Heuristic use or information integration in the estimation of subjective likelihood?

LAWRENCE M. WARD

University of British Columbia, Vancouver, B. C., Canada V6T 1W5

Use of the representativeness heuristic [described by Kahneman & Tversky (1972)] in estimating subjective likelihoods was investigated in three experiments. Procedural factors (number of problems per subject, information available to subjects) were found to account for differences in data from the two-urn situation collected by Kahneman and Tversky (1972) and Leon and Anderson (1974). Subjects use the representativeness heuristic in such problems but seem to modify it by using other sources of information when these are available or demanded by the situation.

This paper reports the results of three experiments on the estimation of subjective likelihoods in the two-urn problem. They represent a preliminary empirical investigation of a model of heuristic use to be detailed in future papers. The experiments were inspired by an apparent conflict between data of Kahneman and Tversky (1972) and Leon and Anderson (1974). Kahneman and Tversky (1972) found that relative likelihood judgments in the two-urn problem were affected only by the ratio of different elements (in their case, X- or 0-marked cards) in the sample and not on their difference or on population proportions as they are in the normative theory. They interpreted these data as an indication of use of the representativeness heuristic by the subjects in making their judgments. In this heuristic, samples are assumed to be representative of the population or process from which they are generated, and decisions concerning from which of the two populations a sample was drawn are based on which of the populations the sample is most similar to in salient attributes (in this case, proportion of element types).

Leon and Anderson (1974) presented evidence that both population proportion and sample size (which affects differences between numbers of different elements in the sample for a constant ratio of element types) affected their subjects' estimates of relative likelihood. They asserted that the conflict between their data and those of Kahneman and Tversky (1972) was due to subjects' misinterpretation of the instructions in the latter experiment, since each subject received only one problem and brief instructions. However, the two experiments were different in a number of other procedural aspects which could explain the conflict. In particular, the fact that Kahneman and Tversky's (1972) subjects made a judgment in only one problem each in a between-groups design, while each of Leon and Anderson's (1974) subjects made at least 392 such

judgments in a repeated measures design, means that the former lacked both a knowledge of the entire design and the requirement for internally consistent but different likelihood judgments across that design. It may be hypothesized that knowledge of the design may induce subjects to use the observed similarity between population and sample proportions in a molecular way, i.e., the more similar the sample to the more likely urn, the more confident will the subject be that he has selected the correct urn. In addition, the requirement for consistency of judgments over the entire design in a repeated measures experiment may induce the subjects to use other information relevant to the representativeness of the sample, namely the difference between the population proportions (the larger this difference, the more confident the subject in his choice), and the sample size (the larger, the more confident). The three experiments reported below present preliminary data relevant to these hypotheses. In addition, subjects in between-groups designs were asked to indicate what they had done and how, after making their judgments, to check on the hypothesis of misunderstanding of instructions and to provide further data on how the subjects were making their judgments.

METHOD

All subjects were students in first or second year psychology classes at the University of British Columbia. In all experiments, each problem seen by the subject consisted of a single sample of students from two high schools (Urban High or City High) who were said to be in a class at one of two universities, Oak or Maple. The subject's task was to choose the university in which it was most probable that the particular class represented by the sample was being held and then indicate on the scale "how confident you are that you are correct." Thus, in the language of the classical two-urn problem, the urns were the universities and the different colored balls were students from different high schools. In all experiments, the design was a 4 by 2 factorial. There were four levels of the more probable university (urn), in which the ratio of students from one high school to those from another high school (or red/white balls) was 60/40, 70/30, 80/20, and 90/10, while the less likely university contained students from the respective high schools in the ratio 40/60.
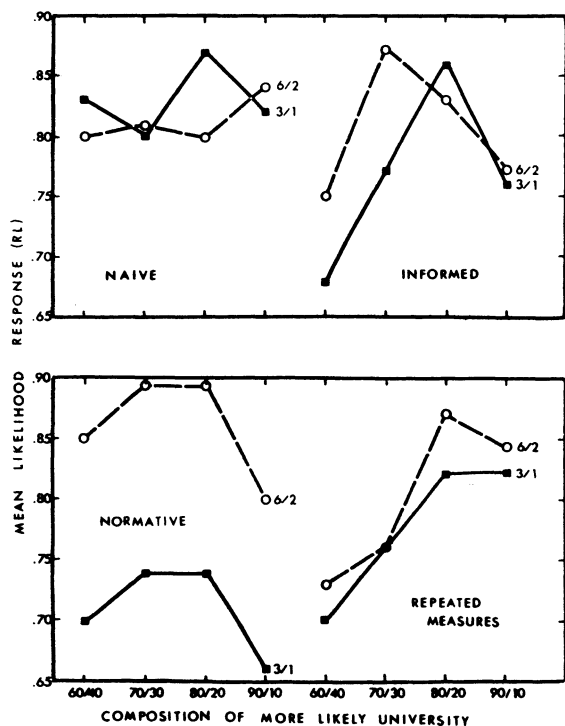
Requests for reprints should be sent to Lawrence M. Ward, Department of Psychology, University of British Columbia, Vancouver, B. C., Canada V6T 1W5.

Figure 1. Beside each curve is indicated the sample size for that set of problems. Naive (Experiment 1): N = 40, 5 per problem (data point); informed (Experiment 2): N = 40, 5 per problem; normative: relative likelihood calculated from Equation 1 and Equation 2; repeated measures (Experiment 3): N = 18.

Names of universities and high schools were counterbalanced in all experiments. There were also two different possible samples (classes) with number of students in the sample from the two high schools of 3 and 1, and 6 and 2, respectively. In answering a problem, subjects made a check mark in one of 20 spaces on a line labeled at the ends with names of the two universities. The right and left halves of the scale were used to indicate confidence (or likelihood) judgments of the more likely university, with the center of the scale indicating a judgment of "equally likely." Again, problems were counterbalanced for which side of the scale was labeled with the name of the most likely university.

Experiment 1

This experiment was done to replicate part of the Kahneman and Tversky (1972) finding for the present conditions. In this experiment, each of the 40 subjects (5 per cell of the design) saw one of the eight problems and made a single judgment of the most likely university and their confidence in that judgment. They did this after the experimenter had explained the problem and the task in general terms. Great emphasis was placed on the task requirements as explained above, and subjects were asked to write down what they had done and how they had done it after they had made their likelihood judgments.

Experiment 2

This experiment was exactly the same as Experiment 1 except that the 40 additional subjects were given additional knowledge about the problem space involved. After the problem and task had been explained, the entire factorial design of the experiment was written upon a blackboard so all could see it. Each subject was told that he would receive only one of the eight different problems but that all of the problems would be answered by someone.[1]

Experiment 3

This experiment is essentially a conceptual replication of the experiment of Leon and Anderson (1974). Thus, each of the 18 subjects received the entire design as problems in one sitting, after having done three practice problems. These subjects were also shown the entire design on a blackboard which remained visible throughout the experiment.

RESULTS AND DISCUSSION

Experiment 1

First, at least 38 of 40 subjects seemed to have correctly interpreted the instructions as determined from their written statements of what they did. The other two responses were ambiguous regarding understanding of the task. No subject said he judged the sample proportion, and it was quite clear from the statements that confidence judgments were not judgments of sample proportion. Second, all 40 subjects correctly chose the more likely university. The presumption is that these subjects understood the task.

Figure 1 shows the data of this experiment. ANOVA supports the impression that there are no significant effects in the data (all $Fs < 1.0$). It would appear that neither population proportion nor sample size affected subjects' likelihood judgments (sample proportion did not vary in this experiment). The choice of the more likely university was apparently based entirely on a judgment as to which of the two universities the sample was most similar with respect to proportion of students from the two high schools. This is supported by analysis of the written comments of the subjects in which 29 of 40 wrote down some form of the representativeness heuristic as a description of how they arrived at this answer.[2] Thus, these data are consistent with those of Kahneman and Tversky (1972) and support the notion that use of the representativeness heuristic indeed explains subjects' behavior in such situations.

Experiment 2

The results of this experiment are similar to those of Experiment 1 with respect to the subjects' written comments, i.e., the subjects seemed to correctly interpret the instructions (40 of 40) and to say they were using a judgment process very like the representativeness heuristic (30 of 40). However, as can be seen in Figure 1, population proportion, at least, seems to have had an effect on the likelihood judgments. This main effect is marginally significant in the ANOVA ($F = 2.67$, $df = 3,32$, $p = .07$), but the effect of sample size and the interaction are not significant ($F = 1.11$ and < 1.0, respectively). The test for the obvious quadratic trend in Figure 1b was significant ($F = 6.78$, $df = 3,32$, $p < .05$), and the quadratic trend accounts for 84.6% of the variance due to population proportion. The test for a linear trend was not significant ($F < 1.0$), and the linear trend accounts for only 15% of that variance.

It appears that although subjects said they were using the representativeness heuristic, an effect of population

proportion was produced by this procedure plus the introduction of knowledge of the entire design. This effect could be due to use of knowledge about the similarity of the sample proportion to the different population proportions. The sample ratio was always 3/1, but the population ratios of the more likely university were 1.5/1, 2.33/1, 4/1, or 9/1. The sample ratio is most similar to the two middle population proportions, and if confidence is affected by *degree* of similarity, then a quadratic trend is the result. However, sample size in this experiment is not related to similarity of proportions, and so no sample size effect should be seen if subjects are using representativeness. Thus, these data are consistent with the assertion that increased knowledge of the situation induced subjects to use differences in similarity of proportions in making their likelihood judgments.

It is also the case that the true posterior odds have the form of a quadratic trend when plotted as in Figure 1. Using the hypergeometric distribution function (Feller, 1968),

$$Q_K = \frac{\binom{n_1}{k} \binom{n - n_1}{r - k}}{\binom{n}{r}} \, , \qquad (1)$$

where Q_K is the probability of getting k red balls in a sample of r balls from an urn which contains n balls, n_1 of which are red, the probability of the sample being drawn from each "urn" for each problem of the design was calculated. The relative likelihood of the more likely urn is

$$RL = \frac{Q_{KML}}{Q_{KLL} + Q_{KML}} \, , \qquad (2)$$

where Q_{KML} is the probability of drawing the sample from the more likely urn and Q_{KLL} is that probability for the less likely urn. Figure 1 displays the relative likelihood calculated from Equation 2 for each problem. As can be seen, both population proportion and sample size affect the relative likelihood. However, even though a quadratic trend is seen in the effect of population proportions, so that subjects might be reflecting true posterior odds in their responses, it is quite different from that actually observed. This difference, coupled with the lack of a sample size effect and the written comments, make it unlikely that subjects were using any form of the hypergeometric to generate their likelihood judgments. Despite some assertions to the contrary, subjects seem to be doing something quite different than would be predicted by the normative probability theory.

Experiment 3

The results of Experiment 3, in which subjects solved all the problems in the design after three practice trials, are displayed in Figure 1. Repeated measures ANOVA of

these data revealed a significant effect of population proportions (F = 5.80, df = 3/51, p < .01) and of sample size (F = 4.94, df = 1/17, p < .05) but not of their interaction (F < 1.0). An analysis for linear trend was significant (F = 14.23, df = 3/51, p < .01), and this trend accounted for 82% of the variance due to population proportion. The quadratic trend was not significant (F < 1.0) and only accounted for 2% of the population proportion variance. Thus, these data replicate those of Leon and Anderson (1974), in that a linear effect of population proportion and a significant effect of sample size were both present. There seems to be no question of the use of the hypergeometric distribution function or the simple version of the representativeness heuristic, for neither predicts a linear effect of population proportion on likelihood estimates.

Leon and Anderson (1974) suggest that subjects may be using what they term a "ratio rule" (actually quite similar to Equation 2) in which the overt response (R) is the ratio of the strength of one response tendency (W_1 or W_2) divided by the sum of the strengths of the two response tendencies, thus:

$$R_1 = W_1 / (W_1 + W_2). \qquad (3)$$

Presumably the weights or strengths of these covert response tendencies are affected by population proportion and sample size in the observed fashion, because subjects perceive that the probabilities that the sample came from each of the two urns vary in that fashion. Leon and Anderson (1974) do not discuss how it is that a judgment of the subjective probabilities is made; apparently, they believe that subjects have arrived at a method which distorts the objective posterior odds in a different way from the representativeness heuristic.

This latter is not a necessary conclusion from these data, however. The repeated measures nature of the experiment may make consistency demands on subjects, to which they respond by using detailed situational information in making their judgments. In particular, since different problems should not be expected (by subjects) to induce the same degree of confidence, information such as sample size may be used as an objective basis for producing differences in confidence over a series of problems. Another type of detailed situational information is the difference between the population proportions. If subjects believed that the more extremely different the possible parent urns were, the more likely the sample was to have come from the urn whose proportion of red and white balls was most similar to that of the sample, the observed linear trend would be predicted. The slight downturn of the curve at the upper end indicates that difference between sample and population proportion may still be used; this also appeared in several of Leon and Anderson's (1974) graphs. In fact, some such idea may have also been operating in Experiment 2, though to a lesser extent, as

seen from one subject's comments: "60/40 and 40/60 are very close. If Maple's data were 90/10, I would know for sure that it was Maple University." Apparently, the knowledge of the design was salient enough in Experiment 2 to produce increased sensitivity to the similarity of sample to population proportion but not to allow differences between population proportions or sample size to significantly affect likelihood estimates. In Experiment 3, knowledge of the design may have been used even more heavily to resolve the increased demands of consistency in judgments of probabilities which were obviously meant by the experimenter to be different. Of course, in Experiment 1, neither the demands nor the knowledge were present.

The data reported above make some progress toward resolving the conflict between the data of integration theorists such as Leon and Anderson (1974), and those who assert that people use heuristics, such as Kahneman and Tversky (1972). It has been shown that the major variable determining which data are produced is procedural, with single problem solutions revealing relatively pure examples of heuristic use and repeated problem solutions giving rise to data which, on first examination, appear totally at odds with the single problem data. Two factors were argued to account for this variance: knowledge of the problem space and situational demands. Future papers will discuss these factors more fully and suggest a quantitative framework which includes a number of situational variations of the two-urn problem.

REFERENCES

Feller, W. *An introduction to probability theory and its applications* (Vol. I, 3rd ed.). New York: Wiley, 1968.

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3, 430-454.

Leon, M., & Anderson, N. H. A ratio-rule from integration theory applied to inference judgments. Journal of Experimental Psychology, 1974, 102, 27-36.

NOTES

1. In addition, these subjects were told that "the larger a sample is, the more reliably it can be expected to reflect the population from which it is drawn; and, the more different the two populations, the more reliably can a sample drawn from one of them be expected to differentiate between them." Although this was an attempt to provide some general knowledge about samples and populations to statistically naive subjects, it may have influenced their judgments. Since no sample size or population proportion difference effects appeared in Experiment 2, and since Experiment 3 replicated essential features of Leon and Anderson (1974), this is unlikely. However, care should be taken in generalizing from.these preliminary data.

2. A typical written comment was: ". . . the ratio in the class is 1:3 . . . (which) is closer to the ratio of City and Urban students at the Oak University, which is 2:3, compared to that at Maple University, which is 3:2." A comment was said to reflect use of the representativeness heuristic if it mentioned the similarity of the proportion (percentage ratio, etc.) of urban to city students in the sample to that in the more probable university. Many comments which did not meet this criterion nonetheless indicated that the choice of "more likely university" had been made because it had more students from Urban or City, just as did the sample (or vice versa). Several students mentioned the "laws of probability" as justification for their choices.