Title
Representativeness Reconsidered

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 4(0)

Author
Bar-Hillel, Maya

Publication Date
1982

Peer reviewed

Representativeness Reconsidered

Maya Bar-Hillel
Hebrew University

People's tendency to rely on representativeness (R) when making judgments of the probability (P) of various events can result in two major kinds of fallacies. Those that are inherent in the very substitution of P by R, and those that accompany the reliance on R as a side effect. By the first I mean fallacies that result from the fact that the logic of similarity differs from the logic of P. Thus, adding detail to the description of some event enriches it, and may thereby enhance its judged similarity to some criterion (Tversky, 1977). But this adding of detail also makes the event more specific, hence necessarily less probable. Kahneman & Tversky (K&T) showed, e.g., that Ss consider it more likely for Bjorn Borg to lose the first set in a tennis match and then win the entire game than merely to lose the first set, though the latter event includes the former. Fallacies that are side-effects of R are those that result when the outcome of judgment by R is not modified or integrated with other relevant considerations.

Early studies of P judgments linked certain common judgmental errors to R causally. In particular, people's tendency to neglect the effects of base rate, sample size and data reliability was seen as resulting directly from the fact that these factors do not affect R. Later studies cast some doubt on this link, for the following reasons.

a. These factors are sometimes ignored even in R-free tasks. Consider, e.g., the Suicide Problem (B-H, 1980)

> A study of suicide among young adults found that the rate of suicide is 3 times higher among singles that among marrieds in this age group. What would be the proportion of singles in a sample of suicide deaths of young adults?

The common response to this problem is 75%.

b. In R-free tasks, these factors sometimes exhibit a systematic effect on judgments of P. E.g., Ss judge it more likely that a large sample would provide an accurate estimate of the population mean than a small sample, ceteris paribus (B-H, 1979).

c. This effect is sometimes manifest even in the presence of R. In one version of the Tom W. prediction task, subjects were lead to expect either high or low predictive accuracy. While both groups gave essentially the same predictions, the low expected accuracy group expressed less confidence in their predictions. Thus, data reliability was not altogether ignored, though it wasn't properly combined with the R considerations either. Rather, it was translated into an expression of confidence in those considerations (K&T, 1973; B-H, 1981).

As a result of such findings, K&T recently moderated their formulation of the R heuristic, saying: "The magnitude of R biases and the impact of variables such as sample size, reliability and base rate depend on the nature of the problem, the characteristics of the design ...", etc.

It is illustrative to consider the role which normative staticstical theory assigns these three neglected factors. Take a prototypical statistical problem, that of reconstructing the parameters of some population on the basis of a sample of data. In the case of pure estimation, statistical theory teaches us that many "essential characteristics" of samples are unbiased estimators of corresponding population parameters. Hence estimation reflects R. So does the statistical notion of goodness-of-fit. When, on the other hand, alternative hypotheses compete, as in hypothesis testing, it is a notorious fact that classical statistical theory (but not Bayesian statistics) has no place for prior probability considerations. Yet these play the role that the base rate plays in prediction tasks such as Tom W.

As to sample size and data reliability, their role in both estimation and hypothesis testing lies in determining the width of a given confidence interval, but not the cenral value around which it is constructed. Analogously, these factors typically seem to effect Ss confidence in their predictions though not the predictions themselves.

In the Bayesian approach, P measures an internal state of uncertainty. Through the subjective filter all sources of uncertainty can be passed and integrated, and thus there is no call for higher order Ps. Psychologically speaking, however, people seem to distinguish between variants of uncertainty (K&T, 1982), and so may hold 2nd order P distributions (e.g., confidence) over 1st order P distributions (e.g., propensities) that are, subjectively, nonintegrable. It is compatible with points a., b. and c. above to hypothesize that R may be a heuristic for assessing 1st order Ps, and that factors which do not affect R may still influence 2nd order Ps. Whether they affect the ultimate P value may depend on the integrability of 1st and 2nd order considerations (B-H, 1982).

It should be apparent that the attempt at drawing analogies between the intuitive treatment of variables and the one formalized by normative theories is in no way an apologia for people's fallacies, which are genuine and worrisome. Cohen (1981) claimed that since the "presence of fallacies in reasoning is evaluated by referring to normative criteria which ultimately derive their credentials from a systematization of the intuitions that agree with them", people's deeply rooted statistical intuitions cannot, in principle, be fallacious. The point is moot, however, since clearly the output of defensible intuitions may itself be indefensible.

So far, I have tried to make the case that R is not just a fundamental feature of lay judgments under uncertainty, but of normative statistical theory as well. A world not governed by R might well be unthinkable. Just try to imagine a breakdown of the "law of averages". Physically uniform coins fall on Heads much more often than on Tails; well shuffled decks of cards yield Hearts more frequently than other suits; repeated independent measurements yield skewed, bimodal distributions; etc. Such a world, to rephrase Einstein, can only be the creation of a God who is not only subtle, but malicious as well.

Even though R may be essential to everyone's basic metaphysics, in particulars an ideal statistician, IS, may apply R more astutely to statistical inference problems than a layperson, L. We will now consider some such particulars, the idea being to show how refining R by simple, qualitative, statistical principles can lead to more appropriate solutions than R "in the raw".

i. Predicting sample features by R. Often the best prediction for an as yet unobserved sample is that it will resemble an already observed one, or the population that is its source. Clearly, however, it is too much to expect every feature of the past sample to be repeated in the future one. Yet, sophisticated respondents believed that, having obtained a just significant result in an experiment with 20 Ss, the chances of now obtaining a significant result on a new sample of 10 is 85% (K&T, 1971). Result significance, however, is a somewhat arbitrary notion. Since it depends on the sample size as well as the mean, expecting the sample mean to replicate (which is reasonable) should lead to more uncertainty about that mean's significance, since sample size was halved.

Other respondents expected a sample (n=50) from a population with mean=100 to have such a mean as well. They held on to that expectation even when told that the first observation was 150. It is impossible for both the unknown portion of the sample (n=49) to repeat the population mean, and for the sample as a whole to do so (K&T, 1972).

In some school, program A consists of 65% boys, while program B of 45% boys. Ss expected classes belonging to Program A to resemble the program's composition more than the other program's. The similarity of some class' proportion of males to 65% versus 45% should be evaluated in terms of standard deviations. Ss seemed to evaluate it in terms of which sex was the majority, thus expecting a class of 53% boys to belong to Program A.

ii. Features of Gestalts versus features of data points. The statistical properties of samples are completely determined by the individual data points of which they are comprised. Features that accrue to the sample as a whole, but not to its constituents (e.g., its mean) are significant insofar as the individual data points are unknown or discarded. Thus, a sample whose mean is near the population mean is more likely, ceteris paribus, than one with a more deviant mean. But this order may be upturned when the specific data points are given. L seems to find it difficult to ignore the emergent properties of samples as Gestalts, even when they are completely specified. IS, on the other hand, would ignore thses emergent properties when specific data points are available. Hence, unlike L, IS, believing that Heads and Tails are equally likely outcomes for the toss of a fair coin, would consider any fully specified sequence of fixed length comprised of equiprobable outcomes to be equiprobable. Similarly, IS would judge the P of a sample of fixed size drawn from a normal distribution to depend on the magnitude of the standardized deviation between the sample points and the population mean, rather than on its directionality. (L's errors are documented in K&T, 1972, B-H, 1980b).

Clearly, the Bjorn Borg example at the beginning of this paper can also be understood in terms of emergent properties. The P value of wholes is derivable from their parts. The R value may not be.

Summary. The reliance on R as a judgmental heuristic is frequntly justifiable, and seldom avoidable. The modification of R considerations by other considerations of relevance, and the refinement of the domain of R, its metric, etc. is a goal to be sought. Inasmuch as the various judgments of R embody much of our substantive knowledge regarding the issue being judged, R can not be eliminated from the probabilistic reasoning process, but the different logic of R and P poses obstacles that must be watched out for.

References

B-H, 1979. The role of sample size in sample evaluation. OBHP.

B-H, 1980. The base rate fallacy in probability judgments. Acta Psychologica.

B-H, 1980b. What features make samples appear representative? JEP:HP&P.

B-H, 1981. Representativeness reconsidered. unpublished manuscript. DR report, Eugene.

B-H, 1982. Ideal evidence, relevance, and second order probabilities. Erkenntnis.

Cohen, 1981. Can human irrationality be experimentally demonstrated? B&BS.

K&T, 1971. Law of small numbers.

K&T, 1972. Subjective probability: A judgment of representativeness.

K&T, 1973. On the psychology of prediction.

K&T, 1982. Variants of uncertainty.

T&K, 1982. Judgmentsof and by representativeness.

(all these can be found in Kahneman, Slovic & Tversky, Judgment under uncertainty, CUP, 1982)

Tversky, 1977, Features of similarity. Ψ Rev.