can be held to have secured very great and novel gains is thus a mark of the thoroughness with which critical philosophy forgets its beginnings, and of the need for the kind of philosophical self-reflection attempted here.

University of Durham

MACKIE'S DEFENCE OF INDUCTION

By P. J. R. MILLICAN

I

HUME's attack on induction is very familiar: *that the past be a rule for the future* cannot be established deductively, whilst any inductive argument to that conclusion would itself depend on it, and so be circular. But Hume takes for granted that there are no other kinds of reasoning available besides deduction and induction, and therefore assumes that the possibilities for justification of the latter are exhausted, since neither of them is adequate to the task. If his rigid dichotomy is accepted, Hume's argument indeed seems compelling. Recently, however, it has been challenged by those who see a third possibility, and I would like to consider, as one representative of this approach, J. L. Mackie's attempt to provide a *probabilistic* justification of induction.[1]

Mackie sets himself the task of justifying a prediction of *general* uniformity for a limited period, a prediction which can then itself be used to support other, particular inductions. He considers first the simple pair of rival hypotheses, that the world's ways of working are completely uniform throughout, and that they are completely random. 'If we had just these alternatives to choose between, it would be reasonable to prefer the former in the light of our observations, unless it was antecedently almost infinitely less probable than the second.' And to assume that it was so vastly improbable would, of course, be question-begging.

Unfortunately, the matter is complicated by the profusion of other possible hypotheses, even if we accept it to be overwhelmingly probable that the world has in fact been completely uniform during the period of our past observations. For it could be that this uniformity has a limited temporal (or spatial) range, beyond which it terminates or gradually fades out. These hypotheses of extensive

[1] J. L. Mackie, 'A Defence of Induction', in G. F. Macdonald (ed.), *Perception and Identity* (London and Basingstoke: Macmillan, 1979).

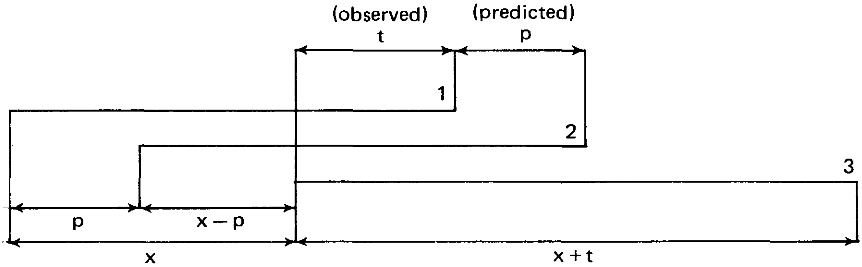
but limited order have yet to be dealt with, and Mackie's central argument is directed to this purpose:

> First, let us consider the indefinitely large set of hypotheses all of which assign the same spatio-temporal extent to uniformity, but locate it differently .... It seems a legitimate application of the principle of indifference to assign equal antecedent probabilities to the various hypotheses of this set. Some of them, however, will have been ruled out by the already-observed spread of uniformity.... Then, of the hypotheses not so ruled out, if the extent which is common to the hypotheses of the set ... is considerably greater than the observed spread of uniformity, relatively few will say that uniformity will terminate either at once or very soon. So, ... we can conclude that, even if uniformity lasts only for this limited extent, it is not likely to end very soon. But, secondly, let us consider the set of sets of such hypotheses .... Clearly, the shorter the extent characteristic of each set of hypotheses, the smaller the proportion of the hypotheses of that set that will cover the observed spread of uniformity.... Consequently, by an inverse probability argument, the observation of a certain spread of uniformity raises the probability that the extent of uniformity is considerably greater than that spread much more than it raises the probability that that extent is equal to or only a little greater than that spread. Now, ... we can appeal to a ... principle of tolerance to justify our not giving a zero antecedent probability to all hypotheses assigning more than a certain extent to uniformity. So long as the greater-extent hypotheses are not ruled out by such an unfair initial assignment, they can come out more probable in the end, their probability being raised more by the observation of some considerable spread of uniformity. And, as we have seen, once we have confirmed such a greater-extent hypothesis, we can go on by a direct probablity argument to infer that the uniformity is not likely to end very soon.

All this seems very plausible, until we examine it mathematically.

Let us assume that one of Mackie's hypotheses of limited order is correct, and that we have already observed uniformity for a length of time t, predicting that it will continue for a further time p. Now, if P(x) is the probability density function representing the initial probability that uniformity will extend over a total time x, then since, by the principle of indifference, all hypotheses assigning an equal extent to uniformity are equally probable, it follows that the probability that uniformity is of duration x and begins at some point within a certain interval of time i is proportional both to P(x) and to i, say kiP(x), where k is some constant. With this information, we want to find the probability that uniformity will continue for at least a further time p, fulfilling our prediction to that effect.

Consider the set of hypotheses whose characteristic extent $(x+t)$ exceeds the observed spread of uniformity t by a time x $(x > p)$:



Lines 1, 2 and 3 represent three possible uniformities of duration $(x+t)$, all of which account for the observed uniformity t. 1 ends immediately, 2 continues for a further time p, while 3, which began at the time of our first observation, finishes after time x. There is, then, a range of possible uniformities of the given duration $(x+t)$ which are consistent with our observations: they all start at different points within the marked period x, and of these, the uniformities starting within the marked period $(x-p)$ will yield successful predictions. As Mackie says, the proportion $(x-p)/x$ of successful predictions will grow considerably as x is increased. In addition – Mackie's second point – a greater value of x will bring a greater range x of possible (starting points for) uniformities, and we can therefore use an inverse probability argument to confirm those hypotheses involving a large value of x, at the expense of those involving a smaller value.

For any *particular* value of x, the initial probability of a uniformity of extent $(x+t)$ which will cover the observed spread of uniformity t is $kxP(x+t)$. So the initial probability of a uniformity of *any* extent which will cover this spread is equal to the sum of the function $kxP(x+t)$ for *all possible values of x* from zero to infinity (let us write this as '$SUM_0(kxP(x+t))$', indicating by the '0' the value of the lower limit of x). In this sum, the multiplication by x will give greater weight to $kP(x+t)$ for larger values of x, and this is the basis of Mackie's inverse probability argument. But it should be noted that this weighting is quite independent of t – no matter how much uniformity is observed, the probability that total uniformity will last for a period which exceeds observed uniformity by an excess time x is raised, on account of that observation, in proportion to x alone.

Let us now move on from Mackie's inverse probability argument to his direct probability argument. As we have seen, if total uniformity exceeds observed uniformity by a time x, then the probability of a successful prediction to the effect that uniformity will continue for a time $p(x>p)$ is $(x-p)/x$. But again, this value is completely independent of t – the probability in no way improves as our observation of uniformity continues. So neither part of Mackie's argument can justify at all taking the past as a rule for the future, for if the past is to be a rule for the future, then, at the very least, a greater observed regularity must give greater strength to a constant prediction. And his argument makes no mention of the extent of the observed regularity: 't' appears only in the argument-place of the function P(x), which Mackie does not discuss.

If induction is to be justified by some method such as Mackie's, this must depend entirely on the behaviour of the function P(x). We can establish the condition for such a justification using Bayes' theorem:

Probability of event, given evidence

$$= \frac{\text{Initial probability of both}}{\text{Initial probability of evidence}}.$$

Our 'evidence' is the observed uniformity t, while the predicted 'event' is a spread of uniformity covering *both* the observed time t *and* the predicted time p. In this case, then, the initial probability of *both* evidence and event is equal to the initial probability of the event, since any example of the latter must be an example of the former as well.

The initial probability of the observed uniformity t is

$$\mathrm{SUM}_0(\mathrm{kxP}(\mathrm{x}+\mathrm{t})).$$

By an exactly parallel argument, the initial probability of the observed *and* predicted uniformity (t + p) is $\mathrm{SUM}_0(\mathrm{kxP}(\mathrm{x}+\mathrm{t}+\mathrm{p}))$. Applying Bayes' theorem, we reach the following result:

Probability of a successful prediction

$$= \mathrm{SUM}_0(\mathrm{kxP}(\mathrm{x}+\mathrm{t}+\mathrm{p}))/\mathrm{SUM}_0(\mathrm{kxP}(\mathrm{x}+\mathrm{t}))$$

$$= \mathrm{SUM}_0(\mathrm{xP}(\mathrm{x}+\mathrm{t}+\mathrm{p}))/\mathrm{SUM}_0(\mathrm{xP}(\mathrm{x}+\mathrm{t})).$$

To justify taking the past as a rule for the future, this probability must grow as t increases. This condition will be satisfied if P(x + p)/P(x) is an increasing function[2]; but this is hardly surprising, since that is the same condition required by Bayes' theorem if it is to be the case that the longer a regularity goes on, the greater the chance of its continuing for a further time p.[3] Clearly, to claim that this is so is to beg the very question at issue: to assume that the past *is* a rule for the future.

[2] Since in that case the functions P(x + t + p) and P(x + t) provide a more favourable ratio as t gets larger while at the same time the minimum arguments for which they are evaluated – i.e. when x = 0, giving minimum arguments of (t + p) and t respectively – increase, thus excluding those lower valued arguments which would provide a less favourable ratio.

[3] More precisely, Bayes' theorem yields the condition that $\mathrm{SUM}_{t+p}(P(x))/\mathrm{SUM}_t(P(x))$ be an increasing function of t, but this comes to the same thing.

II

Mackie's defence of induction fails because he considers only the two possibilities, that the past is *some* positive rule for the future, and that the world is completely random. He can then claim, quite reasonably, that a dogmatic assumption of randomness would be extreme and arbitrary, and that we must therefore 'allow from the start that there is some better-than-zero probability of some not purely random pattern'. In our terms, Mackie assumes that either $P(x+p)/P(x)$ is an increasing function (i.e. that the longer a regularity goes on, the longer it is likely to continue) or that it is a constant (i.e. that the past history of a regularity has no effect whatever on its future – his talk of a simple 'pure randomness' here is misleading, for *any* probability density function of the form $P(x) = a\,e^{-ax}$ will yield this result). A possibility which Mackie does not consider is that $P(x+p)/P(x)$ is a decreasing (or an oscillating) function, and his argument will not succeed unless he can rule this out. There are two ways in which he might attempt to do so: one of these is pragmatic, and the other logical.[4]

First, it might be maintained that we have good reason to proceed upon the assumption that $P(x+p)/P(x)$ is an increasing function, for only if this is so can we draw any conclusions about the future at all. We are unable to justify the assumption in any other way, but must simply adopt it on the principle that even a slight chance of success is better than the certainty of failure. Such a defence of induction, however, relies upon the fact that we do actually predict inductively, and cannot help doing so: it is for precisely this reason that our forecasts can be reliable only if $P(x)$ is such as to validate induction. But if we cannot help reasoning inductively, then any *pragmatic* justification of induction is quite unnecessary in the first place; certainly there is no need whatever to resort to anything as sophisticated as probability theory in order to persuade us to do what we cannot but do!

The second approach is to argue that there is something meta-inductively inconsistent about the suggestion that $P(x+p)/P(x)$ should be a decreasing function. For this would imply that the world was in some way counter-inductive, or at least that the longer uniformity lasts, the less likely it is to continue. Now it is clear that if in such a world inductions consistently fail, then nevertheless one *meta*-induction, to the effect that inductions will *continue* to fail, must be successful. Counter-induction cannot give a coherent direction on all levels simultaneously, and this is sufficient to rule it out as a comprehensive predictive policy.

This reply to the sceptic is open to two objections. First, it is of no assistance to Mackie, for it has force only against an all-embracing

[4] Both were suggested to me by Mr Mackie, for whose helpful discussion I am extremely grateful.

counter-induction, and none against a counter-induction which is applied specifically to predictions of general uniformity (Mackie's chosen battle-ground). Secondly, it makes the pragmatic assumption that even the sceptic must be committed to *some* comprehensive predictive policy, after which it need but show that induction is the most consistent such policy available. This assumption, however, is just what the critic of induction will deny. Why suppose that *any* method of forecasting should be reliable, in the absence of an independent proof that the past is a rule for the future?

III

Mackie's argument fails for two reasons. First, even if a long observed uniformity increases the probability of a longer total uniformity, it does *not* follow that the probability of a certain length p of *unobserved* uniformity is also increased. For as the observed length t is increased, so also the predicted length $(t + p)$ is increased – the prediction is not constant, and is not in any way confirmed by Mackie's argument. Secondly, Mackie has no way of supplementing his argument with information about the function $P(x)$. Even if it is assumed that assignments of 'initial' probability (i.e. in advance of *all* experience!) about the future make sense; even if such problems as Goodman's paradox are overlooked; still there is nothing that can justify the particular assumption that the world's ways of working are inductive rather than counter-inductive. To put the point in a Humean fashion: no deductive argument can do the job, any inductive argument will beg the question, while a probabilistic argument will be circular, simply repeating the problems at a deeper level.

Lincoln College, Oxford