, on the one hand, and from Bayesian standard on the other? Third, how do the belief-states seem to change as a function of increasing experiences in the proportion estimation task. And, fourth, how do individual Ss differ from each other in respect to these properties?

The normative standard distributions were computed by direct application of the Bayes' formula for inverse probability. The a priori probabilities, as announced, were uniformly 3/12 for all four E-states. If the first chip drawn was white, the conditional probability of a white draw, given the E-states 2/7, 3/7, 4/7, 5/7, respectively, would be 5/7, 4/7, 3/7, 2/7. The a posteriori probability of each of the E-states is determined by multiplying the corresponding conditional probability by the a priori probability, and dividing by the sum of these joint probabilities across all four E-states. Thus, the probability of the E-state 3/7 given a white draw, is

$$\frac{3/7 . 3/12}{2/7 . 3/12 + 3/7 . 3/12 + 4/7 . 3/12 + 5/7 . 3/12} = 9/42$$

On subsequent trials the normative probability estimates are computed in exactly the same way except that the new a priori probability is the a posteriori probability for the preceding trial.

Some apology may be in order for the rather heterogenous, not to say heterodox, statistical techniques here employed. The choice of techniques was in part determined by a wish to avoid the use of parametric statistics for indices with unknown or clearly unsavory distributions: it was also influenced, however, by the exploratory mission of this study. The findings noted below are not regarded as conclusive in any way but merely as promising guidelines for further research.

Descriptive Similarity of Subject Belief-States and Bayes Estimates

The first set of results is concerned with the similarity of the S's belief-states to the normative standard. Of course, it is not yet clear what level of similarity might be expected, but the present results give some general impression of this. They also indicate certain specific points of difference that should be noted for further investigation.

Table 1 compares the normative probability vector (top line) with the mean of subjects' vectors at the 14th trial on each of five urns. The 14th trial was selected primarily because 14 is an even multiple of the number of chips in the urn; hence it was thought that any difficulty due to conversion of fractions might be least pronounced at this juncture. And, in fact, the agreement between normative and empirical probability vectors does appear to be surprisingly good, with no obvious pattern among the discrepancies.

The apparent tendency for Ss to overestimate the very low probabilities may be an artifact of the rather coarse units of betting. It may be less distasteful to Ss to assign probability 1/12 to a very small non-zero probability than to call it zero, even though the latter is more nearly correct. In any case, the mean over-estimation of very low probabilities is due to a very few Ss. The standard deviations for the empirical estimates are also included, and, as might be expected, there is a rather close correlation between the mean and the standard deviation.

As a second index of overall similarity between Ss belief-states and the normative vectors, correlations were computed between each S's bet profile and the corresponding Bayes estimate on each trial. Inspection of these correlations indicated that they were very unstable for the first few trials because of the presence of near-zero variances for the uniform bets (3, 3; 3, 3 and approximations). Hence only the correlations for the last 15 trials were used here and in later analyses. The medians, across Ss, of these product moment correlations are presented as Table 2.

One general observation that applies to this table is that the correlations tend to be very high, particularly in view of the fact that they are attenuated by rounding. Except for the highly deviant value at the beginning of the second run, all correlations are above .70, and the overall median is approximately .90.

The following brief illustrative correlation table will give some basis for comparison with these data.

		Belief States		
		B_1	B_2	B_3
B_1	(2, 5, 4, 1)	1.00	.76	.89
B_2	(3, 6, 2, 1)	.76	1.00	.94
B_3	(3, 6, 3, 0)	.89	.94	1.00

Thus a correlation of .90 corresponds to a mean square difference between bet profiles of about 1.0 (depending somewhat on the level of variance).

One other rather firm conclusion that can be drawn from Table 2 is that there is no strong trend in the correlations over time, either within runs or between runs. Inspection of the data gives the impression that the explanation for the absence of a within-runs effect may be a tendency to outrun the evidence as the final trial approaches. This could of course be checked by giving longer trial sequences or sequences of indefinite length. The absence of any clear-cut learning effect between trials may indicate either that no learning took place or that sampling differences among the urns masked any learning that did occur.

An analysis of variance performed on these correlations, using the median for an S on each of the runs, indicated significant differences among Ss ($F = 3.39$ with 23 and 92 degrees of freedom). Differences were also significant between runs ($F = 9.41$ with 4 and 92 df) but these appear to indicate sampling differences rather than any consistent trend.

Confidence and uncertainty

Following an examination of the general congruence between B-states and normative standards, the next step is to investigate more specific aspects of agreement or disagreement. The first of these concerns the amount of confidence S places in his judgement as to the liklihood of various proportions.

Least confidence is of course indicated by betting three points on each of the four states: maximum confidence is reflected by a bet of 12 points on a single proportion. For a given cumulative period of evidence there is a rational degree of confidence, which is precisely defined by the probability distribution of the Bayes' standard.

The index of confidence used here will be the familiar H measure of information theory. This measure has its highest value, 1.92, for the uniform bet distribution and its lowest value, .00, for a bet of 12 points on one proportion. To illustrate the method of calculation, H, for the B-states (2, 5, 4, 1) = -(2/12 log 2/12 + 5/12 log 5/12 + 4/12 log 4/12 + 1/12 log 1/12) = 1.78 bets.

Table 3 presents the average H values for the 24 Ss on each of the 20 trials on each run, and the corresponding normative H value. Both sets of values decrease over the 20 trials on each run, as would be expected: occasional reversals indicate that a surprising (low probability) draw has occurred. The empirical values are clearly lower in general than the normative values. This indicates that Ss tend to be more certain of their judgement than the evidence warrants. The number of Ss whose H values are lower than the normative for at least half the trials is 15 (of 24) on the first run; 24 on the second run; 19 on the third run; 9 on the fourth run and 11 on the 5th run. Although this result may again be due to sampling, there is some indication that Ss tend to lose their surplus confidence toward the later runs.

Finally, analysis of variance on the median H scores for Ss shows that there are consistent individual differences among Ss over the five runs (F = 1.82 with 23 and 92 df) as well as the expected difference among H scores attributable to the runs (F = 13.35 with 4 and 92 df). To single out several extreme cases, one S has a lower H score than the normative estimate on at least 18 trials in each of the five runs: several other Ss have lower H indices than the normative only on the second run and average considerably higher H values on all other runs.

Derived proportion estimate

Because the B-states measured here have as a reference system the proportion of blue and white chips in an urn, they can be converted into a weighted estimate of that proportion. For example, the B-state (1, 2, 4, 5) is converted into a proportion estimate by multiplying its components term by term by the proportions 2/7, 3/7, 4/7, 5/7 respectively, to obtain 49/(12x7) = .583.

The average derived proportions for all Ss are shown in Table 4 with several sets of comparison values. The "observed proportions" are the actual proportions of blue chips that have been drawn in 10 or 20 trials (these points were selected arbitrarily). The "Bayes estimate" is the weighted proportion based on the Bayes vector for the corresponding trials. It will be observed that the Bayes estimate lies between the observed proportion and 50%. This reflects the fact that the Bayes estimate is a compromise between the initial a priori estimate, always 50%, and the observed proportions. The Bayes estimate will approach the observed proportion asymptotically as the number of trials increases.

The chief point of interest with respect to these data is whether Ss tend to be more or less 'conservative' than the Bayes estimates -- that is, whether they are influenced more or less strongly by the observed proportions as compared with their a priori B-state. The evidence is that there is no consistent difference either way. For 6 of the 10 comparisons in Table 4 the mean S estimates are closer to 50% than the Bayes estimates, but a count of Ss above and below the

Bayes estimates reveals no strong trend either way. It will be shown in a later analysis, however, that the correspondence between these derived proportion estimates and the normative estimates may not be quite as close as appears here.

Trial to trial reactions to evidence

Assuming that Ss begin at about the same point as the normative standard (i.e., at maximum uncertainty) and that their final estimates are not greatly different, there are two rather distinct ways in which their intervening "paths" may differ from each other and from the normative. For purposes of visualization, the unidimensional analogs of these are shown in Figure 1. Path I is monotonic and continuous; Path II is monotonic but saltatory; and Path III is fluctuating but continuous. Path I proceeds from the starting point to the terminal point at a steady rate and without any waste motion. Path II always proceeds in the same direction but at an irregular rate. Path III may be relatively continuous but it traverses the same interval (on the y-axis) several times.

Indices of these two aspects of divagation from a simple path -- that is, of non-monotonicity and non-continuity--are rather more complex for the n-dimensional than for the unidimensional case illustrated. However, the use of the trial-to-trial auto-correlations leads to indices of both characteristics. These may be compared, as before, with the corresponding indices for the normative standard.

For an index of monotonicity, it is sufficient to consider the mean autocorrelation over the entire sequence of trials. Thus, while the similarity between initial and final B-states determines the maximum value of the mean trial-to-trial autocorrelation, the minimum value is limited only by the theoretical range of the correlation coefficients. It is possible, in principle, for an S to produce a sequence of B-states that has an almost perfect negative correlation between successive elements, yet terminates at a distribution of bets that is close to normative.

Table 5 presents these trial-to-trial correlations for the Bayes estimates and the mean correlations for the empirical B-states. As will be seen the Bayes autocorrelations are extremely high and the empirical B-state correlations, although very high also, are quite consistently lower than normative for all five runs. There are, however, pronounced individual differences: an analysis of variance[3] of the median autocorrelations for Ss for each of 5 runs resulted in an F ratio of 5.79 for Ss (with 23 and 92df, $P < .001$). The differences among the five runs are not significant.

The implication of these results is that Ss "reverse themselves" more, from trial to trial, than is called for by the evidence, and that there are consistent individual differences in this respect.

The second way in which the B-state 'path' may depart from a smooth monotonic curve is by consisting of a few comparatively large jumps rather than a larger number of small adjustments. This property, however, will be reflected in a large variance among the autocorrelations. That is, there will tend to be very high autocorrelations for those path segments in which little change takes place, and low autocorrelations across the salti.

The variance in autocorrelations was computed for each S on each of the 5

---

3. Although many of the indices considered in this study are not normally distributed, the mean or median values of such variables do not depart seriously from normality and parametric statistics have been applied to those central tendency measures.

runs. There are marked individual differences, ranging from a mean variance of .002 for one S to a mean variance of .379 for another. These data do not invite parametric treatment but a non-parametric test serves to confirm the differences among Ss. Ranking the variances for the first four runs (the mean variance is much lower on the fifth run), a coefficient of concordance of .488 is obtained (corresponding to chi-square of 44.9 with 23 df; $p < .01$). Again, comparing these data with the normative, there do not appear to be consistent differences between the means for Ss and the variances of the Bayes autocorrelations. The former are greater on runs 1, 3 and 4 but less on the other two runs.

Derived proportions

As noted earlier, any B-state implies a certain derived proportion of blue and white chips in the urn. Whenever a chip of one color or the other is drawn, the derived proportion should change in favor of that color. Table 6 shows the actual changes in the derived proportions for blue chip following draws of either color. Only changes of .01 or greater are considered in this tally.

It is clear that there are an appreciable number of anomalous changes, that is, decreases following blue draws and increases following white draws. Of all changes ($>.01$) in derived proportion estimates, the anomalous changes constitute 26.1%, 27.5%, 21,8%, 21.1% and 17.0% of the total on the five successive runs. Thus, there appears to be some reduction in the ratio of anomalous changes with increasing experience.

Examining these anomalous changes in more detail it seems likely that they should not be regarded as reactions to the immediately preceding evidence, but rather as delayed adjustments of the B-state to earlier evidence.

As evidence for this, it is possible to sort the anomalous changes into two classes: those that make the derived proportion closer to the Bayes standard, and those that make the derived proportion less correct than before. If Ss are simply making belated adjustments to earlier evidence, then the former should predominate. If not, then the most likely hypothesis is that Ss are subject to a kind of gambler's fallacy -- that is, they construe the sampling scheme as being partially or temporarily exhaustive.

A count of these anomalous changes for the third run (selected arbitrarily) shows them to be almost equally divided. That is, for about half the anomalous changes there is a decrease in the net discrepancy from the normative proportion estimate. The remainder must be regarded as errors arising from misperception of the draw or from false hypotheses as to the nature of the sampling process as suggested above.

As an additional analysis of the possible tendency for Ss to delay reactions to evidence, Table 7 shows, for run 3, the observed proportion of blue chips, the Bayes estimate, the mean S estimate, and the number of Ss overestimating on each trial. As these data indicate, the majority of Ss tend to overestimate the proportions on trials 1-13, in which the normative proportion is generally decreasing, and to underestimate on the remaining trials in which the normative estimate is increasing. Thus, there appears to be a slight but consistent "hysteresis" effect in the reaction to new evidence.

Commutativity

The final property of the empirical data to be studied will be that of the commutativity of evidence effects. This property hinges upon whether the overall effect of a number of pieces of diverse evidence depends upon the order in which that evidence is received -- in the present case, whether the effect of a certain number of blue and white chips, respectively, depend upon the particular sequence in which they are drawn.

As the present measurement procedure was based on actual chance processes, the investigation of these commutativity effects must depend upon a natural experiment -- that is, a case in which it actually happened that two runs produced the same overall ratio of blue and white chips but for which subsequent totals are markedly different. Fortunately, there are several cases of such coincidences that provide as good a test as can be expected.

For both run 1 and run 2 there are 5 blue chips and 7 white chips after the 12th draw. In run 1, however, only 2 of the 5 blue chips had turned up by the 6th draw, whereas 4 of the 5 blue chips had turned up by the 6th draw in run two. Comparing the Ss' derived proportion estimates on trial 12, they are seen to be higher on run 2 than on run 1 (for 16 of the 24 subjects). This would indicate a "primacy" effect, that is, Ss are influenced more by early evidence than by recent evidence.

It is possible to check this result using a second comparison, between run 1 and run 3. In this case, both urns yielded 7 blue chips and 11 white chips in 18 draws but, for run 1, three of the blue chips had turned up by the 9th draw whereas only 1 had turned up by that time in run 3. With respect to run 1, this comparison is thus in the opposite direction from the comparison with run 2 and the results again bear out the primacy effect: 15 of the 24 Ss offer higher estimates for run 1 in which the blue chips appeared sooner. The results of both these comparisons are presented in Table 8.

4. Discussion

With suitable allowances for the small sample of Ss and sampling runs investigated here, there are several results that appear to warrant further investigation.

1) As indicated by H-values, Ss tend to focus their B-states rather more than is warranted by the available evidence.

2) As indicated by autocorrelations, Ss tend to change B-states more from trial to trial than is required.

3) Subjects in general do not appear to change B-states either more or less smoothly than does the normative standard.

4) Anomalous changes -- those representing a change in derived proportion estimates contrary to the evidence -- are about evenly divided between belated adjustments and genuine errors.

5) There appears to be a slight lag in adjustment to the sampling evidence from trial to trial.

6) Ss in general appear to be more influenced by early evidence than by later evidence.

All in all, it appears that the B-states are surprisingly "good" -- that is, that they correspond quite closely to the normative standards. It is also clear that specific departures from the normative are consistent for individual Ss.

Although the present measurement technique and experimental procedure are generally satisfactory there are several changes that are indicated for future investigation.

One disadvantage of the present technique is the restriction to rather coarse units for expressing the B-states (that is, in twelfths). At present it is not clear that this can be avoided without elaborate instrumentation. Even a modest increase in the number of betting points to be distributed greatly complicates the payoff schedules.

One change which seems quite feasible, however, is an increase in the length of sequences studied, and perhaps a change from runs of constant length to runs of unknown length, so that any "end effects" would be avoided. A second change of procedure would be to base payoff on several unknown trials rather than on a single trial. This might have the effect of diminishing the "sporting" attitude that some <u>S</u>s appear to adopt -- that is, the tendency to make more highly resolved bets than their information warrants.

In order to study a broader class of B-state characteristics, and to measure them more precisely, it will evidently be necessary to work with "rigged" sampling sequences. For example, to investigate further such effects as the primacy effect here noted, it seems essential to contrast sequences that differ sharply from each other in terms of early and late evidence. Again, the investigation of continuous vs. discontinuous reactions to evidence will require sequences that are inhomogeneous in terms of local sampling properties. These and other modifications are being incorporated into research in progress.

Table 1

Normative Standards (x*), Mean Probability Components and Standard Deviations for Each of Five Runs at Trial 14

Run		Component 2/7	3/7	4/7	5/7
	x*	5.08	5.18	1.71	.13
1	$\bar{x}$	4.88	5.21	1.54	.54
	$\sigma\bar{x}$	2.52	2.02	1.38	.63
	x*	2.89	5.53	3.11	.46
2	$\bar{x}$	2.88	6.08	3.08	.25
	$\sigma\bar{x}$	3.15	1.75	1.93	.52
	x*	7.30	3.47	.71	.03
3	$\bar{x}$	7.42	3.87	.46	.33
	$\sigma\bar{x}$	3.20	1.94	1.12	1.40
	x*	.00	.09	1.59	10.32
4	$\bar{x}$	.17	.42	3.71	7.62
	$\sigma\bar{x}$	.80	.91	2.79	3.39
	x*	9.09	2.64	.26	.01
5	$\bar{x}$	9.00	2.33	.17	.21
	$\sigma\bar{x}$	2.61	1.93	.80	1.00

Table 2

Median Correlation Between Subject

Estimates and Bayes Estimates by Trial and Run

	Run				
Trial	1	2	3	4	5
6	.924	.268	.893	.804	.870
7	.913	.721	.929	.826	.901
8	.857	.974	.910	.790	.909
9	.907	.824	.962	.755	.962
10	.833	.943	.987	.828	.986
11	.755	.884	.957	.788	.974
12	.773	.747	.961	.831	.959
13	.896	.919	.937	.860	.938
14	.938	.827	.969	.840	.973
15	.926	.907	.914	.879	.976
16	.887	.943	.949	.945	.985
17	.791	.952	.884	.901	.972
18	.817	.895	.815	.876	.992
19	.868	.852	.727	.943	.986
20	.877	.738	.738	.897	.969

Table 3

Average H (Column S) and Normative H (Column N)

For Twenty Trials and Five Runs

	Run									
Trial	1		2		3		4		5	
	S	N	S	N	S	N	S	N	S	N
1	1.87	1.92	1.83	1.92	1.74	1.92	1.63	1.92	1.84	1.92
2	1.85	1.99	1.81	1.74	1.68	1.74	1.69	1.74	1.70	1.74
3	1.90	1.92	1.84	1.92	1.58	1.53	1.51	1.53	1.50	1.53
4	1.83	1.98	1.74	1.76	1.41	1.31	1.46	1.31	1.39	1.31
5	1.85	1.91	1.69	1.91	1.40	1.57	1.37	1.12	1.13	1.12
6	1.74	1.77	1.70	1.77	1.22	1.37	1.32	1.37	1.27	1.37
7	1.56	1.60	1.66	1.89	1.08	1.19	1.14	1.19	1.04	1.29
8	1.53	1.77	1.66	1.91	0.92	1.04	1.09	1.04	.91	1.03
9	1.38	1.62	1.50	1.86	.82	.89	.98	.89	.89	.89
10	1.55	1.76	1.41	1.86	.90	1.11	1.10	1.11	.98	1.15
11	1.50	1.81	1.34	1.82	.70	.96	.98	.96	.83	.96
12	1.42	1.73	1.29	1.73	.74	.84	.88	.84	.89	1.17
13	1.35	1.62	1.22	1.77	.79	1.03	.77	.84	.83	1.03
14	1.29	1.52	1.20	1.70	.91	1.23	.76	.63	.66	.92
15	1.16	1.39	1.26	1.72	1.07	1.39	.87	.80	.64	.80
16	1.23	1.52	1.25	1.70	1.04	1.27	.78	.70	.70	.98
17	1.27	1.59	1.27	1.66	1.08	1.42	.81	.86	.50	.86
18	1.16	1.51	1.11	1.64	1.04	1.51	.72	.76	.44	.76
19	1.08	1.43	1.11	1.60	1.08	1.55	.66	.66	.34	.66
20	.94	1.34	1.04	1.56	1.04	1.50	.77	.83	.45	.83

Table 4

Derived Estimated Proportions of

Blue Chips with Comparison Proportions

	Run				
Trial 10	1	2	3	4	5
Observed proportion	.40	.50	.20	.80	.20
Bayes estimate	.43	.50	.33	.66	.34
$\bar{x}$ subject estimate	.42	.47	.35	.62	.36
Trial 20					
Observed proportion	.35	.55	.40	.75	.25
Bayes estimate	.37	.54	.41	.68	.31
$\bar{x}$ subject estimate	.38	.47	.38	.65	.31

Table 5

Autocorrelations Between B-states on Trial k and Trial k+1:

Normative Standards (r^*) and Subject Averages ($\bar{r}_s$)

	Run 1		Run 2		Run 3		Run 4		Run 5	
Trial	r^*	$\bar{r}_s$	r^*	$\bar{r}_s$	r^*	$\bar{r}_s$	r^*	$\bar{r}_s$	r^*	$\bar{r}_s$
5-6	.94	.69	.93	.72	1.00	.93	.99	.88	.99	.88
6-7	.98	.82	.82	.76	1.00	.95	1.00	.90	1.00	.91
7-8	.94	.85	.74	.85	1.00	.96	1.00	.90	.82	.92
8-9	.96	.87	.83	.87	1.00	.92	1.00	.93	.93	.95
9-10	.90	.87	.83	.79	1.00	.92	.99	.88	.92	.92
10-11	.85	.89	.88	.94	1.00	.97	1.00	.89	1.00	.93
11-12	.94	.90	.93	.88	1.00	.94	1.00	.90	.99	.94
12-13	.96	.82	.90	.95	.99	.97	1.00	.93	1.00	.95
13-14	.97	.91	.94	.86	.97	.97	1.00	.92	1.00	.87
14-15	.98	.87	.89	.90	.97	.94	1.00	.87	.98	.96
15-16	.94	.90	.94	.90	.99	.92	1.00	.74	.80	.94
16-17	.93	.97	.78	.76	.96	.92	1.00	.95	1.00	.95
17-18	.97	.91	.90	.83	.94	.97	1.00	.94	1.00	.98
18-19	.98	.95	.92	.83	.95	.88	1.00	.85	1.00	.97
19-20	.98	.93	.97	.91	.98	.92	1.00	.87	1.00	.99

Table 6

Changes in Derived Proportion Estimates (for Blue Chips)

Following Blue Chip and White Chip Draws

	Blue Draw			White Draw		
	Increase	Unchanged	Decrease	Increase	No Change	Decrease
Run						
I	51	64	29	43	116	153
II	91	107	42	30	87	99
III	75	85	32	29	91	144
IV	172	127	37	20	59	41
V	42	57	21	25	138	173

Table 7

Trial by Trial Changes in Derived

Proportion Estimates (run 3)

Trial	Observed Proportion	Bayes Estimate	Mean S Estimate	No. of Ss (out of 24) Overestimating
1	.00	.45	.49	19
2	.00	.41	.46	21
3	.00	.38	.43	18
4	.00	.35	.39	21
5	.20	.38	.40	13
6	.17	.36	.38	15
7	.14	.34	.37	16
8	.12	.33	.35	14
9	.11	.32	.35	15
10	.20	.34	.35	11
11	.18	.33	.34	15
12	.16	.32	.35	17
13	.23	.33	.35	12
14	.29	.35	.35	8
15	.33	.37	.36	5
16	.31	.36	.37	13
17	.35	.38	.37	6
18	.39	.40	.39	4
19	.42	.43	.38	2
20	.40	.41	.38	4

Table 8

Estimated Probabilities at Points of Comparison

Indicating Primacy Effects

	Run					
	1		2		3	
Trial	Actual	Estimate	Actual	Estimate	Actual	Estimate
6	.333	.441	.667	.502	---	---
9	.333	.407	---	---	.111	.350
12	.417	.413	.417	.463	---	---
18	.388	.411	---	---	.388	.387

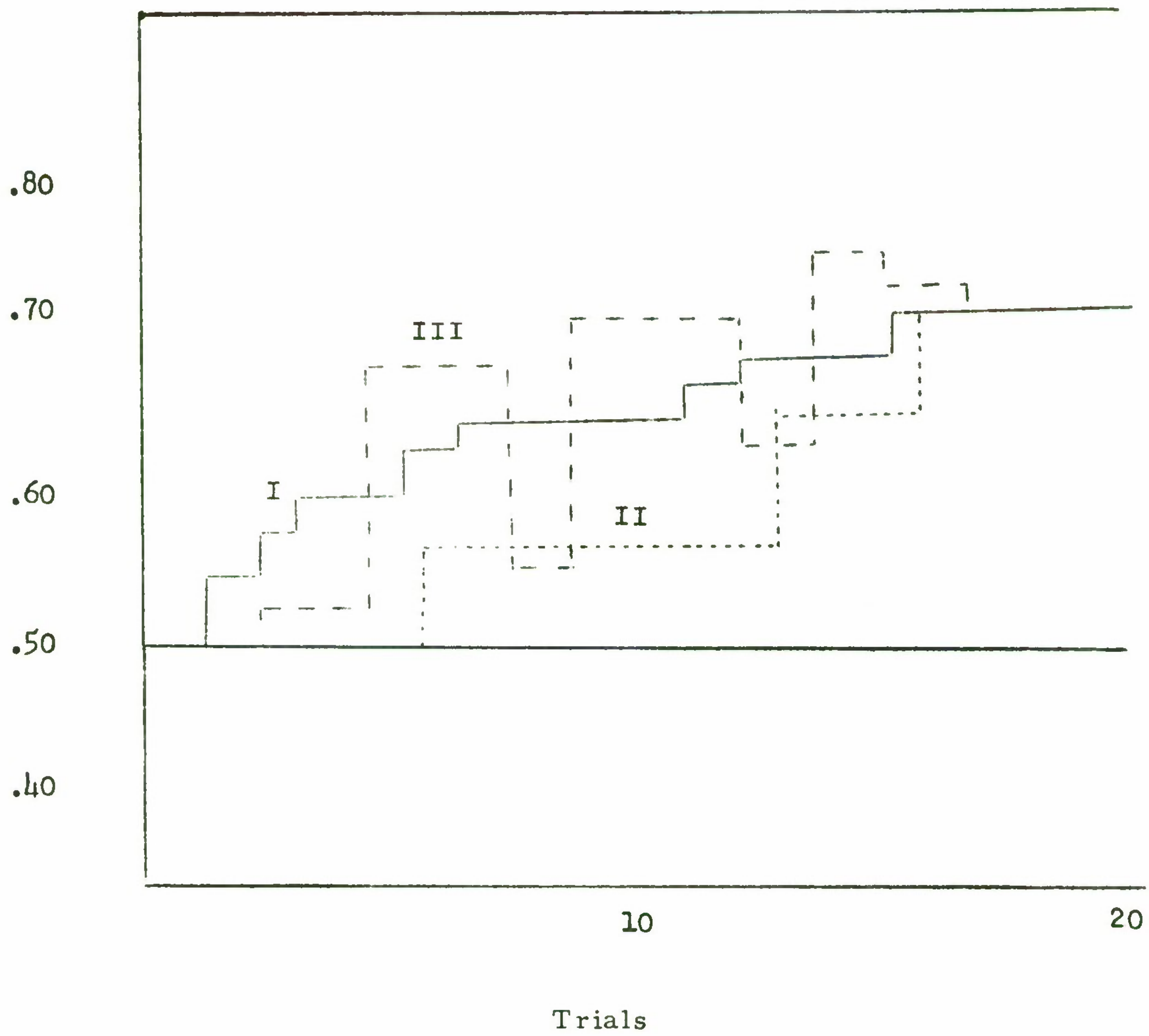

Fig. 1: Schematic comparison of a continuous monotonic (I); a discontinuous path (II); and a non-monotonic (III).

Appendix

Instructions

The following task tests your ability to make certain types of estimates on the basis of incomplete evidence. Your estimate will concern the likelihood of various numbers of blue chips in a box containing seven chips, blue and white mixed. A chip will be drawn at random from the box, you will estimate the number of blue chips, then the chip will be returned to the box. This will be done twenty times. You may never have enough information to be absolutely certain about the numbers of blue and white chips, but you should have a fairly good idea of the most probable distribution by the twentieth draw. A good estimate will be rewarded with real money -- the better the estimate, the bigger the payoff.

The box will contain seven chips in all, and of these seven, either 2, 3, 4 or 5 will be blue. As we have set up the experiment, any number from two to five will be equally likely. You will then distribute a total bet of 12 cents over all of these four possibilities, allotting the largest number of cents to those possibilities which you think most likely, and the fewest or no cents to those possibilities which you think less probable. An example will help to make this clearer.

The attached table shows the possible bets that you can make and the corresponding winnings depending on what the correct number is. Notice, to begin with, that you have exactly 12 cents to bet and you must bet 12 cents each trial or your scores will not count. The four parenthesized columns show how you spread these 12 cents over the 4 possible numbers. For example, 6 4 1 1 means that you bet 6 cents on the number that you considered the most probable; 4 cents on the second most probable; and 1 cent each on the last two which you consider equally improbable (but not impossible). Suppose that you had made this bet (look at your payoff sheet); betting 6 cents on 3, 4 cents on 4, and 1 cent each on 2 and 5. This means that you consider a distribution of 3 blue chips out of 7 slightly more probable than a distribution of 4 blue out of 7 and both of these much more likely than 2 or 5 blue. The payoffs which are shown in the right hand column represent your winnings depending on which of these turns out to be correct (as announced after 20 drawings).

You have an opportunity to revise your bet after each drawing from the box and it is recommended that you revise your bet to correspond to what you consider a reasonable set of probabilities. There will not actually be money on the line on each trial -- in fact there will only be one trial of the twenty that will count. However, you will not know which trial that is until the entire sequence of 20 drawings from one box has been completed. So you should bet on each trial as though it were the critical payoff trial. The critical trial will in fact be chosen at random. Any trial from 1 to 20 is equally likely. The critical trial will change each time we go through the task. Thus your winnings on any particular run will depend on the amount of money that you had bet on the correct number on the critical trial.

Suppose that you actually bet 1 6 4 1 on the critical trial. Then, if the correct number turned out to be 3, you would make 82 cents for having bet 6 cents on the correct number. If the correct number happened to be 4, you would make 54 cents, and if the number was 2 or 5 you would make only 14 cents. Notice that if you had bet all 12 cents on the correct number you would have received $1.00, but you would have received nothing for second best. The system of payoffs used here may seem complicated, but it is based on a mathematical formula, and it is fair if you bet according to your best information. Of course the kind of bet you make is up to you, but your earnings will probably be greatest if you use the more one-sided bets only when you are rather sure of the right answer. In other word, in order to make the most money you should not be either too cautious or too reckless.

You may not erase any bet you have made once you have set it down, so you should decide how you will distribute your 12 cents before you start to write.

(However, if you notice that the sum does not add up to 12 you may change <u>one number</u> by putting a plus or minus correction next to the number in order to cancel the discrepancy. Thus if you find that your total bet adds up to 14 and you decide that you have placed too much of the weight on 3, say, you may place a minus 2 (-2) next to your bet on 3 and this will be an acceptable bet. Any further changes will disqualify you <u>on that trial</u>, but you should continue betting on subsequent trials.)

The first run will be practice to familiarize you with the procedure and to clear up any confusion you may have about the nature of the bets. We will go through the entire task exactly as we will in the 5 runs for money, but you will be permitted to ask questions or to ask for help if you need it on this one trial. (Of course, there will be no payoff on the practice run.)

One further word about your money. We will go through the task for keeps

5 times, and you will be "charged" 50 cents for each run. This is because you could be sure of making 50 cents just by a bet which assumes all four numbers are equally likely (that is, 3 3 3 3). Thus in order to win any money, you must take a chance on one or another of the other bets. Your final earnings will be your winnings on all five runs less the 50 cent charge for each run. (Of course, though, if you come out in the red over all 5 trials you won't have to pay us -- we will call it even.) If you are very lucky or very clever you may make 50 cents on each trial, and win as much as $2.50. However, it is very unlikely that any single individual will make out this well.

Each of the 5 runs is independent of all of the other runs -- that is, whatever the correct number on the first run, the next run may have the same distribution of blue and white chips, or it may have a different one; the number of blue chips in one run does not in any way affect the number of blue chips in any other run. Remember that 2, 3, 4 and 5 are all equally likely.

Name_______________________________ Run Number_______________

Bets on Estimated Number of Blue Chips

Indicate after each drawing of a blue or white chip how you would distribute 12 cents worth of bets among the four possible numbers of white chips listed below. Your bets must total 12 cents. Remember that there are seven chips in the box.

Number of Blue Chips

Trial No.	2	3	4	5
1	____	____	____	____
2	____	____	____	____
3	____	____	____	____
4	____	____	____	____
5	____	____	____	____
6	____	____	____	____
7	____	____	____	____
8	____	____	____	____
9	____	____	____	____
10	____	____	____	____
11	____	____	____	____
12	____	____	____	____
13	____	____	____	____
14	____	____	____	____
15	____	____	____	____
16	____	____	____	____
17	____	____	____	____
18	____	____	____	____
19	____	____	____	____
20	____	____	____	____

Critical trial______ Correct number_______ Your winnings_________

Payoff Table for Bets on Estimated Number of Blue Chips

Most probable Number		Second most probable		Third most probable		Least probable	
Cents Bet	Win if Right	Cents Bet	Win if Right	Cents Bet	Win if Right	Cents Bet	Win if Right
(3)	50	(3)	50	(3)	50	(3)	50
(4)	65	(3)	49	(3)	49	(2)	32
(4)	63	(4)	63	(2)	32	(2)	32
(4)	62	(4)	62	(3)	46	(1)	15
(4)	58	(4)	58	(4)	58	(0)	0
(5)	77	(3)	46	(2)	31	(2)	31
(5)	75	(3)	45	(3)	45	(1)	15
(5)	74	(4)	59	(2)	29	(1)	15
(5)	71	(4)	57	(3)	42	(0)	0
(5)	69	(5)	69	(1)	14	(1)	14
(5)	68	(5)	68	(2)	27	(0)	0
(6)	87	(2)	29	(2)	29	(2)	29
(6)	85	(3)	42	(2)	28	(1)	14
(6)	82	(3)	41	(3)	41	(0)	0
(6)	82	(4)	54	(1)	14	(1)	14
(6)	80	(4)	53	(2)	27	(0)	0
(6)	76	(5)	64	(1)	13	(0)	0
(6)	71	(6)	71	(0)	0	(0)	0
(7)	92	(2)	26	(2)	26	(1)	13
(7)	90	(3)	39	(1)	13	(1)	13
(7)	89	(3)	38	(2)	25	(0)	0
(7)	86	(4)	49	(1)	12	(0)	0
(7)	81	(5)	58	(0)	0	(0)	0
(8)	96	(2)	24	(1)	12	(1)	12
(8)	94	(2)	24	(2)	24	(0)	0
(8)	93	(3)	35	(1)	12	(0)	0
(8)	89	(4)	45	(0)	0	(0)	0
(9)	98	(1)	11	(1)	11	(1)	11
(9)	97	(2)	22	(1)	11	(0)	0
(9)	95	(3)	32	(0)	0	(0)	0
(10)	99	(1)	10	(1)	10	(0)	0
(10)	98	(2)	20	(0)	0	(0)	0
(11)	99	(1)	9	(0)	0	(0)	0
(12)	100	(0)	0	(0)	0	(0)	0

References

Roby, T. B. Belief states, evidence and action. Paper read at Air Force Aerospace Symposium on "Predecisional Processes in Decision Making." April, 1962. (Mimeo copies available on request.)