*Journal of Experimental Psychology*
1965, Vol. 69, No. 5, 522–527

# SAMPLE SIZE AND THE REVISION OF SUBJECTIVE PROBABILITIES [1]

CAMERON R. PETERSON,[2] ROBERT J. SCHNEIDER, AND ALAN J. MILLER
*University of Colorado*

This experiment investigates the relation between subjective probability and mathematical probability. Bayes' theorem provides the correct revision of probabilities as a result of new data. New data, however, may come in different sized samples. This experiment measured the accuracy of subjective probability revision as a function of sample size. The results show that accuracy decreases as sample size increases; the function is negatively accelerated. The gain in amount of data processed by increasing the sample size is at the expense of accuracy.

Behavioral decision theory provides a framework for evaluating the extent to which selected behavior of human *S*s corresponds with ideally consistent behavior as outlined by statistical decision theory (Edwards, 1961). The two primary variables of decision theory are probability and value; the corresponding psychological variables are subjective probability and utility. The present experiment studies the correspondence between subjective probability and mathematical probability.

Recent developments within statistical decision theory focus on the correct revision of probabilities in the light of new information—the problem of revising the probability of a hypothesis as a function of the occurrence of a relevant datum (e.g., see Schlaifer, 1961; or Raiffa & Schlaifer, 1961). These developments revolve around the concepts of personal probability and Bayes' theorem. Personal probabilities are ideally consistent opinions, and conform to the axioms of probability theory; subjective probabilities are the opinions of real persons, and may or may not conform to the axioms of probability theory. Bayes' theorem is a derivation from the axioms of probability theory. A consequence of Bayes' theorem which is relevant to this paper is

$$\frac{P(H_a|D)}{P(H_b|D)} = \frac{P(D|H_a)\,P(H_a)}{P(D|H_b)\,P(H_b)} \qquad [1]$$

or, more simply,

$$\Omega_1 = L\Omega_0 \qquad [2]$$

where $H_a$ and $H_b$ refer to two different hypotheses regarding the state of the world, and $D$ refers to a relevant datum which occurs. The $\Omega_0$ refers to the odds of $H_a$ to $H_b$ prior to the occurrence of $D$. The $\Omega_1$ refers to the corresponding odds posterior to the occurrence of $D$. The $L$ refers to a likelihood ratio; it is equal to the conditional probability of $D$ given $H_a$ divided by the conditional probability of $D$ given $H_b$. Thus Equation 2 specifies that the revision of the odds of $H_a$ to $H_b$ is a function of the occur-

[1] The research reported here was undertaken in the Behavior Research Laboratory, Institute of Behavioral Science, University of Colorado, and is Publication No. 42 of the Institute. A more detailed report of this experiment, including additional analyses, is available as Behavior Research Laboratory Report No. 41 (mimeo). The research was supported by a research grant (M-4977) from the National Institute of Mental Health. The authors are indebted to W. Edwards for his continual guidance. The many suggestions of K. R. Hammond and F. J. Todd and the assistance of Dorothy Boucher in the conduct of the experiment are gratefully acknowledged.

[2] Now at the University of Michigan.

rence of $D$. This revision is accomplished by multiplying the prior odds by the likelihood ratio to yield the posterior odds (for a detailed discussion of Equation 2 see Edwards, Lindman, & Savage, 1963, p. 218).

Experiments carried out by Edwards and his associates indicate that *S*s revise their subjective probabilities in the correct direction, but that the magnitude of revision is less than that specified by Bayes' theorem (e.g., Edwards & Phillips, 1963, 1964). Typically, in these experiments, *S*'s task has been to revise his subjective probability about which $H$ was generating a sequence of data. A revision was made after the presentation of each datum in the sequence.

Information which results in a revision of subjective probabilities need not be restricted to a single datum at a time. Information may be provided in a sample of data. But more data in a sample implies more information processing. For some types of human information processing, an increase in the amount of information presented to an $S$ impairs his accuracy; for other types of information processing, the chunking of more information into a presentation does not impair accuracy (Miller, 1956). The primary purpose of the present experiment was to investigate the function relating accuracy of data processing to the *size* of the *sample.* Accuracy was measured by the degree to which *S*s' revision of subjective probabilities agreed with the correct revision. Sample size referred to the specific number of data in the sample which provided the information for the revision of subjective probabilities.

## METHOD

*Experimental design.*—The experimental design required the manipulation of sample size and an evaluation of the effect upon accuracy in the revision of subjective probabilities. Urns with varying proportions of black marbles and white marbles were the *H*s. Each trial consisted of a random selection of 48 marbles (*D*s) from an urn. After each trial *S*s were shown which urn yielded the marbles.

*Subjects.*—Forty-four volunteer students from an introductory psychology course served as *S*s in groups of four.

*Apparatus.*—Two large beakers served as urns. Each urn contained approximately 1,000 marbles of ½ in. diameter. Large numbers of marbles were used so that, with continual mixing, sampling without replacement could be considered a reasonable approximation of sampling with replacement. Urn B contained three black marbles for every two white marbles; Urn W contained two black marbles for every three white marbles. In this symmetrical situation the ratio of most probable to least probable color was a constant 60:40 for both urns. This relatively uncertain ratio insured that *S*s were not required to make too many revisions near a ceiling of certainty, i.e., subjective probabilities of 1 or 0.

Four sample sizes were selected: 1, 4, 12, and 48. Relatively more small sizes were included because of the expectation that the function would be negatively accelerated. For samples of size 1, $E$ drew single marbles at a time. Dippers were made which would "dip" the appropriate number of marbles for each sample.

The response apparatus for obtaining subjective probability estimates was a 25-in. bar marked off into 100 equal units which were numbered from left to right. A sliding marker permitted $S$ to divide the bar into two sections. The length of the left section represented *S*'s subjective probability that the marbles observed were drawn from Urn B; the length of the right section represented *S*'s subjective probability that the marbles were drawn from Urn W. The length of the entire rod represented a probability of 1, so subjective probabilities for both urns were required to sum to 1. The number under the sliding marker represented the subjective probability of Urn B. In order to preclude problems inherent in responses of complete certainty, subjective probabilities were coded as no more extreme than .001 or .999.

Black drapes divided one side of the experimental room into four cubicles, each with an opening facing $E$. This permitted four individual *S*s to be run simultaneously; but the experimental atmosphere approximated the running of individual *S*s in terms of both communication between *S*s and rapport with *E*.

*Procedure.*—The experiment consisted of eight blocks of four trials each. Each trial consisted of one of the four sample sizes. During each trial *E* drew and displayed successive samples of marbles until the cumulative number of marbles drawn numbered 48. After each sample *S* was required to revise and record his subjective probability that *E* was drawing from Urn B. In order to avoid confounding the effects of sample size and within-block learning, the sample sizes were counterbalanced with respect to order of occurrence during a block.

The *S*s, four or fewer, were seated in cubicles facing *E*. A table in front of each *S* contained a probability bar and an answer sheet. The *E* instructed *S*s with respect to the content of the urns and the use of the probability bar. For example, if *S* expected that it was twice as likely that *E* was drawing from Urn B rather than from Urn W, he would set the marker on the probability bar so that the Urn B side was twice the length of the Urn W side. The *S*s were told that the purpose of the experiment was to compare their intuitive revision of subjective probabilities with the correct revision as calculated by means of a mathematical equation. There was a practice demonstration of the step-by-step procedure.

The procedure of each trial progressed according to the following steps. (*a*) The *E* flipped a coin and selected the designated urn. (*b*) The *E* placed the urn into an opaque box so that *S*s were not able to identify the urn. (*c*) Each *S* set the marker on the probability bar at .50. (*d*) The *E* drew at random a prescribed sample size of marbles and displayed the sample to *S*s. (*e*) On the basis of the sample, each *S* reset his probability marker and recorded his subjective probability of Urn B on his answer sheet. This sequence of draw-reset marker-record was repeated until *E* had drawn 48 marbles from the urn. (*f*) The *E* then showed *S*s which urn he had been drawing from. He then progressed to the next trial by flipping another coin to select the next urn. This procedure continued until the end of the experiment.

*Measure of probability revision.*—In order to compare the subjective probability revision of *S*s with the corresponding probability revision as calculated by means of Bayes' theorem, the data analyses required a common measure of revision. The measure used here was the same as that reported by Edwards and Phillips (1963). It is the $\log_{10}$ of the likelihood ratio (LLR). The likelihood ratio is described above in Equation 2; it is equal to the ratio of the posterior to the prior odds. Thus, the LLR of this experiment was equal to the log of the posterior odds in favor of Urn B minus the log of the corresponding prior odds. Since the LLR increases with the difference between the posterior and prior odds, it must also increase with the difference between the posterior and prior probabilities. It is thus a measure of revision from prior to posterior subjective probabilities; the measure is positive if the revision is upward and negative if the revision is downward.

The LLRs for the present experiment were calculated as follows. Posterior probabilities, both subjective and Bayesian, were obtained after each trial of 48 marbles. These posterior probabilities were converted into odds in favor of Urn B, and were then transformed into log posterior odds. The definition of LLR calls for a subtraction of log prior odds from log posterior odds. However, since the initial probability of Urn B on each trial was .50, the prior odds were always 1 and the log prior odds were 0. Therefore, the log posterior odds were equivalent to LLRs. An LLR calculated from Bayesian probabilities will be termed BLLR; one calculated from subjective probabilities will be termed SLLR. The BLLR in the present experiment was equal to .17609 times the total number of black marbles minus white marbles.

An important advantage of the LLR over the algebraic change in probabilities as a measure of probability revision is that the LLR is proportional to the amount of evidence in favor of one hypothesis over the other. Algebraic change in probabilities, on the other hand, is influenced by a ceiling of certainty; equal increments of evidence result in smaller expected algebraic probability changes near certainty than near uncertainty.

*Relation of SLLR to BLLR.*—Two dependent variables related SLLRs to corresponding BLLRs. The primary dependent variable, the *accuracy ratio*, was the ratio of SLLR to BLLR at the end of each trial of 48 marbles. If *S* changed his probabilities in the same algebraic magnitude as did Bayes' theorem, then his accuracy ratio would be 1. To the extent that *S* changed less than Bayes' theorem, his accuracy ratio decreased.

A second dependent variable, linearity, was the correlation between corresponding within-trial BLLRs and SLLRs. The LLRs were based on probabilities posterior to each sample, i.e., posterior to each probability


FIG. 1. Mean accuracy ratio as a function of sample size.

revision. A correlation was calculated for each trial of sample sizes 1 and 4; there were too few within-trial LLRs for sample sizes of 12 and 48.

## RESULTS

*Analyses of variance.*—In order to evaluate the significance of the effect of sample size upon the accuracy of subjective probability revision, analyses of variance were applied to the accuracy ratio and to linearity. The two main effects for these analyses were sample size and blocks of trials. These analyses of variance are presented in Table 1. Whenever the number of black and white marbles was exactly equal, BLLR had a value of 0 and thus the accuracy ratio was indeterminate. As a constantly conservative procedure, each indeterminate value was replaced by the grand mean of the data in the analysis of variance and the degrees of freedom associated with the highest order error term were reduced by one.

Table 1 indicates that sample size had a significant ($p < .001$) effect upon the accuracy ratio. No other $F$ ratio reached the .01 level of significance.

*Description of effects.*—Table 2 presents the mean value of each dependent variable associated with each level of the independent variables. Figure 1 illustrates the function relating accuracy ratio to sample size. This ratio decreases rapidly at first, and then more slowly, as sample size increases.

Although the LLR is the more adequate measure of probability revision, Fig. 2 further illustrates the effect of sample size by a more direct measure—absolute probability change. The horizontal axis indicates the absolute difference between black and white marbles at the end of each trial. The vertical axis indicates the net absolute revision from the prior probability to the posterior probability at the end of

TABLE 1

ANALYSES OF VARIANCE ON ACCURACY RATIO AND LINEARITY

| Source | Accuracy Ratio | | Linearity | |
|---|---|---|---|---|
| | *df* | *F* | *df* | *F* |
| Sample Size (A) | 3/129 | 13.72*** | 1/43 | 4.04 |
| Blocks (B) | 7/301 | 1.32 | 7/301 | 2.57* |
| A × B | 21/903 | 1.78 | 7/301 | <1 |

* $p < .05$.
*** $p < .001$.

TABLE 2

Mean Value for Each Dependent Variable for Each Level of the Independent Variables

| | Accuracy Ratio | Linearity |
|---|---|---|
| Sample Size 1 | .470 | .915 |
| Sample Size 4 | .361 | .947 |
| Sample Size 12 | .253 | — |
| Sample Size 48 | .255 | — |
| Block 1 | .349 | .929 |
| Block 2 | .342 | .910 |
| Block 3 | .327 | .951 |
| Block 4 | .305 | .917 |
| Block 5 | .283 | .943 |
| Block 6 | .328 | .917 |
| Block 7 | .363 | .940 |
| Block 8 | .384 | .944 |

the trial. The solid curve represents the correct revision as calculated by means of Bayes' theorem. The three dotted lines are smooth curves intended to represent the data points associated with sample sizes of (*a*) 1, (*b*) 4, and (*c*) 12 and 48. The observation base varied across data points, and was relatively small when abscissa values were more extreme than 2 and 22. Figures 1 and 2 tell the same story. Subjective probability revision lagged behind Bayesian revision; amount of revision decreased from sample size 1 to 4, decreased from 4 to 12, and remained constant from 12 to 48.

*Means and individual differences.*—The first two rows of Table 3 present the mean and the standard error of the mean for the accuracy ratio and linearity. In order to evaluate the consistency of individual differences across situations, scores were compared among all levels of sample size. By collapsing over blocks, one score was obtained for each *S* in each sample size. Correlating each pair of sample sizes across *S*s resulted in a matrix of

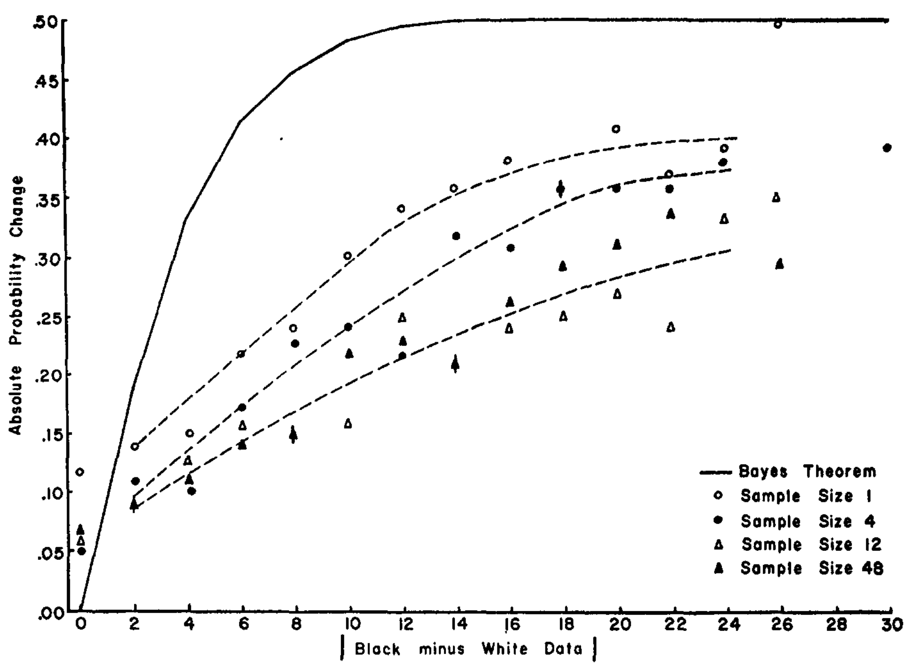


Fig. 2. Net absolute probability change as a function of the absolute difference between black and white data at the end of a trial. (A verticle line through a solid data point indicates coincidence of open and closed points.)

six correlations. The average of these six correlations is presented in the bottom row of Table 3, indicating the extent to which an *S*'s performance in one sample size could be predicted by his performance in another sample size. With $df = 42$, a correlation of .38 is significantly different ($p < .01$) from zero.

TABLE 3

MEAN, VARIABILITY, AND CONSISTENCY MEASURES ACROSS *S*s FOR THE ACCURACY RATIO AND LINEARITY

| | Accuracy Ratio | Linearity |
|---|---|---|
| $M$ | .334 | .933 |
| $\sigma_M$ | .039 | .009 |
| Consistency | .675 | .518 |

## Discussion

The primary purpose of the present experiment was to investigate the shape of the function which relates accuracy in the revision of subjective probabilities to the size of the sample of data upon which the revision is based. The graphs of Fig. 1 and 2 answer this question. Accuracy of subjective probability revision decreases with a negatively accelerated slope as sample size increases.

With reference to the information processing findings discussed earlier, these results provide another instance in which the accuracy of human information transmission decreases as the amount of information to be processed per observation increases. However, even though accuracy decreases, an increase in sample size results in more data processing per observation. Changing from sample size 48 to sample size 1 in the present experiment resulted only in a doubling of accuracy at the cost of 48 times as many observations. Thus, greater revision per observation results from increasing sample size, but at a cost in accuracy.

The lack of effect of any variable upon linearity means that *S*s in the present experiment were equally linear with Bayes' theorem, regardless of the level of sample size or block. Of course, caution is to be used in the interpretation of this meaning of linearity. Any variance in subjective probability revision which was not accounted for by Bayes' theorem may be due either to a systematic nonlinear relation or to error variance around a straight line. However, the relatively consistent magnitude of correlations reported in Table 2 (all mean correlations in the table fall between .91 and .95) implies an important degree of linearity between Bayes' theorem and the revision of subjective probabilities.

Finally, the experimental results provide further support for the previous finding that the revision of subjective probabilities lags behind the corresponding optimal revision as calculated via Bayes' theorem. Fairly high linearity seems to characterize this relation. Some *S*s lag much more than others; the degree of lag is significantly consistent from one sample size to another. Some *S*s are more Bayesian than others.

## REFERENCES

EDWARDS, W. Behavioral decision theory. *Ann. Rev. Psychol.*, 1961, **12**, 473–498.

EDWARDS, W., LINDMAN, H., & SAVAGE, L. J. Bayesian statistical inference for psychological research. *Psychol. Rev.*, 1963, **70**, 193–242.

EDWARDS, W., & PHILLIPS, L. D. Conservatism in probabilistic inference. Paper read at Midwestern Psychological Association meetings, Chicago, 1963.

EDWARDS, W., & PHILLIPS, L. D. Man as a transducer for probabilities in Bayesian command and control systems. In G. L. Bryan & M. W. Shelly (Eds.), *Human judgments and optimality.* New York: Wiley, 1964.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

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

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

(Received March 12, 1964)