ORGANIZATIONAL BEHAVIOR AND HUMAN PERFORMANCE 5, 183–190 (1970)

Sample Proportions and Subjective Probability Revisions[1]

LEE ROY BEACH, JAMES A. WISE, AND SCOTT BARCLAY
University of Washington

The hypothesis was that the proportions of poker chips in the displayed samples influence subjective probability revisions that are obtained in "book bags-and-poker chips" experiments. Subjects revised for simultaneous and sequential samples from two 80%–20% symmetrical binomial populations and two 70%–30% symmetrical binomial populations. Sample proportions account in large part for the revision responses for both kinds of populations for simultaneous samples. For sequential samples, however, proportions appeared to have less influence on revision responses even though 62% of the subjects claimed to use them. The implications are discussed.

In subjective probability revision experiments it is often found that subjects' revisions are less than the amount prescribed by the normative model, Bayes' theorem. This finding, called conservatism (Phillips & Edwards, 1966), has stimulated a good deal of research aimed at finding the reasons for its occurrence (see Peterson & Beach, 1967). The results suggest, as one might expect, that conservatism is not just one thing. Anything that interferes with the subjects' understanding of the task, their willingness to give unbiased answers, their ability to do the appropriate mental or intuitive mathematics, etc., will tend to yield results that can be labelled conservative.

While the amount and kind of conservatism varies from one experiment to another, there is one task for which most experiments obtain essentially the same results. This has come to be called the "book bags-and-poker chips" task, the mechanics of which will be explained in a moment. The point is that this apparently simple experimental task seldom fails to obtain a very predictable degree of conservatism, even though the conservatism obtained in more complex tasks may vary a good deal or be nonexistent. As a result of this pleasing stability the book bags-and-poker chips experiment has become the prototype for investigations of subjective probability revisions and of conservatism.

[1] This research was supported by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant No. 1311-67 and AFOSR-69-1701.

The usual book bags-and-poker chips experiment consists of a hypothetical game that begins by presenting the subjects with two real or imaginary book bags full of poker chips. The subjects are told that one bag contains, say, 70% red chips and 30% blue and the other contains 30% red and 70% blue. Then out of the subjects' sight the experimenter supposedly flips a coin and selects one of the book bags. He then samples from the selected bag and shows the sample to the subjects. In light of the obtained sample, the subjects revise their original .50–.50 subjective probabilities about which bag was chosen to reflect their new opinions about the probabilities of the two bags. Note that the sample can be shown either all at once (simultaneous) or cumulatively, one chip at a time (sequential).

Bayes' theorem, the normative model for revising probabilities in this situation, starts with the .50–.50 prior probabilities that are dictated by the coin flip, and modifies them using the appropriate likelihoods from .70–.30, and .30–.70 binomial sampling distributions. The modified probability is called a posterior probability and it is the number that the subjects would give if they were successfully operating as "intuitive statisticians" (Peterson & Beach, 1967). When the Bayesian values and the subjects' data are plotted, the curve for the latter usually lies below the curve for the former; the data are conservative relative to the Bayesian prescription.

Some of the possible bases for conservatism in the book bag-and-poker chip task have been investigated, e.g., response modes and payoffs (Phillips & Edwards, 1966) and sample sizes (Peterson, Schneider, & Miller, 1965). A simple hypothesis that has received no attention, however, is that subjects may merely use their estimate of the proportion of the predominant color of poker chips in the displayed sample as the basis for their responses.[2] This strategy could not always be applied, of course; for a sample of $n = 1$ red chip, no subject is going to say that the probability is 1.00 that the predominantly red bag was chosen. However, it is possible that for more than three or four chips in a sample that the sample proportion begins to influence the responses. The purpose of this experiment was to investigate the hypothesis that sample proportions can account for at least some of the conservatism in subjects' revision responses.

[2] After this experiment was completed, we found a previous study in which subjects reported using sample proportions as the basis of their judgments (Kriz, J. Der Liklihood-quotient zur Erfassung des Subjectiven Signifikanzniveaus., Forschungsbericht No. 9, Institut für höhere Studien und Wissenschaftliche Forschung, Wien, 1967).

METHOD

Two large urns were drawn on the blackboard (book bags are so small that some subjects assume that the composition changes when the experimenter samples without replacement). These were described as containing a large number of red and blue poker chips and the percentages were written on the urns. Then the coin flipping and the revision procedures were explained and a series of hypothetical games were played. The usual instructions about relying on intuition and refraining from making calculations also were given.

Subjects. One hundred sixty-nine volunteer college men served as subjects in groups of from 1 to 15 at an experimental session.

Urn Compositions. Sixty-four of the subjects were told that the urns contained 70% of one color of chips and 30% of the other; the two urns were symmetrical. Forty-five subjects worked with 80%–20% urns.

Samples. Forty-eight (27/64 and 21/45) subjects were shown simultaneous samples and 61 (37/64 and 24/45) were shown sequential samples. The samples were the same for both kinds of urn proportions.

There were 17 simultaneous samples ranging from $n = 4$ to 10. They were selected to contain all of the possible proportions that could be obtained within the constraints of sample sizes and desirable S-F values (e.g., $n = 4$ can only have proportions of 1.00, .75, and .50 for the predominant color). For each experimental session about one-half of the samples favored one urn and the remainder favored the other urn. The order of sample presentation differed from session to session.

There were three sequential samples of $n = 10$. These were constructed to include the proportions represented in the simultaneous samples. Two of the three samples were presented at each experimental session; one-third of the subjects saw sequences one and two, one-third saw sequences two and three, and one-third saw sequences one and three.

Revisions. Subjects gave their subjective probability revisions by dividing a 100-point scale into two parts, one part represented the subjective probability that the red urn had been selected and the other part represented the probability of the blue urn. One revision was made for each of the simultaneous samples. Ten revisions, one after each chip, were made for each of the sequential samples.

Odds Estimations. It has been argued (Phillips & Edwards, 1966) that odds estimates are the proper response mode for revision experiments. It is possible that the use of the 100-point scale for revision might bias the subjects to depend upon sample proportions. Therefore, another group of twenty-four subjects was shown the 17 simultaneous samples for 70%–

30% urns and asked to state their revised probabilities in the form $X:1$, where X is the number of times more likely the most probable urn was than the least probable urn. These odds estimates were then converted to probabilities for comparison with the other data, $p = X/X + 1$.

RESULTS

Simultaneous Samples. Figure 1 shows the relationship of the subjects' mean revision estimates to the Bayesian posterior probabilities and to the sample proportions for simultaneous samples; medians are almost identical to the means. The estimates are more similar to the proportions than they are to the Bayesian values in not just their "conservative" magnitudes, but also in the pattern of their distributions. Clearly, proportion accounts for the subjects' estimates better than probability does. How-

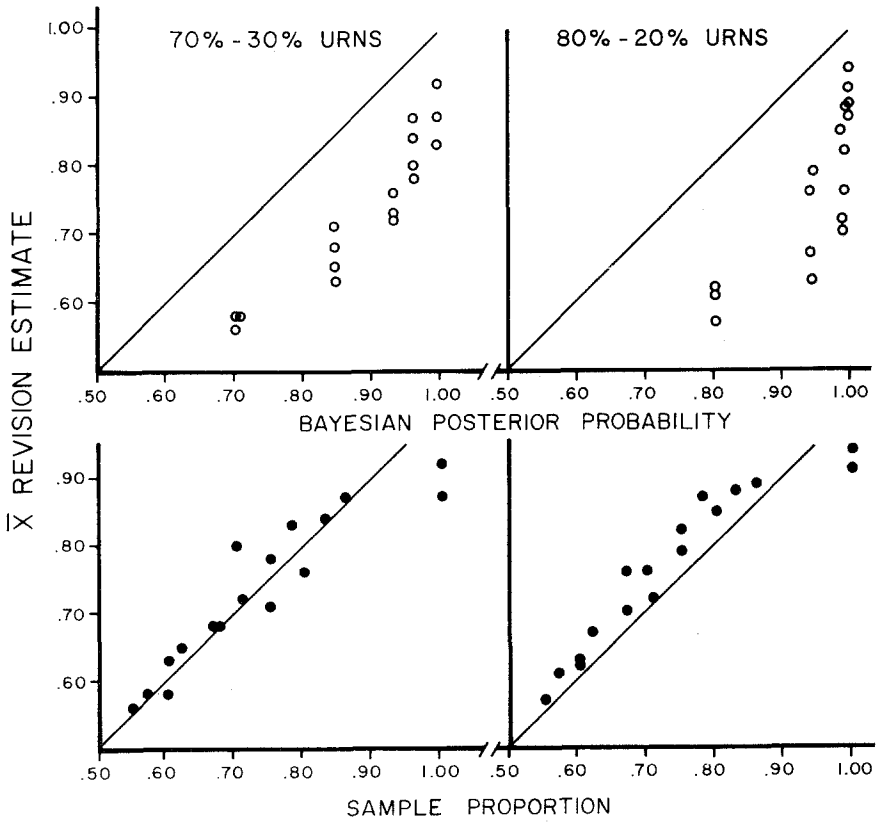

FIG. 1. Relation of mean revision responses to Bayesian posterior probabilities and to sample proportions for simultaneous samples.

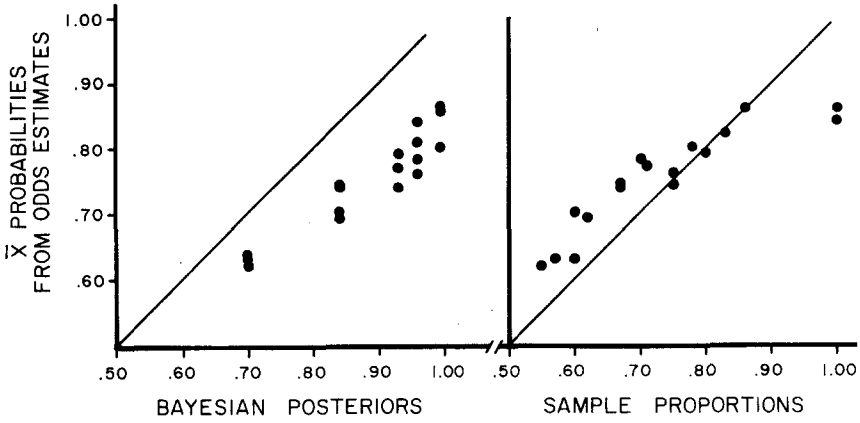

FIG. 2. Relation of mean odds revision (transformed to probability) to Bayesian posterior probabilities and to sample proportions for simultaneous samples from 70%–30% urns.

ever, the composition of the urns was not entirely ignored, the responses for the 80%–20% urns are slightly higher than the proportions.

It is possible that the results in Figure 1 are due to having solicited probability estimates, which are very like proportion estimates, rather than odds. Figure 2 shows the relation of the probabilities inferred from odds estimates to the Bayesian posteriors on the left and to the sample proportions on the right. Clearly the odds, like the revisions in Figure 1, are better accounted for by the sample proportions than by the Bayesian posterior probabilities.

Sequential samples. Figure 3 shows the relationship of the subjects' revisions for sequential samples to the Bayesian probabilities and to the sample proportions; the empty points are for each of the first three chips in each sequence. Even for $n > 3$ chips, the results are not the same as those obtained for simultaneous samples. Here the estimates appear to be a conservative, reasonably orderly version of the Bayesian probabilities, and the conservatism does not seem to be a straightforward function of sample proportion.

Figure 4 provides a more detailed view of the sequential results. It contains a draw-by draw comparison of the Bayesian probabilities for the 70%–30% urns and the 80%–20% urns (B7 and B8, respectively), the corresponding mean revisions (S7 and S8), and the sample proportions (%) for each of the three sequential samples. For $n > 3$ the responses could as well be inflated sample proportions as they could be conservative Bayesian probabilities.

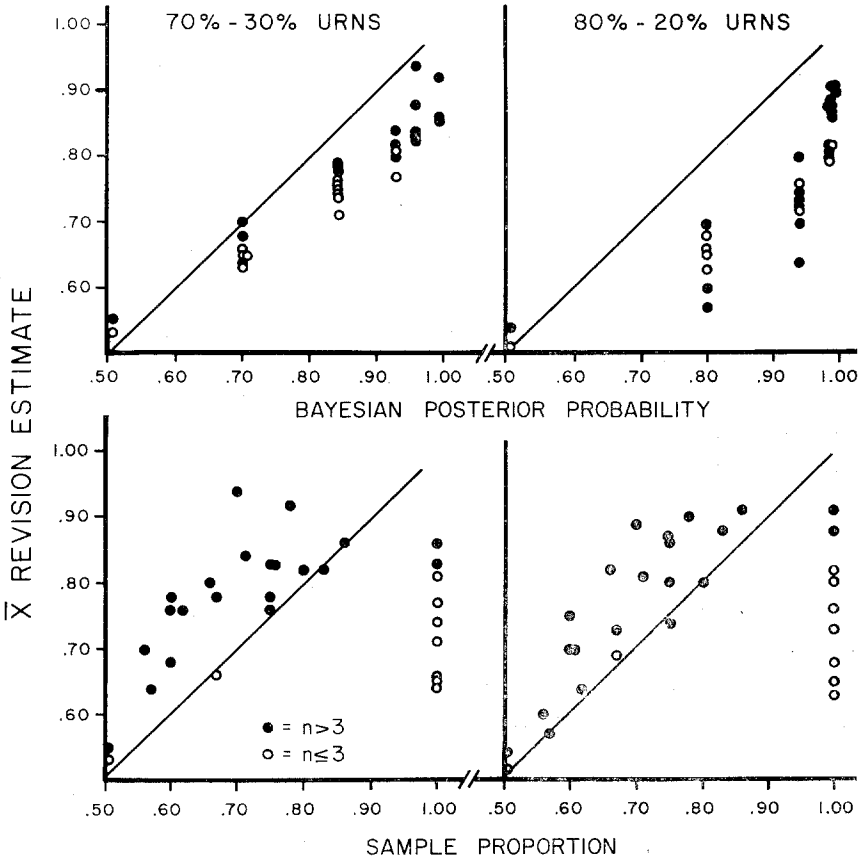

FIG. 3. Relation of mean revision responses to Bayesian posterior probabilities and to sample proportions for sequential samples.

DISCUSSION

The results for simultaneous samples show that sample proportions account for the estimates better than the Bayesian probabilities do. This is not to say that the picture is clear by any means; in the lower part of Figure 1 the results for the 80%–20% urns differ somewhat from those for the 70%–30% urns, which precludes a solely proportion interpretation.

The results for the sequential samples suggest that subjects were ignoring proportions and were attempting to be Bayesian. It would be nice if this were the case because there is a sizeable literature on conservatism in sequential book bags-and-poker chips tasks. However, as Bayesian psychologists we reserve the right to interject our own opinions and the data on which they were formed. After each experimental session with

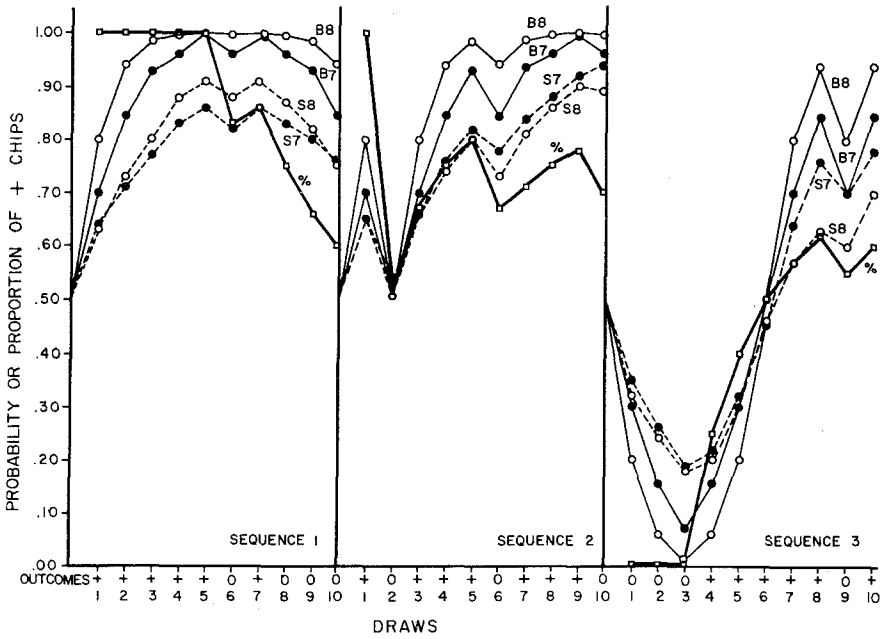

FIG. 4. Draw-by-draw comparison of Bayesian posterior probabilities, mean revision responses, and sample proportions for sequential samples. See text for explanation.

sequential samples, the subjects were asked how they had performed the task. Three options were suggested by the experimenter: (1) intuition, as required by the instructions at the beginning of the experiment; (2) estimation of the proportions of the two kinds of chips in the samples; (3) any other strategy. Then they wrote their answers on their answer booklets and elaborated if they wanted. Sixty-two percent (38/61) of the subjects said that they based their answers on the sequential sample proportions. However, some elaborated by saying that for the first few draws they just revised upward to favor the predominant color. Then they began to use the sample proportion for the middle span of the sequence. Then, toward the end, they exceeded the proportion because the sample was preponderantly one color of chips and they knew that they were only going to see ten draws. Some subjects remarked that sample proportions are very compelling because they were available (and somehow relevant) numbers in a very difficult and foreign task.

It appears, therefore, that the revisions for both kinds of samples may be a combination of dependence on sample proportions and "Bayesian-like" revision of opinion in light of data. While recognizing that more and

better experiments need to be done on this question, we regard these data as sufficient cause to avoid use of the book bags-and-poker chips task and to attempt to devise tasks that do not provide subjects easy access to spurious response strategies.

Finally, it should be made clear that these results do not refute the existence of conservatism in subjective probability revision. Conservatism is found in other more complex tasks (e.g., Schum 1966; Edwards *et al.*, 1968). Moreover, appropriate training in the book bags-and-poker chips task can eliminate fallacious strategies and, even for simultaneous samples, bring about Bayesian behavior (Wheeler & Beach, 1968). The point of this experiment is that some of the data from book bags-and-poker chips experiments may not really speak to the question of subjective probability revision and therefore may hinder rather than aid in the development of the solid theory of subjective probability.

REFERENCES

EDWARDS, W., PHILLIPS, L. D., HAYS, W. L., & GOODMAN, B. C. Probabilistic information processing systems: Design and evaluation. *IEEE Transactions on Systems Science and Cybernetics,* 1968, *SSC-4,* 248–265.

PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistican. *Psychological Bulletin,* 1967, **68,** 29–46.

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.

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

SCHUM, D. A. Prior uncertainty and amount of diagnostic evidence as variables in a probabilistic inference task. *Organizational Behavior and Human Performance,* 1966, **1,** 31–54.

WHEELER, G., & BEACH, L. R. Subjective sampling distributions and conservatism. *Organizational Behavior and Human Performance,* 1968, **3,** 36–46.

RECEIVED: April 28, 1969