Brit. J. Phil. Sci. **45** (1994), 467–476 *Printed in Great Britain*

Bayesian Confirmation: Paradise Regained*

R. D. ROSENKRANTZ

1 AN ANTINOMY OF CONFIRMATION

Despite the counterexamples with which they have been liberally peppered, many philosophers still cling to both *Nicod's criterion* (that AB's confirm and A nonB's disconfirm 'All A are B') and *Hempel's special consequence condition* (that whatever confirms a hypothesis confirms any consequence thereof). In the literature, 'consequence' is broadly construed. For example, Kepler's laws are viewed as consequences of Newton's. Also, from Hempel's 'Studies', we know that nonA nonB's are 'positive instances' on a par with AB's, since the consequence condition entails the equivalence condition (that whatever confirms *h* confirms any equivalent of *h*).

Now Goldbach's conjecture that all even numbers beyond 4 are sums of two odd primes has as a consequence that every odd number is a sum of no more than three odd primes ('little Goldbach'). For N-3 is even if N is odd. Finding an odd number not representable as a sum of fewer than five odd primes (*i.e.* a number not even and not a sum of two primes) thus confirms the Goldbach conjecture, and so also confirms its consequence, the 'little' Goldbach conjecture. We are surprised then to find the latter hypothesis confirmed by one of its negative instances. The upshot is that the consequence condition directly contradicts the Nicod criterion.

Hempel ([1945], note 21) remarked that Nicod's criterion can fail for *relational* hypotheses—*e.g.* $(x)(y)(\sim(Rxy \cdot Ryx) \supset (Rxy \cdot \sim Ryx))$—but maintained that 'if we restrict ourselves to universal conditional hypotheses in one variable . . . then it seems perfectly reasonable to qualify an object as

* I thank the referee for a number of very helpful comments.

confirming such a hypothesis if it satisfies both its antecedent and its consequent'. Here, then, is an example to show that Nicod's criterion can fail even if we do restrict consideration to universal conditionals in one variable.

2 BLOCKING CONTRAPOSABILITY

One common reaction to the paradoxical confirmation of the raven hypothesis by sightings of white shoes is to reject the usual formulation of general laws as universalized conditionals (Aronson [1989]). The point is to block contraposibility, so that 'All nonB are nonA' no longer counts as an equivalent of 'All A are B'. Then odd numbers not sums of fewer than five primes need no longer confirm Goldbach's conjecture, and Nicod's criterion need no longer conflict with the consequence condition. This move is strong medicine—too strong.

First, Goldbach's conjecture *is* a universalized conditional as mathematicians understand it. Aronson's proposal can be saved only by insisting that a different logic of confirmation applies to arithmetical conjectures, like Goldbach's—an onerous thesis—or else by viewing standard universalized conditionals as limiting cases of the 'laws' he wants to formalize.

The more serious objection, though, is that observations of nonB nonAs may well be genuinely confirmatory. Certainly examining a binary star whose members do not attract inversely as the square and finding that magnetic forces are also present would confirm Newton's gravitation law, however construed. On the other hand, if 'All A are B' is a generalized conditional, finding of a nonB that it is nonA will confirm *h* simply by virtue of being a consequence of *h*. Even where we sample a population *at large*, turning up a nonA nonB affords *h stronger* confirmation that turning up an AB provided the considered alternative is that the two traits are unassociated, and that nonA nonBs are less common in the population than ABs (see Suppes [1966]). Finding a rare infected subject who was not vaccinated in an experimental population provides stronger evidence that the vaccine in 100 per cent effective than finding a vaccinated subject who is uninfected.

In a paradigmatic psychological experiment, *Wason's selection task*, cards bear either of the numerals '4' or '7' on one side, and one of the letters 'A' or 'B' on the other side. The subjects are presented with four cards bearing these four symbols on the upturned face and are asked which of the cards must be turned over to decide the truth of the general statement: if 'A' appears on one side then '4' appears on the other side. While most subjects recognize the need to turn over the card bearing 'A', fewer than 10 per cent realize that the card bearing '7' must also be turned over. (The generalization is false if a card bearing '7' on one side bears 'A' on the other side.) On the other hand, in real-life cases as in the defunct British postal rule, 'If a letter is sealed it requires a 5*d*. stamp', subjects do a lot better! At any rate, these examples show that it may be just as essential to check that nonB's are nonA's as that A's are B's.

The source of the raven paradox is, not contraposability, but the move from 'a white shoe is a nonblack nonraven' to 'white shoes confirm . . .' Confirmation by nonblack nonravens does not commit us to confirmation by white shoes or yellow canaries. Relevant conditional probabilities are altered when the observation report is rendered more specific. That is the point of my hatcheck example (Rosenkrantz [1982]). It exploits a quirk of the classic problem of *matches*, namely, that the probability of no matches when N hats are assorted at random among their owners is lower for N-2 than for N when N is odd.

Now the intelligence that two of the men did not receive their own hat surely confirms the hypothesis *h* that no man receives his own. It will raise the probability even more if we are told, further, that neither received the other's hat (which leaves their hats on the heads of two other men). Suppose we are told, on the other hand, that each *did* receive the other's hat? That leaves the remaining N-2 hats assorted *at random* among their owners. Hence, the correct updated probability is just that of no matches with N-2 in place of N. Thus, for odd N, the more specific information has, not merely diminished, but destroyed confirmation.

On Bayesian grounds, there is no reason to suppose that such varnishing of the news will leave confirmation intact—a point first made in this journal by I. J. Good twenty-five years ago (Good [1967]). Since that supposition is precisely what generates the familiar paradoxes of Goodman and Hempel, those paradoxes are stopped cold in Bayesian confirmation theory without special pleading of any sort and without deep changes in our logic or in our usual way of representing general laws. What the antinomy offered in our opening paragraph shows is that such filling out of an observation reporting a nonA nonB can not merely turn confirmation to disconfirmation, but can produce a negative instance.

3 WHAT SURVIVES

As Aronson's paper also illustrates, there are many who find the Bayesian resolution of the paradoxes sketched here, with its rejection of the Nicod and consequence conditions, more paradoxical still. To complete our account, it must be shown what remains intact of these canons.

We have already seen that 'instancehood' survives in the weaker form of consequencehood. Examining known A's (resp., nonB's) and finding them B's (resp. nonA's) confirms 'All A are B' because this finding is a consequence of that hypothesis. And the more improbable a consequence relative to residual background knowledge, $\sim hb$, the stronger the confirmation. Even Popper would agree (while refusing to register confirmation as a plausibility increment), and, in fact, our tendency to equate instances with consequences harmonizes with the Popperian canon that hypotheses are genuinely 'corroborated' only when genuinely 'tested'.

For Bayesians, a *severe test* of h is one that has a low probability on $\sim hb$ of issuing in an outcome e that conforms to h. Watkins [1964] views the passing of such severe tests (thus characterized) as the *sine qua non* of *Popperian* 'corroboration'. This is the common ground between Popperian corroboration and Bayesian confirmation. What seems to distinguish the former from the latter are two additional demands: first, that a genuine test of h expose h to risk of falsification, and, second, that a genuine test represent a 'sincere effort' to refute h (Popper [1959], pp. 401–2, 418). In his disparagement of Freudians and Marxists, Popper has suggested that confirmations of a theory are easily obtained if sought (and signify correspondingly little).

Both these additional demands are *too strong*. In assorting 40 balls at random among 20 cells, the placement of the first 20 balls cannot possibly refute the hypothesis that every cell is occupied, yet it may confirm, even verify, that hypothesis. What does hold (trivially) in Bayesian terms is that one cannot confirm a hypothesis without exposing it to risk of *disconfirmation*.

Again, whether an experimenter 'seeks confirmation' or refutation is irrelevant to the import of what he observes. Confirmation is, to that extent, a *logical* relation. Nor is this a mere quibble. Those who looked for observable effects of the earth's rotation or orbital motion—as Newton, for example, who saw that a stone dropped from a tower would suffer a measurable *eastward*, not westward, deflection—can be described as 'seeking confirmation' of the earth's motion. The inability of rival hypotheses to explain very precise effects of this sort is what makes them so strongly confirmatory, despite being attempted 'proofs' of a hypothesis. In delimiting the theoretical alternatives, it is far more useful to confirm conjectures or promising proto theories than to remove them through apparently decisive falsification. Nor do putative refutations or 'anomalies' seem to me as important a source of radical theory change as the primordial urge to unify and simplify.

4 IRRELEVANT CONJUNCTION

On H–D accounts, h is confirmed by a verified prediction, e, but e is equally a prediction of hk, where the 'tacked on' k may be a quite extraneous hypothesis. That seems a telling objection to any H–D account, especially one wedded to the consequence condition. For then any e confirms any k. There are those who think that this sin of 'irrelevant conjunction' vitiates Bayesian confirmation theory as well. In fact, the problem is to steer a safe passage between the Scylla of consequentialism and the Charybdis of irrelevant conjunction.

Since we do not accord general validity to the consequence condition, the more disastrous implication that anything confirms anything is easily averted.

At the same time, I hope you will agree that the two extreme positions on this issue are equally unpalatable, (i) that a consequence *e* of *h* confirms *hk* not at all, and (ii) that *e* confirms *hk* just as strongly as it confirms *h* alone. For when *h* entails *k*, so that $hk = h$, (ii) should hold, and when *h* excludes *k*, (i) should hold. In general, intuition expects intermediate degrees of confirmation that depend on the degree of compatibility of *h* with *k*.

Measuring degree of confirmation by the *plausibility increment*,

$$dc(e,h) = P(h|eb) - P(h|b)$$

where *b* denotes the assumed background knowledge and assuming that *e* is a consequence of *h*, the product rule yields:

$$\begin{aligned} dc(e,hk) &= P(hk|eb) - P(hk|b) \\ &= P(k|heb)P(h|eb) - P(k|hb)P(h|b) \\ &= P(k|hb)[P(h|eb) - P(h|b)] \end{aligned}$$

or

$$dc(e,hk) = P(k|hb)dc(e,h)$$

which says the dc *e* accords the conjunction of *h* with *k* is the fraction, $p(k|hb)$, of the dc *e* accords *h* alone. This is what we expect.

Some 'counterinductive inferences' can be handled by treating the relevant 'bent' hypothesis as a conjunction. For example, 'All emeralds are grue' is, in essence, a conjunction of 'All examined emeralds are green' with 'All unexamined emeralds are blue'. Given our knowledge that emeralds do not vary in color depending on whether they have been examined or not, the dc which examining an emerald for color and finding it green accords the grue hypothesis is, by our result, a vanishingly small fraction of the dc that finding accords the hypothesis that unexamined emeralds are green. Contrast with this the bent hypothesis, 'All specimens of *Biston bistularia* (a British moth) are blite', where 'blite' applies to black denizens of industrially polluted regions or to light denizens of unpolluted regions. Since trees in polluted areas are lichenless and sooty, dark moths are less visible to predators against their black trunks than light ones, while the reverse is true in unpolluted regions where lichen covers the bark. Here the probability that moths of a *different* color will predominate in unpolluted regions is high, given this background knowledge. Thus, a bent hypothesis may be more confirmable by conforming data than the corresponding 'straight' hypothesis.

To be sure, the green-hypothesis is, equally, a conjunction of 'All examined emeralds are grue' with 'All unexamined emeralds are bleen'. But, presumably, the improbability we assign to a gem's color being dependent on whether or not it has been examined holds irrespective of whether we employ the green-blue or the grue-bleen languages. To be sure, one might challenge this assumption. My point is merely that at a level of analysis where the color of a gem is assumed not to depend on whether or not it has been examined, our confirmation theory can declare unequivocally in favour of the projection of 'All emeralds are green'.

What to incorporate in *b*, the background knowledge, is a matter of real concern. I tend to view confirmation as 'local': the immediate aim is to single out the 'best' of the considered theoretical alternatives. 'Extraneous' hypotheses are earmarked, in the first place, by their exclusion from the relevant Boolean algebra. (Should new connexions warrant their inclusion, that algebra can always be expanded to suit.) In practice, then, the problem of irrelevant conjunction has little more than academic interest. For example, *h* might be a hypothesis about parental genotypes, and *b* an assumed Mendelian model of some mating experiment. Then if *k* were, say, a very *specific* hypothesis about the velocity of neon light in beer, the probabilities assigned to *k* and $\sim k$ in light of *b* would be, patently, highly *diffuse* probabilities. In the absence of pertinent facts, the highly specific *k* would get a low probability and its negation a high probability. Then $P(k|hb) \approx P(k|b) \approx 0$ and $dc(e,hk) \approx 0$, while $dc(e,h.\sim k) \approx dc(e,h)$. In the light of such $b, \sim k$ represents a nearly vacuous addition to *h*. The point is that, under such 'local' construal of *b*, no incompatibility between *h* and *k* is needed to drive dc(e.hk) close to zero. Where *b* is more broadly construed, only the conditon $P(k|hb) \approx P(k|b)$ is left to identify *k* as 'extraneous'. Our haziness about how much of the confirmation of *h* should 'rub off' on its conjunction with *k* is due, at least in part, to our unclarity about what background knowledge is being assumed.

5 THE POPPER–MILLER CONUNDRUM

This analysis largely defangs 'irrelevant conjunction', the claim that Bayesian confirmation is too permissive. At the other extreme is the claim that it is too weak to support genuinely 'ampliative' inferences. In support of this thesis, Popper and Miller ([1983]; henceforth, P–M) urge: (i) $\sim e \vee h$ represents the 'excess content' of *h* over *e*, and (ii) *e* disconfirms $\sim e \vee h$.

The second claim is correct but unsurprising, though, as P–M point out ([1987], p. 576), the more general principle that evidence disconfirms a disjunction whenever it excludes one of the disjuncts is false. Their first claim is advanced on the grounds that $\sim e \vee h$ is the deductively weakest proposition which, together with *e*, yields *h*. Before addressing this claim, I offer a few general remarks.

Induction is being equated here with inference to a conclusion *h* that amplifies the evidence *e* in having no 'deductive dependence' on *e*. Traditionally, the prediction that the next examined *A* is *B* has been thought to amplify the evidence that past examined As are Bs. But P–M flatly deny this

([1987], p. 581), stating that 'in truth this prediction is far from being deductively independent of these instantial statements'. Then they triumphantly conclude that 'this overturns the thesis that is the mainstay of the programme of inductive logic'. *That* conclusion, at least, seems to 'amplify' the evidence. For the inductivist, it is enough that his prediction about unexamined *A*'s is not entailed by the evidence regarding examined *A*'s. P–M seem to be imputing a stronger thesis to inductivists than they have intended.

Much less does their claim about deductive dependence, even if well founded, overturn the *probabilism* I espouse in distinction to *inductivism* (Rosenkrantz [1977], Ch. 3). It will be clear even from the present paper that my Bayesianism eschews central tenets of inductivism, notably Nicod's criterion, the special consequence condition, and even the claim that inductive inferences and straight hypotheses are always preferable to counterinductive inferences and bent hypotheses. Probabilism affirms, first, that the degrees of confidence one may reasonably repose in the competing hypotheses of a partition are adequately represented by probabilities, and, second, that probabilities are to be adjusted in light of new data by Bayes' rule. Bayesian updating, and not the Humean principle of induction, is the basic mode of 'learning from experience'. If it should turn out that conditional probabilities merely reflect hidden deductive relations, all well and good. I know of no better way to mirror these (imputed) deductive dependencies.

Indeed, Bayes' rule is the only *consistent* way to mirror them. For, as I show in Rosenkrantz [1992], this rule for 'inverse probabilities' is a consequence of axioms governing 'direct probabilities', above all, the partitioning formula, $P(h|b)=\Sigma_j P(h|e_j b)P(e_j|b)$. Thus nonBayesian modes of updating are inconsistent with the usual axioms of probability, axioms which are themselves consistency requirements. The direct probabilities of which Popperians and others avail themselves *determine* the inverse probabilities they eschew. And while induction is not thereby reduced to deduction, inductive inferences are subject to consistency constraints.

The distance between my position and P–M's may seem rather small. I suspect the real differences show up when we ask what is the point of their argument? The only point evident to me is to suggest that severe testing of our conjectures, and not anything resembling induction or Bayesian updating, is all that science requires. If, as they aver, 'there is no probabilistic dependence without deductive dependence', then can't we make do with deduction alone?

I think not and refer you to their parting shot about the need to go on testing hypotheses no matter how much (putative) inductive support they have to their credit ([1987], p. 585). Unless we track the changes in our confidence by using Bayes' rule to update inductive probabilities, all unrefuted hypotheses remain equally trustworthy and equally testworthy. Detectives, auto mechanics, medical diagnosticians, search and rescue squads, and chemists in search of a molecular structure, all share a need to track changes in their degrees of belief if only to know what questions to pose, what tests to perform, or where to search, next.

And there is a related point. In P–M's view ([1987], p. 572) that hypotheses are confirmed by their consequences does not provide a probabilistic prop for genuinely ampliative inference. For all hypotheses which have e as a consequence are, they contend, supported by e to the same degree, including bent hypotheses, like 'All swans are black save those in Vienna in 1986'. I have undercut that contention (*v.s.*) by viewing such bent hypotheses as conjunctions. On the other side, irrelevant conjunction and bent hypotheses do pose real problems for the deductivism P–M espouse. Examining a swan for color tests the bent hypothesis as surely as it does its straight counterpart, and more generally, a test of h is, equally, a test of hk. In particular, ascertaining whether two men received their own hat assuredly tests the hypothesis that no man received his own, and the no-match hypothesis survives that test (does it not?) when the two men in question are found to have received *each other's* hat. These considerations fortify my suspicion that Bayesian inference is a more sensitive instrument for scientific and practical purposes than pure deduction alone.

In any case, P–M's attack on inductivism leaves probabilism unscathed (as they doubtless concede); no probabilist need take a stand on the issue they raise. Nevertheless, if forced to play their game, I should approach the explication of 'excess content' in probabilistic terms, viewing, as I do, deductive relations as limiting cases of probabilistic ones. Intuitively, what h says over and above what e says is that part of h about which e says nothing. The excess content of h must incorporate all those consequences k of h to which e is *otherwise irrelevant*, which condition precisely formulated reads:

$$(*)\ P(k|\sim heb) = P(k|\sim hb).$$

The rationale, then, is that we remove h from the background knowledge b and then see what effect the introduction of e has on the probability of k. Only when it has no effect may k be said to 'go beyond' e.

By a partitioning of probability, we have:

$$\begin{aligned} P(k|eb) &= P(k|heb)P(h|eb) + P(k|\sim heb)P(\sim h|eb) \\ &= P(h|eb)[1 - P(k|\sim hb)] + P(k|\sim hb) \end{aligned}$$

using $P(k|heb) = 1$ (since h entails k) and (*). Similarly,

$$P(k|b) = P(h|b)[1 - P(k|\sim hb)] + P(k|\sim hb)$$

and it follows that

$$(**)\ (P(k|eb) > P(k|b) \qquad \text{iff} \qquad P(h|eb) > P(h|b).$$

Thus, (*) is a sufficient condition for e to confirm a consequence k of a hypothesis *h* it confirms. More generally, *e* will confirm a consequence *k* of *h* whenever it is not *negatively* relevant to *k* against the residual background, $\sim hb$. In particular, *e* and *k* might be two 'otherwise unrelated' consequences of *h*, whose role is to forge an inductive link between them. At any rate, the proposed explication of 'excess content' leads to a conclusion diametrically opposed to that of Popper and Miller, namely, that evidence *e* favouring *h* *always* confirms a consequence of *h* that 'goes beyond' *e*.

For a concrete example, consider a roll of two dice, one red and the other white. Let *h* be the event 'sevens', *k* the consequence 'seven or under' (in symbols, $k=[r+w \leqslant 7]$), and $e=[r=3]$ the event that the red die shows a three-spot. For 'normal' dice, *e* is irrelevant to *h* and confirms *k* and (*) does not hold. If we retain independence and the fairness of the red die, then $P(h)=1/6$ still holds and

$$P(k|\sim he)=P(\sim hk|e)/P(\sim h/e)$$
$$=P(1 \leqslant w \leqslant 3)/[1-P(w=4)]$$

while

$$P(k|\sim h)=P(\text{under seven})/[1-P(h)]$$
$$=[P(r=1)P(w \leqslant 5)+\ \ldots\ +P(r=5)P(w=1)]/(5/6)$$
$$=[P(w=1)+P(w \leqslant 2)+\ \ldots\ +P(w \leqslant 5)]/5$$

so that (*) can be satisfied by biasing the white die in favor of higher values. With a little tinkering, I find that the probabilities 0·08, 0·10, 0·12, 0·19, 0·31, 0·20 for $[w=j]$, $j=1, \ldots, 6$, will do. We then have $P(h|eb)=P(w=4)=0{\cdot}19>P(h)$, $P(k|\sim heb)=[0{\cdot}08+0{\cdot}10+0{\cdot}12]/0{\cdot}81=0{\cdot}3704$, while $P(k|\sim hb)=[0{\cdot}08+0{\cdot}18+0{\cdot}3+0{\cdot}49+0{\cdot}8]/5=0{\cdot}3700$, so that (*) holds almost exactly. Finally, $P(k|eb)=P(h|eb)+P(\text{under seven}|eb)=0{\cdot}19+0{\cdot}30=0{\cdot}49>0{\cdot}475=P(h|b)+P(\text{under}|b)=P(k|b)$, so that, in accord with our result, *e* still confirms *k*.

In a straightforward sense, *k* 'amplifies' *e* by requiring that the sum of spots be at most seven, where *e* only constrains it to be at most nine. But to decide if *k* 'goes beyond' *e* in the deeper probabilistic sense that *e* is irrelevant to *k* in light of $\sim hb$, we must look at the joint probability distribution *b* in detail. For a distribution that made the outcome of the white die highly dependent on that of the red die, *k* will be less apt to add content to *e*.

Even where *e lowers* the probability of *k* in the presence of $\sim h$, it may still confirm *k*. Remember, (*) is a *sufficient* condition, not necessary. Nevertheless, it is a near thing. We are going to be in trouble if we provide *carte blanche* to claim confirmation for *any* consequence of a confirmed hypothesis. Equally, we are in trouble if we disallow that transfer of credit in every case. The P–M puzzle quite properly impels us to formulate the conditions under which this transfer does occur.

REFERENCES

ARONSON, JERROLD [1989]: 'The Bayesians and the Raven Paradox', *Nous*, **23**, pp. 221–40.

GOOD, I. J. [1967]: 'The White Shoe is a Red Herring', *British Journal for the Philosophy of Science*, **19**, pp. 322–3.

HEMPEL, C. G. [1945]: 'Studies in the Logic of Confirmation', *Mind*, **LIV**, pp. 1–26, 97–120.

POPPER, K. R. [1959]: *The Logic of Scientific Discovery*. Hutchinson.

POPPER, K. R. and MILLER, D. W. [1983]: 'A Proof of the Impossibility of Inductive Probability', *Nature, Lond.*, **302**, pp. 687f.

POPPER, K. R. and MILLER, D. W. [1987]: 'Why Probabilistic Support is not Inductive', *Phil. Trans. Royal Soc. Lond.*, **321**, pp. 569–91.

ROSENKRANTZ, R. D. [1977]: *Inference, Method, and Decision*. Reidel.

ROSENKRANTZ, R. D. [1982]: 'Does the Philosophy of Induction Rest on a Mistake?' *Journal of Philosphy*, **LXXIX**, pp. 78–97.

ROSENKRANTZ, R. D. [1992]: 'The Justification of Induction', *Philosophy of Science*, **59**, 4, pp. 527–39.

SUPPES, PATRICK [1966]: 'A Bayesian Approach to the Paradoxes of Confirmation' in Jaako Hintikka and Patrick Suppes (*eds.*): *Aspects of Inductive Logic*. North-Holland, pp. 198–207.

WATKINS, J. W. N. [1964]: 'Confirmation, the Paradoxes, and Positivism', in M. Bunge (*ed.*): *The Critical Approach to Science and Philosophy*. The Free Press.