J. Hintikka & P. Suppes (Eds.), Aspects of Inductive Logic.
Amsterdam: North-Holland, 1966, pp. 198-207.

# A BAYESIAN APPROACH TO THE PARADOXES OF CONFIRMATION*

PATRICK SUPPES
*Stanford University, Stanford, Calif.*

**1. Introduction.** What I have to say about the paradoxes of confirmation from a Bayesian standpoint is rather simple. The ideas have been implicitly expressed several times, probably first by Hosiasson-Lindenbaum [1940]. Perhaps the only virtue of the present paper is to make the Bayesian ideas very explicit. The remarks in the last section on the different probabilistic forms of causal and noncausal laws are very likely the most original aspect of the analysis.

The paradoxes arise from two "facts". First the sentence

$$(\forall x)(Ax \rightarrow Bx) \tag{1}$$

is logically equivalent to its contrapositive:

$$(\forall x)(\bar{B}x \rightarrow \bar{A}x), \tag{2}$$

where "$\bar{\ }$" is the symbol of negation (and later of set complementation). Second, the singular sentence

$$Aa \,\&\, Ba \tag{3}$$

seems to confirm (1) in a way that the singular sentence

$$\bar{A}a \,\&\, \bar{B}a \tag{4}$$

does not, but with respect to (2) the roles of (3) and (4) are reversed, even though (1) and (2) are logically equivalent.

**2. Bayesian approach.** On a Bayesian approach, we first look at the four classes and assign each a prior probability in the universe of objects – exactly how this universe is to be characterized I leave open for the moment.

* I am indebted to Ernest W. Adams and Paul Holland for several helpful comments on an earlier draft of this paper. The writing of this paper has been partly supported by the Carnegie Corporation of New York.

Using the familiar notation '$\{x : Ax\}$' for describing the set of objects $x$ that have property $A$, we then have in terms of four mutually exclusive and exhaustive classes

$$P(\{x: Ax \,\&\, Bx\}) = p_1$$
$$P(\{x: Ax \,\&\, \bar{B}x\}) = p_2$$
$$P(\{x: \bar{A}x \,\&\, Bx\}) = p_3$$
$$P(\{x: \bar{A}x \,\&\, \bar{B}x\}) = p_4$$

and

$$\sum p_i = 1 .$$

Also for simplicity I assume throughout that $p_i \neq 0$, for $i = 1, 2, 3, 4$. If we take the familiar example and let '$Ax$' be '$x$ is a raven' and '$Bx$' be '$x$ is black', then $p_4$ should be much larger than $p_1$, $p_2$ and $p_3$ for any very broadly construed universe.

The central question is why we are right in our intuitive assumption that we should look at randomly selected ravens and not randomly selected non-black things in testing the generalization that all ravens are black.

We may consider the general case, representing classes by '$A$' and '$B$' in the obvious way: $A = \{x : Ax\}$, etc. First of all, we note that

$$P(A) = p_1 + p_2 , \tag{5}$$

$$P(B) = p_1 + p_3 , \tag{6}$$

and thus in terms of conditional probability

$$P(B \mid A) = \frac{p_1}{p_1 + p_2} \tag{7}$$

$$P(\bar{A} \mid \bar{B}) = \frac{p_4}{p_2 + p_4} . \tag{8}$$

Now we want to justify the sampling rule that we look at $A$'s rather than non-$B$'s if $P(B|A) < P(\bar{A}|\bar{B})$, i.e., if

$$\frac{p_1}{p_1 + p_2} < \frac{p_4}{p_2 + p_4} . \tag{9}$$

It is an immediate arithmetical truth that (9) is true if and only if

$$p_1 < p_4 . \tag{10}$$

I believe the argument is straightforward for holding that the decision to look at $A$'s rather than non-$B$'s hinges upon the simple inequality (10).

In sampling objects to confirm or disconfirm the general law '$(\forall x)(Ax \rightarrow Bx)$', we want to *test* the law. This, I take it, means that we want to sample items

with a higher prior probability of disconfirming the law. This point is made clear by noting

$$P(\bar{B} \mid A) = \frac{p_2}{p_1 + p_2} \tag{11}$$

and

$$P(A \mid \bar{B}) = \frac{p_2}{p_2 + p_4}, \tag{12}$$

and selection of an $A$ has a higher prior probability of disconfirming the law than does selection of a non-$B$ just when

$$P(A \mid \bar{B}) < P(\bar{B} \mid A),$$

that is, just when once again

$$p_1 < p_4.$$

It should be made clear that the adoption of a rational rule for what to observe or sample does not follow from the prior probabilities alone. Some other ingredient must be added, but the rule that tells us to select an $A$ rather than a non-$B$ when $p_1 < p_4$ follows from any of a number of more general principles. We may, for example, derive it from a general principle of minimizing costs or effort, if the general principle is supplemented with some reasonable and relatively innocent additional assumptions to make the argument watertight – for example, it is approximately as expensive to sample non-$B$'s as $A$'s.

For those who dislike decision-theoretic principles, we may simply begin with the principle that we should always try to take observations so as to maximize the probability of falsifying the hypothesis in question. This principle, which is essentially just a restatement of the earlier remark about testing, should be derivable from more general principles that do not invoke the concepts of decision theory, but it would be a digression if the search for such a derivation were pursued here. It is worth remarking that as far as I know, the question of what observations should be made, or what evidence collected next, has scarcely been discussed in the framework of inductive logic.

As an illustrative numerical example to reinforce these general remarks about testing, let

$$\begin{aligned} P(A \,\&\, B) &= p_1 = 10^{-6} \\ P(A \,\&\, \bar{B}) &= p_2 = 10^{-7} \\ P(\bar{A} \,\&\, B) &= p_3 = 10^{-4} \\ P(\bar{A} \,\&\, \bar{B}) &= p_4 = 1 - 10^{-4} - 10^{-6} - 10^{-7}. \end{aligned}$$

Then

$$P(B \mid A) = \frac{1}{1.1} = 0.9090909 \tag{13}$$

$$P(\bar{A} \mid \bar{B}) = \frac{0.9998989}{0.999899} = 0.999999899\,. \tag{14}$$

The probability of (14) is so close to 1 that it seems ridiculous to try to change it by additional observation. This is far less true for the probability of (13). Consequently we sample ravens rather than nonblack things.

On the other hand, it would be a mistake to think that $p_4$ will always turn out to be close to 1, and therefore that we should always look at $A$'s rather than non-$B$'s. Consider the following example. Suppose that in a certain election district we want to test the generalization that all voters in this district are literate. The universe $X$ of objects we define to be the adult population of the district. Let $V$ be the subset of voters and $L$ be the subset of literate people of $X$. A quite reasonable a priori probability for each of the four classes could be something like the following:

$$\begin{aligned} p_1 &= P(V \,\&\, L) = 0.75 \\ p_2 &= P(V \,\&\, \bar{L}) = 0.05 \\ p_3 &= P(\bar{V} \,\&\, L) = 0.15 \\ p_4 &= P(\bar{V} \,\&\, \bar{L}) = 0.05\,, \end{aligned}$$

and because $p_4 < p_1$, the Bayesian recommendation would be, ceteris paribus, to sample non-$L$'s.

In restricting the universe of objects to the adult population of the district, I seem to have made a rather radical departure from the implicit choice of a universe in the raven case. But a corresponding restriction of the universe in the raven case to the bird population would not have changed the inequality $p_1 < p_4$.

All the same, a serious problem does arise in specifying the universe of objects. Those who discuss the paradoxes of confirmation in the context of confirmation theory generally admit any sort of observation under the $\bar{A}$ & $\bar{B}$-heading, for example, in the raven case, white shoes, red chairs and perhaps even nonblack thoughts. From the standpoint of scientific practice this wholesale inclusion of obviously irrelevant objects is absurd, but naturally what is absurd from a practical standpoint is not necessarily so in a discussion of conceptual issues. The Bayesian viewpoint seems to me to provide a clear and easily understandable argument in defense of the standard scientific practice of limiting the universe of objects. When the universe is not so

limited, the probability $p_4$ becomes extremely close to 1. But when irrelevant objects like nonblack thoughts are excluded from sampling by the probability argument given above, the sample space represents a radically reduced population that does not have $p_4$ absurdly close to 1.

There are several features of the Bayesian viewpoint that stand in sharp contrast to the position of Carnap and others who work mainly in the framework of inductive logic rather than mathematical statistics. On the one hand, the logicians are scandalized at the vague and subjective character of the prior probabilities used by Bayesians. On the other hand, Bayesians are scandalized at the artificiality and simplistic character of nearly all examples considered by the logicians. I don't pretend to be able to offer arguments that will resolve a conflict of this depth, but I would like to say some things that have perhaps already been said elsewhere but that can perhaps be said in a somewhat new way to defend the Bayesian viewpoint. I would claim to do this without too much bias because in two other articles in this volume I am concerned to criticize what I take to be fundamental limitations of the Bayesian approach to induction and problems of rational information processing.

First, the Bayesian is quick to remark that in most systems of inductive logic, the probabilities $p_1$ and $p_4$ in the raven example are assigned the same, or nearly the same value, if the example is analyzed from scratch in terms of the classes or properties $A$ and $B$. (For the Bayesian this *pure* a priori assignment reflects wanton waste of evidence already available.) Now the inductive logician may reply that if he were seriously pursuing this example, he would begin much further back and introduce a number of fundamental elementary predicates into his language, and use them to express explicitly the known evidence about ravens, $A$'s and $B$'s, or what have you.

And this reply leads to the second Bayesian retort. The inductive logicians, to the Bayesian at least, are the heirs to an intellectually distinguished but misguided tradition of logical atomism that begins at least with Hume. The constraints imposed by the atomism of Wittgenstein's *Tractatus* are weak compared to the assumptions of statistical independence built into the Carnapian measures imposed on state or structure descriptions. It is the virtue of Carnap to have pursued the atomistic tradition to its more complete probabilistic version, for what the approach comes to in terms of specific questions of confirmation or observation can be settled in a way that is not possible for the relatively vague and non-categorical doctrines of Hume's *Treatise* or Wittgenstein's *Tractatus*. But for Bayesians it is an approach that is bound to fail for fundamental reasons. First, it is impossible to express in explicit

form all the evidence relevant to even our simplest beliefs. There is no canonical set of elementary propositions to be approached as an ideal for expressing exactly what evidence supports a given belief, whether it be a belief about ravens, gods, electrons or patches of red. Arguments in support of this last assertion are numerous, and it is worth examining the most important ones because of the fundamental nature of this particular issue for any theory of induction or rational behavior.

The simple memory of a computer does provide an example of evidence organized in terms of canonical propositions, but this is because all beliefs or propositions for the computer are categorical. Every issue or belief is an utterly black or white affair. There is no place for partial beliefs, tentative evidence or vague but perceptive hunches. And the lack of these characteristics is perhaps the most salient feature of contemporary computers. At the present time partial beliefs or intuitive hunches can probably be analyzed more thoroughly in terms of probability distributions than in any other fashion. It is not unreasonable to say that the propensity for generating Bayesian distributions is the human facility computers most sorely need.

However, there is no need to bring computers into the argument. Analysis of ordinary beliefs, with appropriate contrast of Bayesian and Carnapian views, will stand on its own feet. If I ask a person who follows politics with any serious interest, what the Republican chances are of winning the 1968 presidential election, he would probably not hesitate to give some sort of qualitative answer, like "Not so good", "Unlikely", or "Very small in my judgment". And if I go on to ask him for the basis of his estimate of the chances, he will probably go on to offer a packet of heterogeneous facts and the reasons for thinking some of them are particularly significant. All this is very standard in political conversation and also very Bayesian. Most political beliefs are not quite pinned down; the evidence is assembled higgledy-piggledy from all kinds of sources that vary widely in reliability and relevance. Now it is not uncommon for an inductive logician to be willing to admit all this, but he may go on to say that while the consideration of hunches or badly formulated beliefs may be an essential part of the discovery, it is not essential in the validation and assessment of hypotheses. The Bayesian is not content with this thin bone of discovery. He will go on to reply that the vague and subjective prior distribution is of importance primarily in summarizing all the information about the experiment or proposed test which lies outside the narrow framework of the experiment itself, but which is still relevant in varying degrees. The assumption of a prior distribution is a systematic way of summarizing a great deal of heterogeneous information.

And here another point arises. The Bayesian is more modest than the inductive logician in what he hopes to express by means of a prior distribution. It is of fundamental importance to any deep appreciation of the Bayesian viewpoint to realize that the particular form of the prior distribution expressing beliefs held before the experiment is conducted is not a crucial matter. If a moderate number of observations is taken in the experiment, the conclusions drawn will be relatively robust, that is, relatively indifferent to moderate variations in the prior distribution; and, the more the number of systematic observations the more robust the conclusion. There is a very general theorem that can be stated here, but I shall not digress to formulate it precisely. It is to the effect that given any two prior distributions drawn from a large class of possible distributions, there is, for a broad class of experiments, a sufficiently large number of observations to bring the two posterior distributions as close together as is desired. For the Bayesian, concerned as he is to deal with the real world of ordinary and scientific experience, the existence of a systematic method for reaching agreement is important. To him it is hopeless to strive for an atomistic expression of the total relevant evidence in terms of elementary observation sentences. The well-designed experiment is one that will swamp divergent prior distributions with the clarity and sharpness of its results, and thereby render insignificant the diversity of prior opinion.

The Bayesian does not believe that we can find ways to express these diverse prior opinions in logically tight, explicit form. The task of the theory of rationality, for the Bayesian, is to understand how to conceive and design experiments that will eliminate or reduce diversity of opinion about serious questions, and part of the task of this theory is being clear about puzzling matters like the paradoxes of confirmation. I hope that these more general remarks will have defined more sharply the framework within which I have proposed to resolve the paradoxes.

**3. Causal versus noncausal laws.** I have some additional specific remarks to make about the paradoxes. From a standard statistical viewpoint the analysis already given goes only part of the way. A well-defined sample space for a well-defined experiment was not constructed for either the raven or voting example. What bothers me is that the construction of the appropriate sample space does not seem at all natural. In trying to determine why this is so, what has struck me most is the complete artificiality of the problem. The experimental literature of biology, psychology and to some extent even physics is full of meaningful experiments testing meaningful hypotheses in a statistical

fashion, but essentially none of those hypotheses is of the form "All ravens are black". The main point seems to be that no one applies systematic statistical procedures for making inductive inferences or testing hypotheses when the hypothesis in question asserts a nonprobabilistic implication about a discrete classification.

In the case of physics particularly, statistical procedures are used to test deterministic hypotheses, but the hypotheses are about continuous quantities, and statistical questions enter mainly in discussing errors of measurement. The reason for this attitude toward deterministic hypotheses seems clear. The assessment of evidence for or against such a hypothesis is a quite simple-minded affair. A single observation will falsify the hypothesis; all positive instances will confirm it. We have no serious or systematic statistical problem of assessing evidence.

LaPlace, I am sure, would have considered the raven sort of problem rather silly, because he thought the apparatus of probability theory was to be applied to the determination of the complex causes of phenomena when no simple or deterministic scheme would work in practice. More can be said on this point, but let us see how the main thrust of these remarks bears on the paradoxes of confirmation moved to a more realistic setting. The paradigm-sort of hypothesis is now:

Smoking tends to cause cancer.

Put in terms of classes $A$ and $B$, we move from

For all $x$, if $x$ is $A$ then $x$ is $B$

to:

$$P(B|A) > P(B|\bar{A}) \tag{15}$$

or:

$$P(\text{cancer}\,|\,\text{smoking}) > P(\text{cancer}\,|\,\text{nonsmoking}). \tag{16}$$

The first thing to note is that the obvious form of the paradox of confirmation disappears for in general

$$P(B|A) \neq P(\bar{A}|\bar{B}),$$

i.e., the direct analogue of contraposition is not valid in terms of conditional probability. On the other hand, it reappears in another form, which is innocuous in many applications. We need the usual $2 \times 2$ contingency table to bring out the point. The distribution of the population (or sample) is shown by the numbers $n_{ij}$.

$$\begin{array}{c|cc} & B & \bar{B} \\ \hline A & n_{11} & n_{12} \\ \bar{A} & n_{21} & n_{22} \end{array} \tag{17}$$

We may use this table to show that (15) holds if and only if

$$P(\bar{A}|\bar{B}) > P(\bar{A}|B), \tag{18}$$

and (18) is a sort of probabilistic contrapositive of (15). Using (17), we have

$$\begin{aligned}
P(B|A) &> P(B|\bar{A}) && \text{if and only if}\\
\frac{n_{11}}{n_{11}+n_{12}} &> \frac{n_{21}}{n_{21}+n_{22}} && \text{if and only if}\\
n_{11}n_{22} &> n_{12}n_{21} && \text{if and only if}\\
n_{11}n_{22}+n_{21}n_{22} &> n_{12}n_{21}+n_{21}n_{22} && \text{if and only if}\\
\frac{n_{22}}{n_{12}+n_{22}} &> \frac{n_{21}}{n_{11}+n_{21}} && \text{if and only if}\\
P(\bar{A}|\bar{B}) &> P(\bar{A}|B), &&
\end{aligned}$$

which establishes the desired equivalence.
In terms of smoking and cancer, we have:

$$P(\text{cancer}|\text{smoking}) > P(\text{cancer}|\text{nonsmoking}) \text{ if and only if}$$

$$P(\text{nonsmoking}|\text{noncancer}) > P(\text{nonsmoking}|\text{cancer}),$$

and not only does this seem reasonable, but it also seems reasonable to sample either the causes (smoking) or the effects (cancer) and their absences in establishing a probabilistic causal law. We may sample by looking at smokers and nonsmokers, or by looking at persons with cancer and those without cancer. (For detailed design of an experiment, the question of precisely what class seems a priori most appropriate to sample or, more realistically, in what proportions classes of individuals should be sampled, would follow the same line of analysis pursued earlier in discussing the raven example, and will not be considered in detail again.)

However, a subtle point has been illegitimately smuggled in, and the situation changes when we consider something closer to the raven case, i.e., a noncausal law.

We may entertain the noncausal probabilistic law:

$$\text{Most ravens are black.} \tag{19}$$

The natural probability expression of this hypothesis is *not* the analogue of (15):

$$P(B|R) > P(B|\bar{R}), \tag{20}$$

but rather

$$P(B|R) > P(\bar{B}|R), \tag{21}$$

and without further assumption the apparent "contrapositive" probability analogue of (21) is not necessarily equivalent to it. To be explicit, (21) is not necessarily equivalent to

$$P(\bar{R}|\bar{B}) > P(R|\bar{B}), \tag{22}$$

as may be seen from using table (17) as before, and with this observation, the paradoxes of confirmation vanish for (19). (It may be argued that the bare inequality of (21) does not reflect the exact meaning of *most* and that a stronger form of inequality should be used, but meeting this criticism is not crucial for the present discussion.)

As far as I know, the relevance for the paradoxes of confirmation of the sharp distinction between causal and noncausal laws, particularly the relevance of the different probabilistic forms of such laws, has not been previously noticed. It should be apparent that the kind of causal law pertinent to this discussion is probabilistic rather than deterministic in character, and is of the sort ordinarily tested in biological, medical and psychological experiments and reported in contingency-table data.

A certain lack of clarity in the distinction between causal and noncausal laws is also to be found in the terminology used in the statistical literature. Statisticians have developed measures of *association* for contingency-table data and the probabilistic causal laws tested by the tables. It would seem more natural to reserve the term *association* for testing the noncausal laws, but such tests are not ordinarily discussed in the same detailed fashion, undoubtedly because of the greater importance of causal laws from both a practical and conceptual standpoint.

I do not mean to suggest that inequality (15) offers a very profound analysis of the probabilistic notion of cause. My limited objective in this paper has been to point out the conceptually sharp distinction between causal and noncausal laws when they are expressed in a probabilistic form. The ideas used here go no deeper than what I would call the level of naive causes. The identification of genuine causes, which to me seems necessarily relative to a particular conceptual scheme, requires a more elaborate probabilistic structure than I have introduced here. But the introduction of additional structure would not change what I have said about the non-existence of the paradoxes of confirmation for either causal or noncausal laws of a probabilistic sort.

## References

HOSIASSON-LINDENBAUM, 1940, *On confirmation*, J. of Symbolic Logic, vol. 5, pp. 133–148