PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Why Probabilistic Support is not Inductive

K. Popper and D. W. Miller

Phil. Trans. R. Soc. Lond. A 1987 **321**, 569-591
doi: 10.1098/rsta.1987.0033

Phil. Trans. R. Soc. Lond. A **321**, 569–591 (1987) [569]
Printed in Great Britain

WHY PROBABILISTIC SUPPORT IS NOT INDUCTIVE†

By K. POPPER[1], F.R.S., and D. W. MILLER[2]

[1] *London School of Economics, Houghton Street, Aldwych, London WC2A 2AE, U.K.*
[2] *Department of Philosophy, University of Warwick, Coventry CV4 7AL, U.K.*

(*Received* 5 *June* 1986)

CONTENTS

		PAGE
1.	Logical and probabilistic dependence	570
2.	Main results on counterdependence	575
3.	Discussion: logical content	579
	Appendix	585
	References	590

This paper attempts to provide both an elaboration and a strengthening of the thesis of Popper & Miller (*Nature, Lond.* **302**, 687f. (1983)) that *probabilistic support is not inductive support*. Although evidence may raise the probability of a hypothesis above the value it achieves on background knowledge alone, every such increase in probability has to be attributed entirely to the *deductive connections* that exist between the hypothesis and the evidence. We shall also do our best to answer all the criticisms of our thesis that are known to us.

In 1878 Peirce drew a sharp distinction between '*explicative, analytic*, or *deductive*' and '*amplifiative, synthetic*, or (loosely speaking) *inductive*' reasoning. He characterized the latter as reasoning in which 'the facts summed up in the conclusion are not among those stated in the premisses'. *The Oxford English Dictionary* records that the word 'ampliative' was used in the same sense as early as 1842, and that in 1852 Hamilton wrote: 'Philosophy is a transition from absolute ignorance to science, and its procedure is therefore ampliative.'

This was the background to our letter to *Nature* on 21 April 1983. It was there shown that, relative to evidence e, the content of any hypothesis h may be split into two parts, the disjunction $h \vee e$ (read h or e) and the material conditional $h \leftarrow e$ (read h if e); and the 'ampliative' part of h relative to e was identified with this conditional $h \leftarrow e$; that is, with the deductively weakest proposition that is sufficient, in the presence of e, to yield h. We then proved quite generally that if $p(h, eb) \neq 1 \neq p(e, b)$ then

$$s(h \leftarrow e, e, b) = -ct(h, eb)\, ct(e, b) < 0. \tag{0.1}$$

Here $s(x, y, z) = p(x, yz) - p(x, z)$ is a measure of the *support* that y gives to x in the presence of z, and $ct(x, z) = 1 - p(x, z)$ is a measure of the *content* of x relative to z. Relative to b, both h's content and its support by e may be added over the two factors $h \vee e$ and $h \leftarrow e$:

$$ct(h, b) = ct(h \vee e, b) + ct(h \leftarrow e, b), \tag{0.2}$$

$$s(h, e, b) = s(h \vee e, e, b) + s(h \leftarrow e, e, b). \tag{0.3}$$

† *In memoriam* Alfred Tarski, 1902–1983.

What (0.1) establishes is that *the* ('ampliative') *part of the hypothesis h that goes beyond the evidence e is invariably countersupported by the evidence*. In other words, *probabilistic support is not inductive support*.

1. Logical and probabilistic dependence

We shall write $p(x,y)$ for the *logical probability* of the proposition x *given* y, or *relative to* y, or *in the presence of* y. Several equivalent axiomatic systems for relative probability are to be found in appendixes *iv and *v of Popper (1959), and we shall have recourse to some of their theorems on a number of occasions. All these systems assume that with any elements x, y of the set S on which p is defined there is associated an element xy (which is standardly interpreted as the *conjunction* or *product* of x and y) and with each element z there is associated an element z' (the *negation* or *complement* of z; in Popper (1959) it is written $\bar{z}$). It is of some significance, however, that most of the results that we shall present here can be proved without assuming the presence of any operation of complementation in the set S, provided that we assume the existence of a *disjunction* or *sum* $x \vee y$ for each x and y. What is crucial is that we have at our disposal the familiar generalized *multiplication* and *addition* laws of the calculus of probability,

$$p(xy,z) = p(x,yz)\,p(y,z), \tag{1.1}$$

$$p(x \vee y,z) + p(xy,z) = p(x,z) + p(y,z), \tag{1.2}$$

to which we shall sometimes refer by name, as well as three other scarcely controversial theorems:

$$0 \leqslant p(x,y) \leqslant 1, \tag{1.3}$$

$$p(xy,z) \leqslant p(xz,z) = p(x,z), \tag{1.4}$$

$$y \vdash x \quad \text{iff} \quad p(x,yz) = 1 \text{ for every } z. \tag{1.5}$$

Here and elsewhere $y \vdash x$ can be taken as an abbreviation of

$$(z)\,[p(y,z) \leqslant p(x,z)]. \tag{1.6}$$

It is indeed the familiar relation of logical (or formal) implication. It is a partial (pre)ordering of the elements of the set S.[1]†

Note that even in the presence of negation (defined, say, by $p(x',y) = 1-p(x,y)$ provided $p(z,y) \neq 1$ for some z) it is not required that y in $p(x,y)$ be a consistent or non-zero element of the set S: $p(x,yy')$ is a perfectly acceptable term. Nor are there any other conditions imposed on the elements of S, beyond those embodied in the probability axioms. (It is not even necessary to insist separately that the elements of S conform to the laws of classical logic, because this is a consequence of the probability axioms.) Because of this the treatment throughout the paper is easily converted, if required, into an immaculately formal one. Our conclusions accordingly hold with the greatest generality.

Two propositions x and y may be said to be probabilistically dependent on each other in the presence of z if the probability $p(xy,z)$ is different from the product of the probabilities $p(x,z)$ and $p(y,z)$; *positively dependent* or simply *dependent* if $p(xy,z) > p(x,z)\,p(y,z)$ and *negatively*

† Numbers given in this form refer to the notes at the end of the text.

dependent, or *counterdependent*, if $p(xy, z) < p(x, z)\, p(y, z)$. This relation is obviously symmetric in x and y, and it is natural to introduce the difference

$$d(x, y, z) = p(xy, z) - p(x, z)\, p(y, z) \tag{1.7}$$

as a measure of the mutual dependence of x and y. This is a very simple measure of dependence, though perhaps not entirely adequate for all purposes when either $p(x, z)$ or $p(y, z)$ is equal to 0; for then, by (1.4) above, $p(xy, z)$ must be 0, so that $d(x, y, z)$ equals 0 as well. When $d(x, y, z) = 0$ we shall normally call x and y probabilistically *independent* (in the presence of z). It may seem excessive to have to say that any proposition with zero probability, even the contradiction yy', is probabilistically independent of every proposition (including itself). We need not, however, worry too much about a zero case and, for the next few paragraphs, we shall tacitly exclude such propositions from consideration. It is perhaps a little worse that if either $p(x, z)$ or $p(y, z)$ equals 1 then x and y are probabilistically independent (given z). (For suppose $p(x, z) = 1$. By (1.4) $p(x \vee y, z) = 1$, so by (1.2) $p(xy, z) = p(y, z) = p(x, z)\, p(y, z)$.) We shall shortly introduce an alternative measure of probabilistic independence that is marginally more satisfactory in the way it handles propositions of zero and unit probability.

Our chief concern is the connection between probabilistic dependence on the one side and *deductive* (or *logical*) dependence on the other. The latter relation, if held to obtain between x and y if and only if $x \vdash y$, would not be a symmetric one; for x may imply y but y not imply x. And for this reason we may be inclined to call x and y deductively dependent whenever either of them implies the other. Now this sort of deductive dependence, though symmetric like probabilistic dependence, is unlike it in not admitting degrees. Since our ambition in this paper is to lay bare the extent to which any relation of probabilistic dependence (often called *probabilistic support*) between two propositions is no more than a reflection of a relation of deductive dependence between them, it will be useful here to indicate how the relation described above may be generalized.

The matter is actually quite straightforward. For clearly the greatest degree of deductive dependence between x and y is obtained when one of them implies all that the other implies; the lowest degree when neither implies any non-trivial proposition that the other implies (all propositions imply all logical truths). It seems unavoidable therefore that we should reckon the extent of the deductive dependence of x and y on each other by how much of the other each of them implies. This is still a rather loose formulation (it will be tightened below). But for the moment it should suffice.[2]

Let b, h, e be any propositions whatever. It must be stressed that for the purpose of our derivations we assume absolutely nothing material about these propositions, though for the sake of easy comprehension we shall often refer to b as *background knowledge*, to h as an arbitrary *hypothesis* or conjecture, and to e as empirical *evidence*. It should be noted too that we shall state explicitly any assumptions we may have occasion to make concerning the existence of relations of deductive implication among b, h, and e.

Suppose first that in the presence of b the hypothesis h implies the evidence e; that is, that $hb \vdash e$. Then, by (1.3) and (1.5) above, $p(e, hb) = 1 \geqslant p(e, b)$. By the multiplication law (1.1),

$$p(eh, b) = p(e, hb)\, p(h, b) \geqslant p(e, b)\, p(h, b);$$

so if $p(e, b) < 1$ (and $p(h, b) > 0$) then e and h are probabilistically dependent on each other as well as being deductively dependent on each other. This is not very surprising, of course.

By another use of (1.1), moreover, the probabilistic dependence between h and e can be written also:

$$p(h, eb) > p(h, b)$$

(it still being understood that $p(e, b) < 1$, and, where necessary, that some of the other probabilities exceed 0). This simple and well known result is sometimes expressed by saying that, except in extreme cases, the consequences of a hypothesis h lend it positive support. It seems to be this property of positive dependence (or positive support) that has encouraged the dream that the theory of probability is the key to the problem of providing a working system of inductive logic. For the probability of a hypothesis h cannot be lowered, and may well be raised, each time a logical consequence e of it is verified in experience.

Our view is that it is a mistake to interpret this theorem in this way. Given the symmetrical character of probabilistic dependence, the result is a direct reflection of the deductive dependence of e on h and b. There is no sense in which e, by raising the probability of a hypothesis h from which it follows, points beyond itself. This should be obvious when we consider that such an e raises the probability of every proposition h from which it follows in the presence of b.

Indeed, matters are considerably more acute than this. By (1.1), if the hypothesis h logically implies the evidence e in the presence of b (so that he is equivalent to h) then $p(h, eb)$ is proportional to $p(h, b)$; in fact, $p(h, eb) = p(h, b)/p(e, b)$. So suppose that e is some such evidence statement as 'All swans in Vienna in 1986 are white', h the supposedly inductive generalization 'All swans are white' and k the counterinductive generalization 'All swans are black, except those in Vienna in 1986, which are white'. Then $p(h, eb)/p(k, eb) = p(h, b)/p(k, b)$. No matter how h and k generalize on the evidence e, this evidence is unable to disturb the ratio of their probabilities (see Popper 1983, pp. xxxvii–xxxix; Popper 1985, pp. 312–316). Supporting evidence points in all directions at once, and therefore points usefully in no direction.

For this reason it seems sensible to try to eliminate from each hypothesis h that part of it that is no longer hypothetical given eb; that is, to isolate that part of h that is in no manner deductively dependent on eb. If we do this, we come to a very unexpected result. For we find that what is left of h once we discard from it everything that is logically implied by e, is a proposition that in general is counterdependent on e; which is to say that there is no positive probabilistic dependence of h on e except where there is also positive deductive dependence.[3]

This constitutes our main thesis, and will be greatly elaborated in the remainder of the paper. We now discard the assumption that hb implies e; from this point on h, e, b will be entirely arbitrary propositions.

A simple illustration of our thesis is provided by the conditions for the probabilistic independence of h and e given b, which was defined above by the equation $p(he, b) = p(h, b)\,p(e, b)$. It is of some interest that this equation is obtained if and only if there is a balance between two terms: one is the measure of deductive dependence between h and e; the other can be thought of only as a measure of non-deductive counterdependence between h and e. (Perhaps it would be better to say that these two kinds of dependence, the non-deductive and the deductive, annul each other or cancel each other out.) What we are able to prove is that for any h, e, b

$$d(h, e, b) = d(h \vee e, e, b) + d(h \leftarrow e, e, b), \tag{1.8}$$

where $h \leftarrow e$ abbreviates $(h'e)'$ and $h \vee e$ abbreviates $h \leftarrow e'$ (endnote 4). (Here it is taken for granted that the product and complement conform to the restrictions of classical logic.) Each of $h \leftarrow e$ and $h \vee e$ is a consequence of h and it is just when they depend on e to equal and opposite degrees that h and e are themselves probabilistically independent; that is, that $d(h, e, b) = 0$. It will be readily apparent that the proposition $h \vee e$ is a consequence of e also; indeed, that $h \vee e$ is simply what h and e have in common deductively. Because of this the first summand on the right of (1.8) is, we submit, a fair measure of the degree of deductive dependence between h and e. It is, as we have already noted above, never negative (because $h \vee e$ is a consequence of e) and will be positive provided that $p(h \vee e, b) < 1$. In the latter case (1.8) shows that h will be probabilistically independent of e only if there is a part of it (namely $h \leftarrow e$) that is probabilistically counterdependent on e. But this counterdependence is clearly quite separate from the deductive dependence between h and e. Indeed, $h \leftarrow e$ and e are themselves deductively independent propositions (provided that neither is a logical truth); but though the degree of deductive dependence between them is 0, their degree of probabilistic dependence $d(h \leftarrow e, e, b)$ is negative.

The proof of (1.8) is comfortably straightforward. Because for any e and h

$$(h \vee e)\,(h \leftarrow e) \text{ is equivalent to } h,$$

$$(h \vee e) \vee (h \leftarrow e) \text{ is equivalent to a tautology,}$$

and by (1.5) tautologies have probability 1 relative to anything at all, we have by the general addition law (1.2)

$$p((h \vee e)\, e, b) + p((h \leftarrow e)\, e, b) = p(he, b) + p(e, b) \tag{1.9}$$

$$p(h \vee e, b) + p(h \leftarrow e, b) = p(h, b) + 1. \tag{1.10}$$

If we multiply (1.10) throughout by $p(e, b)$ and subtract it from (1.9) the result is proved.

To simplify the mathematical presentation of the rest of the paper we shall henceforth suppress explicit mention of the background knowledge b in expressions of the form $p(x, yb)$ and $p(x, b)$, writing them instead as $p(x, y)$ and $p(x)$. The latter expression may be thought of as an abbreviation of $p(x, t)$ where t is any tautology. The adoption of this notational convention will not in any way affect the correctness of any of our results.[5]

Thus we shall write $d(h, e)$ for the degree of probabilistic dependence between h and e as defined above. Other measures of dependence will be introduced shortly, and they too will be written with the term b suppressed.

The rest of this section will be an illustration of the charming remark of Levi (1984) that 'probabilistic support speaks with many tongues'.

In the presence of e the proposition h is equivalent to many others; indeed, provided that the condition $he \vdash x \vdash h \leftarrow e$ is satisfied h will be logically equivalent to x in the presence of e. It follows that he and xe are logically equivalent, as are $h \leftarrow e$ and $x \leftarrow e$. Thus for any proposition x in this range, $d(x, e)$ is equal to $p(he) - p(x)\,p(e)$, which is to say that it increases as x gets stronger (and $p(x)$ accordingly smaller). In response to this simple observation we may define maximal and minimal measures of the dependence of h on e:

$$d^{\mathrm{m}}(h, e) = d(he, e) \tag{1.11}$$

$$d_{\mathrm{m}}(h, e) = d(h \leftarrow e, e). \tag{1.12}$$

It may be noted that these measures are no longer symmetric in h and e. But clearly we shall have

$$\text{if} \quad he \vdash y \vdash x \vdash h \leftarrow e \quad \text{then} \quad d^{\mathrm{m}}(h, e) \geqslant d(y, e) \geqslant d(x, e) \geqslant d_{\mathrm{m}}(x, e) = d_{\mathrm{m}}(h, e). \tag{1.13}$$

Because in the presence of e every such x or y is, as already noted, logically equivalent to h, one may well ponder why they depend probabilistically on e to such markedly different degrees. The answer, we suggest, is simply that the various different xs have different degrees of *deductive* dependence on e, ranging from a stark and obvious dependence in the case of he at the top down to total deductive independence in the case of $h \leftarrow e$ at the bottom. For when y and x are in this range $y \leftarrow e$ is equivalent to $x \leftarrow e$, so by (1.8) above if $d(y, e)$ exceeds $d(x, e)$ it is entirely because $d(y \vee e, e)$ exceeds $d(x \vee e, e)$. That is, y and e are deductively more dependent on each other than x and e are.

This makes d_m of particular significance as a measure of the non-deductive dependence of h on e; for it is the only measure that is not artificially boosted by the presence of some deductive dependence. Of all the propositions x that together with e suffice to imply h, $h \leftarrow e$ is the only one for which $d(x \vee e, e)$ is obliged to be zero.

We must now take proper notice of the main disadvantages of the measure $d(h, e)$, that it is zero whenever either $p(h)$ or $p(e)$ is zero and whenever either of them is unity. These shortcomings may be partly corrected by moving to the measure $s(h, e)$ defined by

$$s(h, e) = p(h, e) - p(h). \tag{1.14}$$

When $p(e)$ is positive, $s(h, e)$ is of course equal to $d(h, e)/p(e)$; but it will be defined even when $p(e)$ is zero, and may be positive, negative, or zero under these conditions. If $p(h)$ is zero but $p(e)$ is positive $s(h, e)$ will also be zero. It is to be observed that $s(h, e)$ is no longer symmetric in h and e, so may be thought of as measuring the dependence of h on e, rather than a mutual dependence. Traditionally s has been widely esteemed as a measure of probabilistic *support*.

Almost everything that has been done with d above can be done also with s. In particular, the direct parallel to the additive principle (1.8) can be established in the same way:

$$s(h, e) = s(h \vee e, e) + s(h \leftarrow e, e); \tag{1.15}$$

and $s(h \vee e, e)$ stands out as a far from inadequate measure of the mutual deductive dependence between h and e (endnote 6). Furthermore, maximal and minimal measures s^m and s_m can be defined as in (1.11) and (1.12); and the analogue of the monotony principle (1.13) continues to hold good. All this is quite obvious, because when e is fixed d is a constant multiple of s (zero if $p(e)$ is zero). Because it vanishes less often we shall in the sequel concern ourselves wholly with s, rather than with d.

To sum up the argument of this section: if there were to be some genuinely inductive dependence between a hypothesis h and some evidence e, it could hardly change if h were replaced by some hypothesis x equivalent to h (given e, or equivalent to h in the presence of e). This much has recently been argued at length by Levi (1986). We wish, however, to go further and to stress that, whatever it is that $d(he, e)$ and $s(he, e)$ measure, it cannot be *pure* inductive dependence. And, in general, the same goes for $d(h, e)$ and $s(h, e)$; unless h happens to be deductively independent from e, the values of $d(h, e)$ and $s(h, e)$ are deductively contaminated. If there is such a thing as pure *inductive* dependence at all, there seems nothing for it but to measure it by something like $s(h \leftarrow e, e)$ or $d(h \leftarrow e, e)$.

However, as we shall show, $s(h \leftarrow e, e)$ is never positive (so far we have shown this only where h and e are mutually independent), and only in uninteresting circumstances is $s(h \leftarrow e, e)$ not negative. Inductive dependence is counterdependence. This overturns the thesis that is the mainstay of the programme of inductive logic.

2. Main results on counterdependence

In this section we shall prove a number of theorems on the theme that any probabilistic dependence $s(h, e)$ of a hypothesis h on evidence e, or probabilistic support of h by e, is entirely attributable to the extent of the deductive dependence of h on e. All non-deductive dependence, we aim to show, is counterdependence.

Our first and central theorem is proved in a way similar to (1.15) above. It calculates the degree of probabilistic dependence on e of a hypothesis that has no deductive dependence on e, and shows that it is never positive.

Theorem. 1. *If* $\vdash k \vee e$ *then*

$$s(k,e) = -[1-p(k,e)]\,[1-p(e)] = -ct(k,e)\,ct(e).$$

Proof. By the addition law (1.2) both

$$p(k,e)+p(e,e) = p(ke,e)+p(k \vee e,e)$$

and

$$p(k)+p(e) = p(ke)+p(k \vee e).$$

By subtraction,

$$s(k,e)+s(e,e) = [p(ke,e)-p(ke)]+s(k \vee e,e),$$

so

$$\begin{aligned} s(k,e)-s(k \vee e,e) &= [p(ke,e)-p(ke,e)\,p(e)]-[1-p(e)] \\ &= -[1-p(ke,e)]\,[1-p(e)] \\ &= -[1-p(k,e)]\,[1-p(e)]. \end{aligned}$$

By hypothesis, $p(k \vee e,e) = p(k \vee e) = 1$; thus $s(k \vee e,e) = 0$ and the result follows. ∎

Corollary. 1. $s(h \vee e,e)-s(h,e) = [1-p(h,e)]\,[1-p(e)]$
$= ct(h,e)\,ct(e).$

Proof. Immediate from the proof above. ∎

Corollary. 2. $s(h,e) \leqslant s(h \vee e,e) = 1-p(h \vee e)$
$= ct(h \vee e).$

Proof. Immediate. ∎

It is worth noting that theorem 1 and its corollaries have been proved without any assumption concerning the existence of an operation of complementation in the set S. Although stated rather generally, what the theorem really establishes is that if k is a consequence of h that has no non-trivial consequences in common with $h \vee e$ then $s(k,e) \leqslant 0$. For the latter condition amounts to saying that $\vdash k \vee h \vee e$; if $h \vdash k$, that is to say, then $\vdash k \vee e$. So $s(k,e) \leqslant 0$. In other words, if k is a consequence of h that has no deductive dependence on $h \vee e$, then k is countersupported by e. This result was first noted in Popper (1983, part II, section 16, p. 326, note 2). (See also Popper 1985.)

The logically strongest k that is a consequence of h and is deductively independent of $h \vee e$ is the proposition $h \leftarrow e$.

Theorem 2. $s(h \leftarrow e,e) = -[1-p(h,e)]\,[1-p(e)]$
$= -ct(h,e)\,ct(e).$

Proof. Because $\vdash (h \leftarrow e) \vee e$, it follows at once from theorem 1 that $s(h \leftarrow e,e)$ is equal to $-[1-p(h \leftarrow e,e)]\,[1-p(e)]$. By (1.4), $p(h \leftarrow e,e)$ is equal to $p(he,e)$, and so to $p(h,e)$. ∎

It is perhaps appropriate to draw explicit attention to a way in which the counterdependence of $h \leftarrow e$ on e cannot be proved. On p. 20 of the introduction to his Italian translation of Popper & Miller (1983) Marcello Pera (Pera 1983) suggests that because $h \leftarrow e$ is logically equivalent to $e' \vee h$, it is not to be wondered at that this proposition is counterdependent on e, 'seeing that the probability of a disjunction goes down if one of its disjuncts is negated'. A similar suggestion seems to be made by Redhead (1985, p. 188), who writes that 'the countersupport of $h \leftarrow e$ depends just on the fact that knowing e rules out the cases of $h \leftarrow e$ being true in virtue of e being false'. If this is meant to be an appeal to the persuasive sounding principle that evidence that knocks out one disjunct will countersupport a disjunction, then it is largely in vain. The principle is false. A simple counterexample is provided by selecting a fair die and writing k for the proposition 'five is thrown', h for 'six is thrown', and e for 'A number divisible by three is thrown'. It is easily checked that although e knocks out k, the probability of $h \vee k$ is raised from $\frac{1}{3}$ to $\frac{1}{2}$ by the discovery of e. Theorem 2, though not hard to prove, is therefore not quite as trite as some writers have made it out to be. Nor is it as lacking in significance as some writers (for example, Gaifman (1985)) have suggested.[7]

The result announced in theorem 2 is very little different from the calculation (first carried out by one of the authors in 1938) of the excess of the probability of the conditional $h \leftarrow e$ over what is usually referred to as the conditional probability $p(h,e)$. For because $p(x,y) = p(xy,y)$ and $(h \leftarrow e)\,e$ is equivalent to he it follows at once that $p(h \leftarrow e, e) = p(he, e) = p(h, e)$. The value of this never negative excess was first published, as far as we know, in Popper (1963, p. 396, formula 22) (see also Popper 1966, p. 307), and has since been noted by Lewis (1976, p. 306). The next theorem shows that the inequality can be considerably refined; not only is $p(x \leftarrow y)$ always at least as big as $p(x,y)$; in addition, $p(x, x \vee y)$ always lies between these two extremes.

LEMMA 1. *$p(x, x \vee y) - p(x,y)$ is equal to*

$$[1-p(x,x \vee y)]\,[1-p(y,x \vee y)]/p(y,x \vee y) \quad \textit{if} \quad p(y,x \vee y) > 0,$$

and to $1-p(x,y)$ *if* $p(y, x \vee y) = 0$.

Proof. Note that by (1.1) and (1.2)

$$\begin{aligned}[p(x,x \vee y) - p(x,y)]\,&p(y,x \vee y)\\ &= p(x,x \vee y)\,p(y,x \vee y) - p(x,y(x \vee y))\,p(y,x \vee y)\\ &= p(x,x \vee y)\,p(y,x \vee y) - p(xy,x \vee y)\\ &= p(x,x \vee y)\,p(y,x \vee y) - [p(x,x \vee y) + p(y,x \vee y) - p(x \vee y,x \vee y)]\\ &= [1-p(x,x \vee y)]\,[1-p(y,x \vee y)],\end{aligned}$$

by (1.4). Thus if $p(y, x \vee y) > 0$ the announced equality is proved. If, on the other hand, $p(y, x \vee y) = 0$ then the final product above is also 0, so its first factor is 0. Thus $p(x, x \vee y) = 1$. ■

LEMMA 2. *$p(x \leftarrow y) - p(x, x \vee y)$ is equal to*

$$[1-p(x \leftarrow y)]\,[1-p(x \vee y)]/p(x \vee y) \quad \textit{if} \quad p(x \vee y) > 0,$$

and to $1-p(x, x \vee y)$ *if* $p(x \vee y) = 0$.

Proof. Here we use the fact that the disjunction of $x \leftarrow y$ and $x \vee y$ is a tautology:

$$\begin{aligned}[p(x \leftarrow y) - p(x,x \vee y)]\,p(x \vee y) &= p(x \leftarrow y)\,p(x \vee y) - p(x(x \vee y))\\ &= p(x \leftarrow y)\,p(x \vee y) - p(x)\\ &= p(x \leftarrow y)\,p(x \vee y) - [p(x \leftarrow y) + p(x \vee y) - 1]\\ &= [1-p(x \leftarrow y)]\,[1-p(x \vee y)].\end{aligned}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

If $p(x \vee y) > 0$ we may divide by it. Otherwise, in the same manner as before we may conclude that $p(x \leftarrow y) = 1$, and the proof of the lemma is complete. ■

THEOREM 3. $p(x, y) \leqslant p(x, x \vee y) \leqslant p(x \leftarrow y)$.

Proof. By (1.3) all probabilities lie between 0 and 1 inclusive. Thus the theorem follows at once from the two preceding lemmas. ■

Our next theorem improves significantly on corollary 2 to theorem 1. Like lemma 1 it is proved without any assumption concerning the existence of an operation of complementation in the set S.

THEOREM 4.
$$\begin{aligned} s(h, e) &\leqslant s(h, h \vee e) \\ &= [1 - p(h \vee e)]\, p(h, h \vee e) \\ &= s(h \vee e, e)\, p(h, h \vee e) \\ &\leqslant s(h \vee e, e). \end{aligned}$$

Proof. By lemma 1, $p(h, e) \leqslant p(h, h \vee e)$. Subtraction of $p(h)$ from both sides yields the first inequality. The rest of the theorem follows by the theorems (1.1), (1.5) and (1.3) adduced above. ■

COROLLARY. *Provided that* $p(h) > 0$, *so that* $p(h \vee e) > 0$,
$$p(h, e)/p(h) \leqslant p(h \vee e, e)/p(h \vee e).$$

Proof. If each side of the inequality
$$p(h, e) \leqslant p(h, h \vee e)$$
is multiplied by $p(h \vee e)$, we obtain
$$p(h, e)\, p(h \vee e) \leqslant p(h).$$
By division by $p(h \vee e)\, p(h)$,
$$p(h, e)/p(h) \leqslant 1/p(h \vee e).$$
As $p(h \vee e, e) = 1$ (by (1.5)), the result is immediate. ■

This corollary is not without interest; what it tells us is that if probabilistic dependence or support is measured not by $s(h, e)$ but by $\sigma(h, e) = p(h, e)/p(h)$, as mooted by Redhead (1985, p. 190), then the inequality corresponding to corollary 2 of theorem 1 holds:

$$\sigma(h, e) \leqslant \sigma(h \vee e, e). \tag{2.1}$$

The analogue of (1.13) also holds:

$$\text{if} \quad he \vdash y \vdash x \vdash h \leftarrow e \quad \text{then} \quad \begin{aligned} &\sigma(he, e) \geqslant \sigma(y, e) \geqslant \sigma(x, e) \\ &\geqslant \sigma(x \leftarrow e, e) \geqslant \sigma(h \leftarrow e, e). \end{aligned} \tag{2.2}$$

The measure σ, no less than the measure s, sustains the view that differences in probabilistic dependence are to be attributed entirely to differences in degrees of deductive dependence.[8]

The representation of a hypothesis h as the product (or conjunction) of $h \vee e$ and $h \leftarrow e$ is unique in that for no other factorization is the sum (or disjunction) of the factors a tautology. Indeed, provided that neither of these factors is itself a tautology, they are maximally independent of each other. We have seen in corollary 2 to theorem 1, and in theorem 2, that the probabilistic dependence of these factors on e is very fiercely constrained: $s(h \vee e, e)$ is never less than $s(h, e)$, whereas $s(h \leftarrow e, e)$ is never greater than 0. Indeed, as we know:

$$s(h, e) = s(h \vee e, e) + s(h \leftarrow e, e). \tag{1.15}$$

Now theorem 4 shows that if we perform a similar factorization on the evidence statement e rather than on the hypothesis h, a similar result is obtained for $h \vee e$: $s(h, h \vee e)$ is never less than $s(h, e)$. The next theorem provides the complementary result for $h \rightarrow e$, but shows that in general there is no additivity corresponding to (1.15).

THEOREM 5. (i) $s(h, h \rightarrow e) = -[1-p(h, h \rightarrow e)][1-p(h \rightarrow e)]$.

(ii) $s(h, e) = s(h, h \vee e) + s(h, h \rightarrow e)$

if and only if either $p(he) = p(h')p(e)$,

or $p(h \vee e) = 1$, *so that* $s(h, h \vee e) = 0$,

or $p(h \vee e) = p(e)$, *so that* $s(h, h \rightarrow e) = 0$.

Proof. We may prove (i) immediately from theorem 2 by writing $h \rightarrow e$ for e and using the equivalence of $h \leftarrow (h \rightarrow e)$ with h.

The proof of (ii) involves some elementary but not very interesting algebra. It is omitted. ■

Although part (i) of theorem 5 nicely matches theorem 2, we are not aware of any significant interpretation of part (ii). It is included only in order to complete the record.

To conclude this section we turn our attention to factorizations that are probabilistically rather than deductively independent. It will become clear at the end of the next section that for some writers on the subject such factorizations are thought to be of some importance.

Theorem 6, which also is free from any assumption concerning complementation, seems to be rather powerful.

THEOREM 6. *Provided that each of the conditions*

(*a*) $$p((x \vee y)(x \vee z)) = p(x \vee y)p(x \vee z)$$

(*b*) $$0 < p(y)$$

holds, $s(x \vee z, y) \leqslant 0$. *The inequality is a strict one provided that, in addition,*

(*c*) $$p(x \vee z) \neq 1,$$

(*d*) $$p(y) \neq p(x \vee y).$$

Proof. If $p(x \vee z) = 0$ then $p(x \vee z, y)$ does too, because $p(y)$ exceeds 0. Thus the first conclusion is immediate.

We may therefore assume that $p(x \vee z)$ is greater than 0. Because $(x \vee y)(x \vee z)$ is equivalent to $x \vee y(x \vee z)$, it follows from (*a*) and the addition law (1.2) that

$$p(x) + p(y(x \vee z)) - p(xy(x \vee z)) = [p(x) + p(y) - p(xy)]p(x \vee z).$$

Now $xy(x \vee z)$ is equivalent to xy, so collecting terms,

$$\begin{aligned}[1-p(x \vee z)][p(x)-p(xy)] &= p(y)p(x \vee z) - p(y(x \vee z)) \\ &= p(x \vee z)[p(y) - p(y, x \vee z)] \\ &= -p(x \vee z)s(y, x \vee z).\end{aligned}$$

The two factors on the left are never negative; thus $s(y, x \vee z) \leqslant 0$. As $p(y)$ is not 0 it follows that $s(x \vee z, y) \leqslant 0$; and this proves the first part of the theorem.

The addition law and (*d*) ensure that $p(x) \neq p(xy)$; so if (*c*) holds, the left side of the above equation is positive. Thus $s(y, x \vee z) < 0$. ■

COROLLARY. *If* $h \vdash f$ *and, in addition,*

(*a*) $$p((h \vee e)f) = p(h \vee e)\, p(f),$$

(*b*) $$0 \neq p(e) \neq p(h \vee e),$$

(*c*) $$p(f) \neq 1,$$

then $s(f, e) < 0$.

Proof. If $h \vdash f$ then f is equivalent to $h \vee f$. The result is then immediate by writing h, e, f for x, y, z respectively in the theorem. ∎

This corollary holds in particular when $(h \vee e)f$ is equivalent to h itself. So whereas theorem 1 showed that (except in extremity) a factor f of h that is (maximally) *deductively* independent of $h \vee e$ is counterdependent on e (countersupported by e), the last corollary shows that the same is true if f is *probabilistically* independent of $h \vee e$. In the appendix to this paper we shall show how this result may be used to answer a challenge issued by Good (1984).

3. Discussion: logical content

Inductive inference, if it is anything, yields conclusions that go beyond or transcend the available evidence. As we noted at the outset of the paper, Peirce (1878) (CP 2.680), called such inferences 'amplifiative', though the more usual term these days (see, for instance, Salmon (1967, ch. 1)) is 'ampliative'. Inductive inferences amplify the evidence, one must suppose, and it is pertinent to ask how they do this; to ask, in particular, what it is that an inductive inference amplifies the evidence with. It seems that enthusiasts of inductive logic have rarely been prepared to answer this question. By failing to do so they have signally failed to explain why they think of the process of inductive inference as consisting of a constructive step that starts from the evidence and terminates with the conclusion.

More recent writers, especially those who label themselves Bayesians, have abandoned the view that there is any such process of inductive inference; as Levi (1986) notes, they 'want to preserve the appearance of induction without its substance'. Hypotheses, they admit, do not grow out of the evidence; they are inventions; and only when they have appeared do they call for assessment in probabilistic terms. But the question of how a hypothesis h amplifies the evidence e is not as easily set aside as this. As we saw in §1, there are many propositions, not deductively equivalent on their own to h, that in the presence of e are equivalent to h. The strongest of them, he, implies e, and at the other extreme there is $h \leftarrow e$, which is deductively quite independent of e. There is presumably something that all these different propositions have in common, something that explains why in conjunction with e they all become equivalent.

It is our thesis that this thing that they have in common is simply what they say in addition to e. That is why they say the same when conjoined with e. But what $h \leftarrow e$, which shares no consequences with e, says in addition to e is everything it says: the proposition $h \leftarrow e$, we venture to suggest, precisely encapsulates what each of these propositions says beyond e (a conclusion reached also by Hudson (1974)). The ampliative component of each of them is just the proposition $h \leftarrow e$.

In order to formulate this thesis in the crispest possible way we now introduce the familiar idea of the *logical content* of a proposition.

Let X be any set of propositions. We shall, following Tarski (1930), represent by $Cn(X)$ the set of all X's logical consequences, its *consequence class*. Where X is a unit set $\{x\}$ we normally write $Cn(x)$ instead of $Cn(\{x\})$; we shall refer to it as the *logical content*, or just the *content*, of the proposition x. Tarski calls any set X that is the consequence class of some set X a *deductive system*, and shows that the class of all deductive systems of a language has the structure of a lattice: the meet or product $\boldsymbol{XY}$ of two deductive systems is provided by their set-theoretical intersection; their join or sum $\boldsymbol{X} \dotplus \boldsymbol{Y}$ is provided in general not by their set-theoretical union $\boldsymbol{X} \cup \boldsymbol{Y}$ but by the set $Cn(\boldsymbol{X} \cup \boldsymbol{Y})$. The smallest system is the set $\boldsymbol{L}$ of all logical truths; the largest is the set $\boldsymbol{S}$ of all propositions. Because $\boldsymbol{L}$ is a subsystem of every system it is often treated as though it were empty; this is a device that can be made technically quite respectable by redefining $Cn(X)$ as the class of X's non-trivial consequences; it will be taken for granted for the rest of the paper. By identifying the proposition x with the system $Cn(\{x\})$, which for obvious reasons is called *finitely axiomatizable*, we may make logically equivalent propositions identical to each other. The class of all finitely axiomatizable systems is not only a lattice but also a Boolean algebra *dually* isomorphic to the algebra of propositions of the language. (It is dually isomorphic because the *product* of the contents $Cn(x)$ and $Cn(y)$ is $Cn(x \vee y)$, the content of the *sum* $x \vee y$; likewise, the sum of $Cn(x)$ and $Cn(y)$ is $Cn(xy)$.) In all but the most cruelly stunted languages, however, there will be systems that are not finitely axiomatizable, and the lattice of systems for such languages is not Boolean. For the moment we shall take account of finitely axiomatizable systems only, but shall return to the more general case later.

The problem of the ampliative character of induction may be put in terms of contents as follows:

What is the difference in content between the evidence e and the hypothesis h, and to what degree is this excess content supported by, or probabilistically dependent on, e?

Our answer is:

The difference in content between e and h is $Cn(h \leftarrow e)$, and this excess content is never positively dependent on e, usually being strongly counterdependent on e.

The explanation for this answer has already been given, but is worth repeating: $Cn(h \leftarrow e)$ is just what needs to be added to $Cn(e)$, without duplicating anything already there, in order to yield the system $Cn(h)$. In other words, $Cn(h \leftarrow e)$ is the unique content whose sum with $Cn(e)$ is the system $Cn(h)$ and whose product with $Cn(e)$ is the system $\boldsymbol{L}$.

There is an obvious objection to understanding the difference between $Cn(h)$ and $Cn(e)$ in this way, an objection that forms the backbone of Redhead (1985) and is sympathetically endorsed by Howson & Franklin (1985). The objection is that it is only in uninteresting circumstances that $Cn(h \leftarrow e)$ is identical with the set-theoretical difference $Cn(h) - Cn(e)$; only if e implies h, indeed, when each is empty (that is, is equal to $\boldsymbol{L}$). For all interesting h and e the set $Cn(h \vee e) \cup Cn(h \leftarrow e)$ is a proper subset of $Cn(h)$, so that there are usually many consequences of h that are to be found neither in $h \vee e$ nor in $h \leftarrow e$. For an example Redhead (p. 187) takes h as the kinetic theory of gases and e as Boyle's law, and notes that Graham's law of diffusion is a consequence of h that surely does not follow from either $h \vee e$ or $h \leftarrow e$. An even simpler example, one at the heart of the belief in induction, takes h as 'All ravens are black' and e as some conjunction of instantial statements of the form 'The ith observed raven was black'. It is clear that the prediction 'The next raven to be observed will be black', though a consequence of h, is in neither $Cn(h \vee e)$ nor $Cn(h \leftarrow e)$. (Jeffrey (1984) discusses an example that is entirely similar to this one, and so does Gaifman (1985, p. 19); Gillies (1986, p. 112), makes some use of this example to criticize σ as a measure of support.) In neither of these cases is there any cause to suppose that the prediction particularized is bound to be counterdependent on e. Redhead concludes, as does Jeffrey, that we have failed to show that the part of h that goes beyond e is always countersupported by e.

The facts are of course as stated, but the objection is devoid of merit. In each example let the prediction that is in neither $Cn(h \vee e)$ nor $Cn(h \leftarrow e)$ be called k for short. As $h \leftarrow e$ does not imply k, neither does e', so $k \vee e$ is not a logical truth; $Cn(k \vee e)$ is not empty. This is just to say that k shares content with e, that it is to some degree deductively dependent on e. It is only because of this, indeed, that it has the chance of being supported by e. Thus the objection misses its mark: it fails to show that positive probabilistic dependence can be achieved in the absence of some degree of deductive dependence. Indeed, absolutely nothing at all is shown by introducing the prediction k at this point; for h itself is a consequence of h that appears in neither $Cn(h \vee e)$ nor $Cn(h \leftarrow e)$, and it could easily replace k throughout the argument. There is a strong temptation, granted, to think that 'The next raven to be observed will be black', which (unlike 'All ravens are black') appears to be concerned only with the future, must be quite independent of any conjunction e of instantial statements all of which report occurrences that befell in the past. But in truth this prediction is far from being deductively independent of these instantial statements: a proposition like this in the future tense has a host of consequences that can be established by evidence formulated entirely in past tenses. This is one reason why it is misleading to formulate the problem of induction wholly in terms of the distinction between past and future. The crucial distinction is between what is settled by the evidence and what wholly transcends it. Inductive logic, as we understand its aims, proposes that there is some logical connection between the two.

The objection (also to be found in Good (1984, 1985*b*)) that the content of h that goes beyond e should be taken as the set-theoretical difference $Cn(h) - Cn(h \vee e)$, which is not a content at all, rather than $Cn(h \leftarrow e)$, which is, is a thoroughly unhealthy one. (The proposal it makes is reminiscent of the definition in Popper (1963, pp. 233f.), of the falsity content of a proposition as the difference between its content and its truth content. We do no more here than recall the unfitness of that definition.) To ignore in this way the directives of the calculus of contents is in effect to dismiss out of hand the problem of the ampliative character of inductive inference. To see that this is so, compare the case with the remarkably like case of the addition and subtraction of ordinal numbers under the von Neumann reading, where each ordinal is identical with the set of all its predecessors. What is the difference between $5 = \{0,1,2,3,4\}$ and $2 = \{0,1\}$? Although one answer is doubtless the set-theoretical difference $\{2,3,4\}$, it is patently not the only answer; and in an arithmetical context, it is not even a correct one. The correct answer is $3 = \{0,1,2\}$. The set $2 \cup 3$ is a distinctly proper subset of $2+3$. But to recognize that some arithmetical operations on ordinals differ from corresponding set-theoretical ones is not to challenge the construal of ordinals as sets. Likewise, we are not repudiating the definition of the content of h as the set of its consequences by insisting that in a context like the present one the difference between two contents is rarely equal to their set-theoretical difference. We are simply taking seriously the idea that there is something like ampliative content.

What separates Popper and Miller on the one hand, and Jeffrey, Good, Redhead, and others on the other, seems in the end to have quite a lot to do with where to draw the line between deductive *independence* and deductive *dependence*. It may be put succinctly in this manner: we say that f is *dependent* on e unless it is maximally *independent* of e; they say that f is *independent* of e unless it is maximally *dependent* on e. This is not just a verbal dispute, because we should be more than happy to restrict 'dependent' to the maximally dependent, provided that 'independent' is likewise restricted to the maximally independent. Any proposition f that is neither dependent on e nor independent of it could be called 'partly dependent' or (of course) 'partly independent'. Probabilistic dependence comes in degrees, and it is hard to understand why deductive dependence should not do so too. What we cannot accept is the view that h has no deductive dependence on e when it is not implied by e. A table is not wholly dependent on any one of its four legs; but that does not mean that it is independent of each of them.

We suspect that at this point some Bayesians, and perhaps other inductivists, may concede our thesis that probabilistic dependence is dependent on deductive dependence, but that they will insist nonetheless that their views on the possibility of inductive logic are not in any way damaged. We may be reminded that induction has often been thought of as a converse of deduction; that inductive inference has for long been epitomized as a kind of partial entailment (Salmon 1969). If probabilistic dependence turns out just to be a reflection of deductive dependence, so be it. Induction is, after all, nothing but the way the probability of the hypothesis h alters in the face of evidence. It cannot be denied that this happens (provided, of course, that $p(h) > 0$).

However, this concession, far from being innocuous, leaves the thesis of inductivism in tatters. Whereas any deductive inference is of course deductively correct, it cannot seriously be maintained that the converse of any deductive inference is inductively correct. Some effort has to be made to distinguish inductive inference from outright fallacy. But by conceding that there is nothing to probabilistic dependence but deductive dependence, the inductivist removes any way of making this distinction. Jeffrey (1984), after acknowledging that there are cases (involving predicates like 'grue' and 'bleen') where 'e supports h ($= ef$) only because it supports itself', concludes: 'But it is not always so.' Yet the only way in which the difference between the two contrasting evaluations can be explained on inductivist principles is by designating as the part of h that goes beyond e a proposition f (roughly: 'All unobserved ravens are black') that is deductively highly dependent on e. This is how Jeffrey explains it. What he does not explain is how such a proposition f can characterize the part of h that goes beyond e.

As we stressed in §1, for any h and e there is a multitude of propositions k that, conjoined with e, amplify it into a proposition equivalent to he. Where these various k differ is in how much $Cn(k)$ overlaps $Cn(e)$; in how much of e's content k repeats; in the extent of $Cn(k \vee e)$; in how dependent deductively k is on e. (These formulations all say the same.) What is common to the content of each k is $Cn(h \leftarrow e)$. What distinguishes h from rival amplifications of e is, one can only assume, this part of it that contains all the real novelty. But this part of h is just the part that in all circumstances is countersupported by the accumulated evidence e.

In connection with this disagreement it must be mentioned that, in addition to the standard syntactical definition of the content of h that we have here endorsed, there is an equally intuitive definition that is semantical: according to this second definition, the content $Ct(h)$ of a hypothesis h is the class of all models or structures in which h fails. (To keep the size of $Ct(h)$ within bounds it is customary to identify models that are elementarily equivalent: that is, models in which exactly the same sentences hold.) Under this definition, whose origins lie in the definition in Popper (1934) of the empirical content of h as the class of basic statements

h excludes (see Popper 1934, 1959, 1984, sections 31 and 35), the union of two contents is also a content: indeed because hk fails in a model if and only if one of h and k does, $Ct(h) \cup Ct(k) = Ct(hk)$. The class of contents actually forms a field of sets to which the original algebra of propositions is dually isomorphic. It follows, and is easily checked, that the set-theoretical difference between $Ct(h)$ and $Ct(h \vee e)$ is exactly $Ct(h \leftarrow e)$. This indicates clearly enough that there is nothing outrageous about the views we have been advancing about ampliative content. Ct and Cn capture the same properties of logical content. The only difference between them is that Ct manages with the standard set-theoretical operations, whereas Cn needs to be a bit more involved. The use of Ct, rather than Cn, incidentally, should appeal especially to those who embrace what is called the non-statement (or structuralist) view of scientific theories.

The function Ct also helps to clarify the appropriateness of the use of spatial terminology ('going beyond', 'overlapping', etc.) in the discussion of content. It can without difficulty be agreed that the parts of a (simple) table that go beyond its top, say, are just its four legs. But tables are poor analogues of contents in that their separate parts do not themselves have parts in common: they do not (spatially) overlap. As we have seen, contents are quite different; and a more appropriate spatial analogy is provided by something like the human body. How are we to characterize those parts of the human body that go beyond the chin (say)? The position of Redhead, and those others we have associated with him, would appear to encourage the suggestion that the (intact) head is one such part, on the grounds that it is not itself part of the chin. We, in contrast, would insist that only the chinless head, not the unmutilated head, can be thought of as properly going beyond the chin. In brief: we accept that the set-theoretical difference $x-y$ unambiguously represents the difference between the sets x and y only when they are both sets of *points*. In the corporeal example, points are just spatial points. In the case of propositions, however, they have to be taken as models; these are the *elements* of content in the sense that their unit sets are the smallest non-empty contents. (If $Ct(x)$ is a unit set then x has almost minimal content; for $Ct(x)$ is the set of models where x fails.) It must be realized that individual consequences cannot be taken as (equal) elements of content. For some of them have very much more content than do others. And only in quite extreme cases can contents be decomposed into conjunctions of sentences of minimal content, what Tarski (1936) calls *irreducible systems*. His theorem 37 shows that it is only in thoroughly finite languages that every system can be represented as the logical conjunction of irreducible systems.

At this point we may mention a much more sophisticated objection to the proposal that the proposition $h \leftarrow e$ represents precisely that part of the content of h that goes beyond $h \vee e$. It is that within Tarski's calculus of deductive systems the conditional $\boldsymbol{X} \leftarrow \boldsymbol{Y}$ is not universally definable. That is to say, the remainder operation $\boldsymbol{X} - \boldsymbol{Y}$ is not defined for every pair of unaxiomatizable deductive systems. We have to be a little careful about terminology here because the sum $\boldsymbol{X} \dotplus \boldsymbol{Y}$ of two systems or contents, being more extensive than either, corresponds to their logical *product*, rather than to their sum, whereas $\boldsymbol{XY}$, their set-theoretical intersection, must be written as a disjunction $\boldsymbol{X} \vee \boldsymbol{Y}$ in order to display its logical force. (As was noted above, the lattice of finitely axiomatizable systems is *dually* isomorphic to the algebra of propositions from which it springs.) Thus the system that would have the logical force of the conditional $\boldsymbol{X} \leftarrow \boldsymbol{Y}$ would be the remainder $\boldsymbol{X} - \boldsymbol{Y}$: together with $\boldsymbol{Y}$, it would just suffice for $\boldsymbol{X}$. But, as noted, this remainder may not exist once systems that are not finitely axiomatizable are included.[9] This certainly is something of a surprise when first encountered. There is, for instance, no system that precisely characterizes the difference in content between Zermelo–Fraenkel set theory $\boldsymbol{ZF}$

and Zermelo set theory $\boldsymbol{Z}$; strange, but true. In the present context, however, the difficulty is not unduly severe because the remainder $X - Y$ always exists when Y is finitely axiomatizable. Thus whenever $h \vdash e$, and in addition the evidence e can be enshrined in a finite set of propositions (it is not important whether h is finitely axiomatizable or not), we may be assured that there is a conditional system $Cn(h) \leftarrow Cn(e)$, namely the intersection $Cn(h) \vee Cn(e')$. The same applies in the more interesting case in which e is not a consequence of h, even when the common part $Cn(h) \vee Cn(e)$ fails to be axiomatizable. For here too the product of the systems $Cn(h)$ and $Cn(e')$ just suffices, in the company of $Cn(h) \vee Cn(e)$, to deliver $Cn(h)$. (To strip away some of the cumbersome notation: if e is finitely axiomatizable, it has a negation e'; and $h \vee e'$ is the required conditional.)

In the general case in which e is not finitely axiomatizable there may be no way of characterizing as a system the difference between h and $h \vee e$; no way, that is, of delineating the extent to which the hypothesis h goes beyond the evidence e. The fact remains that every consequence of h that does go wholly beyond e will be countersupported by e. There seems no escape from the conclusion that probabilistic support is not inductive.

None of this is to say, however, that there may not be some other measure of inductive dependence or inductive support, not too closely related to probabilistic support, according to which $h \vee e$ can be less well supported by e than h is, and $h \leftarrow e$ not countersupported. It is worth noting that neither of the measures c of degree of corroboration of Popper (1959, p. 400) qualifies here. For each of these measures we have $c(h \leftarrow e, e) \leqslant 0$. Given the interpretation of degree of corroboration as the degree to which the hypothesis has been subjected to severe and uncompromising tests, these results are far from alarming. It is obvious that $h \leftarrow e$, which shares no consequences with e, has therefore not been tested by it; in consequence, we really have no right to interpret c as degree of corroboration at all (see the addition, marked by a *, to note 8 on p. 402). It is also the case for both measures that $c(h \vee e, e) = s(h \vee e, e) = 1 - p(h \vee e)$. But it does not appear to be impossible that this should be less than $c(h, e)$.

Finally, we turn to measures of content. The simplest measure $ct(h)$ of the content of a hypothesis h, and one that has been widely adopted, is given by

$$ct(h) = 1 - p(h). \tag{3.1}$$

This has the attractive property that

$$ct(h) = ct(h \vee e) + ct(h \leftarrow e), \tag{3.2}$$

which agrees absolutely with our decomposition of h into these two factors. It follows as well from (3.1) that the term $s(h \vee e, e)$, which we have earlier suggested as a measure of the mutual deductive dependence between h and e, is equal simply to $ct(h \vee e)$, the content that they share. This too accords well with what we have said earlier: *the probabilistic support of h by e is dominated above by the measure of their common content.*

As Howson & Franklin (1985, p. 426) observe, an identity very closely related to the addition property (3.2) was stated by Hintikka (1968, p. 314, formula (8)). Being persuaded by Redhead's view that there is more to h's content than what may be extracted from $h \vee e$ and $h \leftarrow e$, Howson & Franklin see (3.2) as a reason to reject (3.1) as a measure of content; in particular, to reject $ct(h \leftarrow e)$ as a measure of how much h's content surpasses that of $h \vee e$. They propose (p. 429) that the (information) content of a hypothesis h be measured instead by $\mathrm{Inf}(h) = -\lg p(h)$. This as it stands is hardly objectionable (see Popper 1959, p. 402); $-\lg p(h)$

is, after all, a monotonic function of $ct(h)$, and could well be used in its stead provided appropriate adjustments are made elsewhere. Howson & Franklin are, of course, not minded to make any such adjustments. In particular they see Inf as a fundamental improvement on ct because in contrast to the addition property (3.2) we have in general the inequality

$$\mathrm{Inf}(h) \geqslant \mathrm{Inf}(h \vee e) + \mathrm{Inf}(h \leftarrow e), \tag{3.3}$$

with identity only in relatively special and rather uninteresting cases.

Howson & Franklin actually make it a requirement on any satisfactory measure of content that 'if the knowledge of a's truth should not alter one's state of belief concerning b's truth, then the content of, or the information supplied by, the conjunction ab is the sum of the contents of a and b'. (See p. 429 of their paper. We have used our own notation.) They interpret this condition to mean that a content measure ϕ should be additive on a and b when these propositions are probabilistically independent; and so interpreted it virtually compels ϕ to be the function Inf defined above. Howson & Franklin go on immediately to propose that when $h \vdash e$ the excess content of h over e is to be measured by $-\lg p(h) + \lg p(e)$, which equals $-\lg p(h, e)$. More generally, we may suppose, the excess content of h over e would be measured by the difference of $-\lg p(h)$ and $-\lg p(h \vee e)$, which is equal to $-\lg p(h, h \vee e)$. They accordingly conclude that 'the excess content of h relative to e is *not*...represented by any truth-function of h and e'. Indeed they assert, apparently with satisfaction, that the excess content of h over e so measured is not represented by any proposition or deductive system at all.[10]

The measure Inf is just a very simplified version of a measure of information championed by Good in several publications (see, for example, Good 1977, section 2; 1983, pp. 220f.). In Good (1984) (discussed in our appendix) it is suggested that relative to evidence e the appropriate decomposition of the content of a hypothesis h is into an evidential part $h \vee e$ and a part that is probabilistically independent of e; in Howson's and Franklin's words, that evidence 'should not alter one's state of belief' in the content of h that transcends that evidence. It is surely astonishing to find inductivists voicing such a non-inductivist sentiment, however mutedly.

The thought that there is no such thing as probabilistic induction – even, perhaps, no such thing as induction – may distress some of our readers. But it must be remembered that there is nothing in this world that can be settled beyond all doubt. Thus the duty to go on testing a hypothesis severely, with the use of all our imagination and intelligence in our search for loopholes, is not discharged even if the hypothesis has been established by some process called induction. (Aspirin has been tested countless millions of times, yet there may be a significant, not yet suspected, side effect.) As we have shown, no amount of testing can provide a hypothesis with anything like inductive support. And no amount of probabilistic support, or even inductive support, can release us from our responsibility to impose further tests.

Appendix

In this appendix we shall do our best to respond to the challenge issued by Good (1984) in the following words:

> ...when Popper and Miller claim to prove the impossibility of inductive probability this argument needs to be watertight if it is to be believed. Their argument falls short of achieving the potentially watertight (but unachievable) form...[that may be

> stated thus]. Let h be any hypothesis and e any event such that $p(e) \neq 1$ and $p(h,e) \neq 1$, and suppose that h can always be written in the form $h = ab$ where (1) e deductively implies b, (2) e probabilistically undermines a, that is $p(a,e) < p(a)$, and (3) a and b are probabilistically independent. All this is to be true even if $p(h,e) > p(h)$. [It is clearly taken for granted that $p(e) \neq 0$, so $p(h) \neq 0$ too.] This situation would I think be 'completely devastating to the inductive interpretation of the calculus of probability', to quote Popper and Miller. Popper and Miller in effect define a as $h \leftarrow e$ and b as $h \vee e$. Then all conditions are satisfied, except that their condition (3) is replaced by the weaker condition: (3a) a and b are probabilistically independent when e is given.... Because I believe that inductive probability cannot be refuted I predict that Popper and Miller will not be able to redefine a and b to satisfy the conditions [(1), (2), (3)].... Thus, in accordance with Popper's standard and justifiable requirements, I have stuck my neck out and have handed him an axe.

At the outset we must make plain that, interesting as Good's challenge is, it arises from a misunderstanding of our thesis in Popper & Miller (1983). (Good acknowledged this later in Good (1984), and more decisively in Good (1985*b*).) At no stage do we suggest that h be decomposed into probabilistically independent factors; our point, on the contrary, is that *any factor of h that is deductively independent of e will be counterdependent on e.*

The challenge, given an arbitrary hypothesis h, is to find factors a, b that (G0) are jointly equivalent to h and (G3) probabilistically independent of each other, (G1) b being a consequence of the evidence e and (G2) a being counterdepent on e. We shall show that except in some extreme cases it is always possible to write h in the form $(h \vee e)f$, where f is both independent of $h \vee e$ and counterdependent on e. Indeed, the proposition f can be constructed in a uniform manner from the propositions h and e.

All that is needed for the construction of f is the availability for any r in [0, 1] of a proposition g that has probability r and is probabilistically independent of the other propositions we are concerned with. If $r = m/n$ the proposition 'The first ball drawn from an urn with m white balls and $n-m$ black balls will be white' (or some proposition of similar structure) will suffice for g. If r is irrational, a suitable g may be constructed with the help of a random variable ξ that is uniformly distributed on [0, 1], g being the proposition '$\xi < r$'. Any set S of propositions rich enough to contain even a fraction of the propositions of current physical science will allow the construction of such an f (which may, if necessary, describe a remote event in the past or future). We shall call any such set S an *informed* set of propositions. That the set of propositions under consideration be an informed set is more than enough for our purposes.[11]

THEOREM 7. *Let the set S to which h, e, and their logical combinations belong be an informed set. Then there is a proposition f satisfying*

$$h \text{ is equivalent to } (h \vee e)f, \tag{G0}$$

$$p(h) = p(h \vee e)\,p(f). \tag{G3}$$

Moreover, provided that

(i) $$p(h) \neq 0 \neq p(e),$$

(ii) $$p(h,e) \neq 1 \neq p(e,h),$$

every such proposition f satisfies Good's second condition

$$s(f,e) < 0 \tag{G2}$$

(*whilst* (G1), *that* $e \vdash h \vee e$, *is trivial*).

Proof. Because S is an informed set we can find a proposition g that is probabilistically independent of $h \vee e$ and itself has probability equal to $p(h, h \vee e)$. Thanks to condition (i) it follows easily that g and $(h \vee e)'$, that is $h'e'$, are probabilistically independent.

We now take for f the proposition $h \vee e'g$. It is trivial that (G0) is satisfied.

To prove (G3), note that by the addition law

$$\begin{aligned} p(f) &= p(h)+p(h'e'g) \\ &= p(h)+p(h'e')\,p(g) \\ &= p(h)+[1-p(h \vee e)]\,p(h, h \vee e) \\ &= p(h)+p(h, h \vee e)-p(h) \\ &= p(h, h \vee e). \end{aligned}$$

Thus $p(h) = p(h \vee e)\,p(f)$ as required.

To prove (G2) we apply the corollary to theorem 6. By (i) and (ii), $p(h) \neq 0$ and $p(e,h) \neq 1$; so $p(he) \neq p(h)$, whence by the addition law $p(h \vee e) \neq p(e)$. By (i) again $p(e) \neq 0$, so condition (b) of the corollary is satisfied. By (ii) again $p(h,e) \neq 1$, so by the addition law $p(h \vee e) \neq p(h)$. Because $p(h \vee e) \neq 0$, it follows from (G3) that $p(f) \neq 1$; so condition (c) holds as well. By the corollary, $s(f,e) < 0$. ∎

Conditions (i) and (ii) thus ensure that $h \vee e$ may be complemented by an independent factor f, and that every such independent factor f will be countersupported by the evidence e. To ensure only that there is such an f that is not supported by the evidence, that is, that $s(f,e) \leqslant 0$, it is enough to require that $p(e) \neq 0$. This follows at once from theorem 6. Thus even in the extreme case there is no support.

It must be noted, however, that (i) and (ii) together are rather stronger than the conditions envisaged by Good, in that we have required that $p(e,h) \neq 1$. (Good's condition that $p(e) \neq 1$ follows from (i) and (ii).) This is an inequality that is doomed to fail whenever the hypothesis h implies the evidence e; and trivially if such an implication holds we cannot have $h \vee e$ and f probabilistically independent and also $s(f,e)$ negative. The question arises whether it might be possible to replace $h \vee e$ by some weaker consequence of e; that is, to write h as the conjunction of probabilistically independent factors g and f where e implies g, $p(g)$ is strictly greater than $p(e)$, and f is counterdependent on e. (This would suffice to answer Good's challenge in this case.) Interestingly enough, this turns out not to be possible. For we would have $(h \vee e)f$ equivalent to h, and $(h \vee e) \vee f$ equivalent to $e \vee f$, so

$$\begin{aligned} p(h) &= p(h \vee e)+p(f)-p((h \vee e) \vee f) \\ &= p(e)+p(f)-p(e \vee f) \\ &= p(ef). \end{aligned}$$

Thus $$p(f)\,p(e) < p(f)\,p(g) = p(h) = p(f,e)\,p(e) = p(fe),$$

and f turns out to be positively dependent on e.

Before those who wish to defend probabilistic inductive logic take too much satisfaction in this result it should be pointed out that for no genuine hypothesis h could we expect to have $p(e,h) = 1$. In particular, no hypothesis h ever gets near to implying all the evidence e. We cannot help observing that Good (1967) is one of the most celebrated defences of *the principle of total evidence*, the principle that in the calculation of probability values all the available

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

evidence should be used. If this principle is adhered to, there can be no denying that the case excluded by $p(e,h) \neq 1$ is as extreme as the other components of (i) and (ii).

Our stipulation that the set S be an informed one seems equally harmless. True enough, in a countable language most real numbers do not have names, which means that for most r there will not exist any proposition g of the form '$\xi < r$'. But this scarcely matters, as the existence of such a proposition g is needed only when r is a value adopted by the function p; indeed, equals $p(h, h \vee e)$. In a countable language there can be only countably many such values. Thus we easily avoid the spectre of uncountably many different values of p.

But though when S is thought of as a genuine scientific language the condition of informedness, so understood, is unimpugnable, it is of some interest, from a formal and even from an intuitive standpoint, that it can be considerably weakened; in particular that we can eliminate from it all reference to probabilistic independence. Let it be agreed that r is a real number in the range of p. Then instead of requiring that for each proposition y there be a proposition g in S, *independent* of y, for which $p(g) = r$, we need require only that

if $y \vdash x$ and $p(y) \leqslant r \leqslant p(x)$ then
there exists g in S such that $y \vdash g \vdash x$ and $p(g) = r$.

We shall call a set S of propositions *dense with respect to the measure* p when this condition holds. It would indeed be enough if x were limited to being a tautology. (This is left as an exercise.)

THEOREM 8. *Let S be dense with respect to p. Then there is a proposition f satisfying*

$$h \text{ is equivalent to } (h \vee e) f, \tag{G0}$$

$$p(h) = p(h \vee e)\, p(f). \tag{G3}$$

Moreover, provided that

(i) $$p(h) \neq 0 \neq p(e)$$

(ii) $$p(h,e) \neq 1 \neq p(e,h)$$

every such proposition f satisfies Good's second condition

$$s(f,e) < 0. \tag{G2}$$

Proof. The proposition h implies both $h \vee e$ and $h \leftarrow e$. Thus $p(h)$ is no greater than $p(h, h \vee e)$, which, by theorem 3, is itself no greater than $p(h \leftarrow e)$. We thus have both

$$h \vdash h \leftarrow e,$$

$$p(h) \leqslant p(h, h \vee e) \leqslant p(h \leftarrow e).$$

By the density condition, there is therefore a proposition f such that

$$h \vdash f \vdash h \leftarrow e \quad \text{and} \quad p(f) = p(h, h \vee e).$$

Because $(h \vee e) f$ implies $(h \vee e)(h \leftarrow e)$, it implies h, and so is equivalent to h. The probabilistic independence of $h \vee e$ and f is immediate. As before, (G3) follows from the corollary to theorem 6. ■

We are indebted to Dr D. A. Gillies, Professor I. J. Good, Dr P. Schroeder-Heister, Mr C. Howson, Professor I. Levi, Dr A. Mura, Professor M. Pera, Professor M. L. G. Redhead and Professor R. G. Swinburne for their comments on earlier versions of this paper.

Notes

[1] It seems to be the thesis of Wassermann (1985) that our proof of the impossibility of inductive probability is 'doomed from the start' because one of its assumptions is that logical probabilities are 'numerically well defined or definable' (pp. 130f.). Because, in his words, 'no method is cited for attributing numerical values to $p(e, hb)$', Wassermann goes so far as to assert, in direct contradiction to (1.5), that the term $p(e, hb)$ is 'vacuous' even when e is deducible from h. But though the assumption that numerical values can be assigned to inductive probabilities may be made by some defenders of probabilistic induction, it is not made by us. We shall therefore not pursue this point here.

[2] This relation of deductive independence is what Sheffer (1921, pp. 32f.) called *maximal independence*. It should not be confused with the somewhat misleadingly named relation of *complete independence*, the relation (whose characterization can be traced back at least as far as Moore (1910, p. 82) that is postulated to hold between the atomic propositions of Wittgenstein (1921): a set of propositions is completely independent if any of them can be true and the remainder false simultaneously. (Maximal and complete independence are mentioned by Tarski (1956, p. 35); for more information on these relations see Miller (1974, pp. 185–188).) Note that propositions that are completely independent of each other cannot in general be probabilistically independent of each other. To see this, let $\{x, y, z\}$ be a completely independent set of propositions with probabilities strictly between zero and one. It is easily checked that $\{x, yz\}$ and $\{xy, xz\}$ are also completely independent sets. If completely independent sets were always probabilistically independent, we should have $p(x)p(y)p(z) = p(x)p(yz) = p(x(yz)) = p((xy)(xz)) = p(xy)p(xz) = p(x)p(y)p(x)p(z)$, which is impossible unless one of x, y, z has probability zero or x has probability one.

In an earlier version of this paper it was asserted that there is no obvious way in which the relation of complete independence admits of degrees. But Dr Alberto Mura has kindly pointed out to us that this is incorrect. By a theorem of de Finetti (1974, p. 112), called by him 'the fundamental theorem of probability', when the probability of a proposition x is not determined by the probabilities of the members of the set X, it is anyway restricted to a closed subinterval of $[0, 1]$. A. Mura (unpublished results) suggests that the length of this subinterval is a suitable measure of the degree of (complete) independence of x on X; when x is completely independent of X the interval has length one; when the truth value of x is determined by the truth values of the members of X the interval has length zero; in all other cases, it may be somewhere in between.

[3] Those who put their trust in Mill's methods of induction, especially the method of difference, will presumably be sympathetic to this move as a way of showing that probabilistic dependence should be ascribed wholly to deductive connections. We ourselves prefer to look on the argument as a test of the imputation, but not a confirmation of it. Note that because the ratio $p(h, eb)/p(k, eb)$ equals the ratio $p(h, b)/p(k, b)$ whenever both hb and kb imply e, these ratios equal also $d(h, e, b)/d(k, e, b)$; they equal $s(h, e, b)/s(k, e, b)$ too, where s is the measure of probabilistic dependence introduced in the text below. In other words, in these circumstances the relative probabilistic dependence of h and k on e (however measured) is fixed once their initial probabilities are. If $k \vdash h$ as well, this implies that $d(k, e, b) = p(k, hb)\, d(h, e, b)$, and a similar identity for s in place of d. What this all amounts to is just this: as hypothesis h goes more and more beyond the evidence, its probabilistic dependence on the evidence (measured by either d or s) decreases in direct proportion. Rather disarmingly, Howson & Franklin (1985), writing from what they say is a Bayesian viewpoint, take satisfaction in this result, suggesting that it shows that the value of $s(hh', e, b)$ depends on 'the structure of h'', and accordingly escapes what they call the tacking paradox (p. 430). But all the result as they state it really shows is that evidence e fails to point far beyond itself. What it quite clearly does not show is that it points in any particular direction beyond itself.

[4] The arrow here has a perfectly familiar meaning, even if it may have a slightly unfamiliar sense: it is simply the material conditional of elementary logic. The commentary Blandino (1984), which starts with 'a few considerations on the expression $(h \leftarrow e)$, when it is used to indicate that the hypothesis h, not directly evident, is made inductively probable by the occurrence of evidence e', and continues with this construal throughout, accordingly fails to come into proximity, let alone contact, with our position. A measure of the extent to which Blandino misunderstands us is provided by his stigmatizing as one of the defects of our letter to *Nature* 'the admission that $p(h \leftarrow e)$ is equivalent to $p(h, e)$' (p. 193). It is virtually our main result that these two quantities are in general different. See theorem 3 of §2.

[5] Provided that b is a consistent element of S all theorems of probability theory that hold unrelativized hold also when relativized to b. In most cases below there will be other conditions imposed that ensure that b must be consistent. Where there are not, the theorems cited will hold even for inconsistent b.

[6] Because h is equivalent to $he \vee he'$ we may show with even less difficulty what might be called the dual of (1.14):

$$s(h, e) = s(he, e) + s(he', e).$$

This has been pointed out by Peter Schroeder-Heister, and also by Dunn & Hellman (1986) (in a note originally communicated to us by Noretta Koertge). Although the first summand here is never negative, the second never positive, we feel no inclination to follow Dunn & Hellman in the suggestion that $s(he, e)$ could be taken to measure the non-deductive dependence of h on e and $s(he', e)$ to measure deductive counterdependence. A similar suggestion seems to be at the heart of the curious arithmetical juggling of Wise & Landsberg (1985*a*). See our reply (Popper & Miller 1985), and their response (Wise & Landsberg 1985*b*).

The grounds for this odd judgement seem to be the odder one that $s(x, y)$ should be described as deductive if its sign is independent of the measure p. Dunn & Hellman go so far as to conclude, more provocatively than persuasively, that '*all countersupport is "purely deductive"*'. It appears to be their view that there is nothing to choose between this position and ours, and so neither can be correct. But unlike $h \vee e$ and $h \leftarrow e$, neither he nor he' is a part of the content of h. Our thesis that e countersupports any part of h that is deductively independent of e is therefore not impugned.

[7] An example in which e is not implied by k is provided by the case of two dice, where h is 'At least one four is thrown', k is 'At least one five is thrown' and e is 'A total of eleven or twelve is thrown'. It is easily checked that $p(h \vee k) = \frac{5}{9} \leqslant \frac{2}{3} = p(h \vee k, e)$.

[8] The suitability of the function σ as a measure of support is criticized by Gillies (1986), in the passage following formula (3), and endorsed by Good (1986). There are, of course, measures of probabilistic dependence for which the analogue of corollary 2 of theorem 1 fails. It is interesting that the next most obvious choice after s and σ, namely $\Sigma(h, e) = s(h, e)/\sigma(h, e)$, is not one of them. Nor indeed is the measure of weight of evidence $w(h : e) = \lg(p(e, h)/p(e, h'))$ defined in Good (1985*a*, p. 251 and elsewhere). On the other hand, the analogue of (1.13) and (2.2) holds for all measures that increase when $p(h, e)$ increases and decrease when $p(h)$ increases. This is simply because the condition $he \vdash x \vdash h \leftarrow e$ is necessary and sufficient for he and xe to be equivalent; it is sufficient, therefore, to ensure that $p(h, e)$ and $p(x, e)$ be equal.

[9] To take a very simple example, suppose that $\boldsymbol{T}$ is a complete non-axiomatizable system; it might perhaps be the class of all true sentences of the language being considered. Then it is easily seen that there is no system $\boldsymbol{X}$ just adequate to yield the contradictory system $\boldsymbol{S}$. Any $\boldsymbol{X}$ at all that contains a false consequence will satisfy $\boldsymbol{T} \dotplus \boldsymbol{X} = \boldsymbol{S}$, but every such $\boldsymbol{X}$, indeed every false proposition x, will share non-trivial consequences with $\boldsymbol{T}$. For if $\boldsymbol{T}$ and x had only logical truths as common consequences, then for every true y the disjunction $x \vee y$ would be a logical truth. But then x' would imply y, whatever true proposition y was. Because x was assumed to be false, we should in this way have shown that $Cn(\{x'\}) = \boldsymbol{T}$, contrary to the hypothesis that $\boldsymbol{T}$ is not finitely axiomatizable. See also Miller (1978, section 3).

[10] They call 'h, e' an incomplete symbol, having meaning only in the context '$p(h, e)$'. (Again we use our own notation.) It does not follow from this that there is no proposition x such that $p(h, e) = p(x)$. Indeed, in the Appendix we show that under fairly weak conditions such a proposition x will exist. But it needs to be noted that Lewis (1976) has proved that in general x varies with the background knowledge or other assumptions b: in general there is no x such that $p(x, b) = p(h, eb)$ for all b. That is to say, the conditional probability is not open to uniform transformation into the absolute probability of a conditional (or of anything else). To this extent what Howson & Franklin say about the propositional formulation of their excess content is correct.

Recent discussion of the probabilistic representation of conditionals has been marked by its almost total disregard for Copeland (1950). This paper presents axioms governing two new operations in a Boolean algebra: the cross product $x \times y$ and the conditional $x < y$. It is shown that in algebras that obey the axioms the identity $(x < y) < z = x < (y \times z)$ holds, and that the probabilistic equation $p(x, y) = p(x < y)$ is also valid. Thus we shall have $p(h < e, b) = p(h, e \times b)$ for every b. None of this vitiates Lewis's results, of course, but it does show that there may exist conditionals whose probabilities are closely related to what are usually called conditional probabilities.

In another context, incidentally, Howson (1984, p. 246) makes essential use of the difference $K - \{e\}$ between background information K and a proposition e contained within it. He even says that '$\{e\}$ will be taken to include everything in K dependent on e'. When K is finitely axiomatizable, by k perhaps, this presumably contains k. If the difference $K - \{e\}$ so understood is to be something to which probabilities may be relativized, as Howson demands, it is not easy to see how it can avoid being identified with $k \leftarrow e$ (or more generally $K \leftarrow e$).

[11] Our proof is based on the demonstration that every proposition (not just 'six is thrown with each of two dice', but also 'six is thrown with one die' and even 'Snow is white') can be expressed as the product of probabilistically independent factors. We record our gratitude to Dr A. O'Hagan for showing us the proof of this apparently unknown result.

References

Blandino, G. 1984 Critical remarks on an argumentation by K. Popper and D. Miller. Discussion about induction. *Epistemologia* **7**, 183–206.

Copeland, A. H. Sr 1950 Implicative Boolean algebra. *Math. Z.* **53**, 285–290.

De Finetti, B. 1974 *Theory of probability*, vol. 1. London: John Wiley & Sons. (A translation of *Teoria delle probabilità*, Giulio Einaudi editors s.p.a., Torino (1970).)

Dunn, J. M. & Hellman, G. 1986 Dualling: a critique of an argument of Popper and Miller. *Br. J. Phil. Sci.* **37**, 220–223.

Gaifman, H. 1985 On inductive support and some recent tricks. *Erkenntnis* **22**, 5–21.

Gillies, D. A. 1986 In defense of the Popper–Miller argument. *Philosophy Sci.* **53**, 110–113.

Good, I. J. 1967 On the principle of total evidence. *Br. J. Phil. Sci.* **17**, 319–321. (Reprinted as ch. 17 of Good (1983).)

Good, I. J. 1977 Explicativity: a mathematical theory of explanation with statistical applications. *Proc. R. Soc. Lond.* A **354**, 303–330. (Reprinted as ch. 23 of Good (1983).)

Good, I. J. 1983 *Good thinking: the foundations of probability and its applications.* Minneapolis: University of Minnesota Press.
Good, I. J. 1984 The impossibility of inductive probability. *Nature, Lond.* **310**, 434.
Good, I. J. 1985*a* Weight of evidence: a brief survey. *Bayesian statistics* (ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley & A. F. M. Smith), vol. 2, pp. 249–264. Amsterdam: North-Holland.
Good, I. J. 1985*b* Probabilistic induction is inevitable. *J. statist. Computat. Simulat.* **20**, C216, 323f.
Good, I. J. 1987 A reinstatement, in response to Gillies, of Redhead's argument in support of induction. *Philosophy Sci.* (In the press.)
Hintikka, K. J. J. 1968 The varieties of information and scientific explanation. *Logic, methodology and philosophy of science III* (ed. von Rootselaar, B. & Staal, J. F.), pp. 311–331. Amsterdam: North-Holland.
Howson, C. & Franklin, A. 1985 A Bayesian analysis of excess content and the localisation of support. *Br. J. Phil. Sci.* **36**, 425–431.
Howson, C. 1984 Bayesianism and support by novel facts. *Br. J. Phil. Sci.* **35**, 245–251.
Hudson, J. L. 1974 Logical subtraction. *Analysis* **35**, 130–135.
Jeffrey, R. C. 1984 The impossibility of inductive probability. *Nature, Lond.* **310**, 433.
Levi, I. 1984 The impossibility of inductive probability. *Nature, Lond.* **310**, 433.
Levi, I. 1986 Probabilistic pettifoggery. *Erkenntnis* **25**, 133–140.
Lewis, D. K. 1976 Probabilities of conditionals and conditional probabilities. *Phil. Rev.* **85**, 297–315.
Miller, D. W. 1974 On the comparison of false theories by their bases. *Br. J. Phil. Sci.* **25**, 178–188.
Miller, D. W. 1978 On distance from the truth as a true distance. *Essays on mathematical and philosophical logic* (ed. Hintikka, K. J. J., Niiniluoto, I. & Saarinen, E.), pp. 415–435. Dordrecht: D. Reidel.
Moore, E. H. 1910 Introduction to a form of general analysis. *The New Haven mathematical colloquium*, pp. 1–150. Yale University Press.
Peirce, C. S. S. 1878 The probability of induction. *The Popular Science Monthly* **12**, 705–718. (Reprinted in Hartshorne, C. & Weiss, P. (eds), 1931–5. *The collected papers of Charles Sanders Peirce*, paragraphs CP 2.669–2.693. Harvard University Press.)
Pera, M. 1983 La miseria dell'induzione. *Scienza duemila* **5** (7/8), 18–20.
Popper, K. R. 1934 *Logik der Forschung*. Vienna: Springer.
Popper, K. R. 1959 *The logic of scientific discovery*. London: Hutchinson.
Popper, K. R. 1963 *Conjectures and refutations*. London: Routledge & Kegan Paul.
Popper, K. R. 1983 *Realism and the aim of science. The postscript* (ed. Bartley, W. W. III), vol. 1. London: Hutchinson.
Popper, K. R. 1984 *Logik der Forschung*, 8th edn. Tübingen: J. C. B. Mohr (Paul Siebeck).
Popper, K. R. 1985 The non-existence of probabilistic inductive support. *Foundations of logic and linguistics* (ed. Dorn, G. & Weingartner, P.), pp. 303–318. New York: Plenum.
Popper, K. R. & Miller, D. W. 1983 A proof of the impossibility of inductive probability. *Nature, Lond.* **302**, 687f.
Popper, K. R. & Miller, D. W. 1984 Reply to Levi (1984), Jeffrey (1984) and Good (1984). *Nature, Lond.* **310**, 434.
Popper, K. R. & Miller, D. W. 1985 Reply to Wise & Landsberg (1985*a*). *Nature, Lond.* **315**, 461.
Redhead, M. L. G. 1985 On the impossibility of inductive probability. *Br. J. Phil. Sci.* **36**, 185–191.
Salmon, W. C. 1967 *The foundations of scientific inference*. University of Pittsburgh Press.
Salmon, W. C. 1969 Partial entailment as a basis for inductive logic. *Essays in honor of Carl G. Hempel* (ed. Rescher, N.), pp. 47–82. Dordrecht: D. Reidel.
Sheffer, H. M. 1921 The general theory of notational relativity. (Mimeographed paper.)
Tarski, A. 1930 Über einige fundamentale Begriffe der Metamathematik. *C. r. Séances Soc. Sci. Lett. Varsovie* (cl. iii) **23**, 22–29. (Transl. Woodger, J. H. as ch. III of Tarski (1956).)
Tarski, A. 1936 Grundzüge des Systemenkalkül, Zweiter Teil. *Fundam. Math.* **26**, 283–301. (Transl. Woodger, J. H. as §§4–5 and the appendix of ch. XII of Tarski (1956).)
Tarski, A. 1956 *Logic, semantics, metamathematics*. Oxford University Press.
Wassermann, G. D. 1985 On the nature of inductive probabilities. *Methodology and science* **18**, 128–139.
Wise, J. & Landsberg, P. T. 1985*a* Has inductive probability been proved impossible? *Nature, Lond.* **315**, 461.
Wise, J. & Landsberg, P. T. 1985*b* On the impossibility of inductive probability. *Nature, Lond.* **316**, 22.
Wittgenstein, L. 1921 *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.

Additional reference

Hamilton, W. 1852 Discussions on philosophy and literature, education and university reform, chiefly from the Edinburgh review. Edinburgh: Blackwoods and London: Longman, Brown, Green and Longmans.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY A
PHILOSOPHICAL TRANSACTIONS OF