Philosophers' Imprint

volume 7, no. 2
march 2007

A PROOF OF INDUCTION?

Alexander George

Department of Philosophy
Amherst College

David Hume's argument that inductive inference cannot be justified consists of two halves. In the first half, he argues that no demonstration — which is to say, no mathematical or logical proof — can justify extrapolations from past observations to unobserved cases. In the second, he argues that no reasoning from experience can provide such a justification, either. If these methods exhaust our tools for justification then we must acknowledge, as vexing as it might seem, that we lack justification for inductive inferences. The second half of Hume's argument, involving as it does the thrilling circularity charge, has garnered the lion's share of attention. The first has generated less interest. This is partly because Hume there offers an argument in terms of conceivability that leaves many readers unimpressed; and partly because, notwithstanding his argument's weakness, its conclusion, that no demonstrative proof can justify induction, appears so obviously correct in the first place. A recent mathematical result, however, presents a challenge to these appearances.

It is not entirely clear how to state the problem of induction. I shall try to approximate it in mathematical terms so as to bring it into contact with the result to be discussed below. We shall describe a possible world by a function from the reals ($\mathbb{R}$) to the reals. We can think of the function's arguments as moments of time and of its values as encodings of the total state of the universe. Thus we describe a possible world by a function that tells us, for each time t, what the state of the universe is at t. The actual world is described by one specific such function, say f_a. Might the problem of induction be that of determining whether there is a strategy that would lead to correct prediction of the value $f_a(t)$ for each t? This cannot be our question, for it is trivially answered in the affirmative: let the strategy simply be "Make your predictions using the function f_a." This strategy is guaranteed to generate correct predictions for every argument t.

This is clearly a non-starter as far as a justification of induction is concerned, but it is worthwhile spelling out why. One unsatisfactory feature of the strategy is that it is simply a lucky guess: after all, if the world had been at all different, this strategy would have generated

incorrect predictions. We would like to know whether there are inductive strategies that are robust, that is, whose accuracy does not depend on our world's being just as it is. A second unsatisfactory element is that this strategy cannot be implemented without knowledge of all the facts about the universe. Obviously, if any rule for prediction helps itself to all facts about the future, its accuracy is assured. But if something is to count as an interesting strategy of extrapolation, it must be one whose input requirements are modest relative to the predictions it makes. Our own everyday inductive extrapolations are both robust and modest in these senses: we believe that our extrapolating techniques would continue to be reliable even if the world were not just so, and the basis on which we make inferences at t involves at most information about the world prior to t.

With this in mind, it seems that our question about induction is better put this way: Is there a rule for by-and-large correct prediction of the values of f, where f is any function from reals to reals, and for each t, our rule takes as the basis for its prediction of $f(t)$ just the values of f at all $s < t$? Actually, the demand that a rule work for *any* function seems like an impossibly strong way to articulate the desideratum of robustness; for it seems impossible that there could be a procedure for prediction so hardy that it would deliver reliable results regardless of the kind of world in which we might live. And yet, as we shall soon see, precisely this desideratum can be met.

This very strength suggests differences between application of such a rule and familiar instances of inductive inference. For surely we do not think that our inductive extrapolations would yield correct results however wild our world; put another way, a rule that yielded correct but very counter-intuitive predictions would not resemble anything we would call "induction". By contrast, there are other respects in which application of a rule that satisfies the above conditions seems *more* restrictive than our actual inductive practices. For one thing, the input to such a rule is infinitary, whereas we do not extrapolate from more than a finite number of instances. For another, on the basis of (finite) information about the past, we often extrapolate to the distant or at least not immediate future, whereas a rule that answers to the above characterization will, on the basis of data at points $s < t$, issue in a prediction regarding events at t only. Despite these differences, such a rule is akin to our familiar induction in being recognizably one of prediction, one that licenses ampliative inferences. And so to justify the existence of such a rule would be to justify the validity of a form of extrapolation.

A surprising new result in mathematics seems to do just that.[1] The result establishes that there is a rule that will correctly predict most of the values of any function on the basis of its past behavior. Such a justification of a form of ampliative inference is doubly unexpected. For by now (to paraphrase Hume) surprises in these matters were not to be expected. And furthermore, the justification on offer takes the form of a *mathematical* demonstration.

So, to repeat, it can indeed be proved that there is a rule—the "Hardin-Taylor rule", I shall call it—that will, for any arbitrarily chosen function f, correctly predict most values of f on the basis of its past behavior; that is, for most t the rule will correctly predict $f(t)$ on the basis of f's values at all $s < t$. I shall first sketch the proof's central idea and then turn to assess the result's philosophical significance.

We begin by well-ordering $\mathcal{R}$, the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. (Here the Axiom of Choice must be employed, a fact to which we shall return.) That is, we select a linear ordering $\preccurlyeq$ for $\mathcal{R}$ such that every non-empty subset of $\mathcal{R}$ has a least element in the $\preccurlyeq$ ordering. Assume now that one has arbitrarily fixed a function f in $\mathcal{R}$ and that one wishes to predict its value at t. The Hardin-Taylor rule would have us consider the collection of all functions g in $\mathcal{R}$ that agree with f on all arguments less than t. That is, we consider:

$$A(f, t) = \{\, g \in \mathcal{R} \mid g(s) = f(s), \text{ for all } s < t \,\}.$$

Since $A(f, t)$ is a non-empty subset of $\mathcal{R}$, it contains a $\preccurlyeq$-least element;

1. "A Peculiar Connection Between the Axiom of Choice and Predicting the Future", Christopher Hardin and Alan D. Taylor, *American Mathematical Monthly*, forthcoming.

call this function $m_{A(f,\,t)}$. The Hardin-Taylor rule's prediction for $f(t)$ is $m_{A(f,\,t)}(t)$.

How good a rule is this for predicting the values of f? Very surprisingly, it can be shown that this rule's predictions will almost always be correct. The core of the proof is as follows. Consider the set of arguments of f for which the rule delivers the wrong results. That is, consider:

$$W_f = \{\, t \in \mathbb{R} \mid f(t) \neq m_{A(f,\,t)}(t) \,\}.$$

What we need to show is that almost all elements of $\mathbb{R}$ fail to belong to W_f.

We begin by establishing a relationship between the $<$ ordering on W_f and the $\prec$ ordering on $\mathcal{R}$. Let $u, v \in W_f$ and assume $u < v$. First we note that:

$$(*) \quad m_{A(f,\,u)} \neq m_{A(f,\,v)}.$$

This follows because we know that $m_{A(f,\,v)}(u) = f(u)$, since $m_{A(f,\,v)}$ agrees with f at all points less than v, and also that $m_{A(f,\,u)}(u) \neq f(u)$, since $u \in W_f$. Next we note that:

$$(**) \quad m_{A(f,\,u)} \preccurlyeq m_{A(f,\,v)}.$$

Since $m_{A(f,\,v)}$ agrees with f at all points less than v and $u < v$, $m_{A(f,\,v)}$ must agree with f at all points less than u. Consequently, $m_{A(f,\,v)} \in A(f, u)$. But $m_{A(f,\,u)}$ was chosen to be the $\preccurlyeq$-least element in $A(f, u)$, and so (**) follows. From (*) and (**), we infer that $m_{A(f,\,u)} \prec m_{A(f,\,v)}$. We arrived at this on the basis of our assumption that $u < v$, where $u, v \in W_f$. Pulling this together, the upshot is that if two elements of W_f are ordered by the $<$ relation, then their corresponding functions (*i.e.*, the ones that the Hardin-Taylor rule supplies to generate a choice of value for f at those points) will be similarly ordered by the $\prec$ relation.

This last observation enables us to argue that $<$ must be a well-ordering on W_f. For let E be any non-empty subset of W_f. We can show that E must contain a least element in the $<$ ordering as follows. Consider $\mathcal{E}$, the non-empty subset of $\mathcal{R}$ that contains all and only the functions that correspond to the elements of E. Since $\mathcal{R}$ is well-ordered by $\prec$, there must be a function $m_{A(f,\,s)}$, for some $s \in E$, that is the $\prec$-least function in $\mathcal{E}$. We can now show that s is the $<$-least element in E. We argue for this by contradiction. Assume there is an $r \in E$ such that $r < s$. But then, by the upshot of the previous paragraph, $m_{A(f,\,r)} \prec m_{A(f,\,s)}$, contradicting the fact that $m_{A(f,\,s)}$ is the $\prec$-least element in $\mathcal{E}$. Hence, any non-empty subset of W_f must have a $<$-least element; that is, W_f is well-ordered by $<$.

Now, it is a familiar fact that every subset of $\mathbb{R}$ that is well-ordered is countable. Therefore, W_f is countable. What this means is that the collection of "misses"—of points at which the Hardin-Taylor rule fails to deliver a correct prediction for the value of f—is at most the size of the natural numbers. It is in this sense that the rule's predictions for f are correct almost all of the time. Since f was arbitrarily chosen, this conclusion holds for every function in $\mathcal{R}$.

This already striking result can be made even more startling. In predicting the value of a function f at t, one does not actually need to know the behavior of f at *all* points less than t. In fact, no particular values of f need to be known. To predict $f(t)$, all that an appropriate rule would require are the values that f takes on in the interval $(t-\varepsilon, t)$, for any $\varepsilon > 0$. However much information one thought one would need for by-and-large accurate predictions is in fact more than one needs. Not quite Lady Macbeth's "The future in the instant", but as close to it as one pleases.

The result, as the authors soberly put it, is "peculiar". It might also seem positively paradoxical. The proof tells us that every function is such that the Hardin-Taylor rule will be right in predicting its values almost all of the time. But surely, each time the rule is employed it will generate incorrect predictions for just about any world one might be in. Knowing the behavior of an arbitrarily chosen function at points less than t offers one no way at all to work out its value at t: the information given is just irrelevant to the conclusion drawn. Now, these two claims certainly appear to conflict. To put the matter another way, the proof tells us that if one takes any particular function and asks,

"At how many points will the Hardin-Taylor rule correctly predict the value of this function?", the answer is "Almost every." And yet it seems obvious that if one considered any particular point and asked, "How many functions will have their values at this point correctly predicted by the rule?", the answer would be "Almost none."[2]

While this does appear paradoxical, an outright contradiction is in fact skirted. In order to see this, let us restate the upshot of the proof by making explicit its quantificational structure:

(1) *Every function f* from $\mathbb{R}$ to $\mathbb{R}$ is such that, *for almost every point t* in $\mathbb{R}$, the Hardin-Taylor rule will correctly predict the value of f at t.

Now, what seems obviously to be the case is that:

(2) *Every point t* in $\mathbb{R}$ is such that, *for almost every function f* from $\mathbb{R}$ to $\mathbb{R}$, the Hardin-Taylor rule will incorrectly predict the value of f at t.[3]

(1) is the upshot of the proof, and (2) is a more careful formulation of what seems obvious. But we can now see that there is no paradox, for (1) and (2) do not even on their face contradict one another: the quantificational prefixes of (1) and (2) differ.[4] That these two claims can both be true is indeed surprising, but it is not an outright paradox.[5]

Ludwig Wittgenstein issued a salutary warning against being suckered by a "puffed-up proof".[6] Hume likewise cautions us against proofs that are "mere sophistry and illusion".[7] But such proofs may nevertheless be worth thinking about; at the very least, our understanding will be improved by discovering how the puffing-up takes place. Is this such a proof?

We have just seen that reflecting on Hardin and Taylor's result does indeed help us to appreciate differences that we may not have previously suspected. Thus, one might have thought that to establish something like (2), namely the pointwise absence of a priori justifications for any predictive rule, would be to establish the absence of any a priori justification of a predictive rule that (more or less) uniformly held. Indeed, this is a helpful way of looking at Hume's reasoning: he argues (correctly, if not entirely persuasively) that at any particular moment of time, no amount of information regarding prior events could form the basis of a proof about what to expect next. And from this, he infers to the conclusion of (what I called above) the first half of his argument, that a priori demonstrations of the general reliability of inductive extrapolations are not available. But surprisingly, this inference is mistaken: it turns out that (2)'s truth is not incompatible with (1)'s. At the very least, then, the proof teaches us differences.

But does the Hardin-Taylor proof of (1) provide us with a justification of induction? If not, this must be either because (1) fails after all to express what we seek to justify, or because the proof of (1) fails to offer a justification of the kind we want. I argued at the outset that,

2. "*Almost* none" because the rule will deliver correct results for those functions which happen to accord with the rule.

3. One might hope to precisify and justify (2) by defining a measure on the set of all functions and then showing that, for any t, the measure of the set of all functions for which the Hardin-Taylor rule correctly predicts $f(t)$ is equal to 0. Christopher Hardin has shown that this is not possible (at least for countably additive measures). (I am grateful to Hardin for this personal communication.) In fact, for almost all t, the collection of functions for which the Hardin-Taylor rule correctly predicts $f(t)$ is not measurable (at least, this is so for any reasonable measure). Measure theory thus does not provide a way of justifying one's strong intuition that, for any given t, the collection of functions whose values at t are correctly predicted by the Hardin-Taylor rule is small. (I am grateful here, and elsewhere, to Daniel J. Velleman for many helpful conversations.)

4. As the previous footnote makes clear, (1) and (2) differ not only in their quantificational structure but also in their mathematical interpretability.

5. One might wonder whether a contradiction could be generated from the following claim: "If every x is such that, for almost every y, $R(x, y)$, then every y is such that, for almost every x, $R(x, y)$". But this claim does not seem plausible and in fact it is not provable in ZF, if ZF is consistent.

6. *Remarks on the Foundations of Mathematics*, Third Edition, Basil Blackwell, 1978, p. 132.

7. According to Hume, this is the fate of any demonstration that concerns not merely "quantity and number" (*An Enquiry Concerning Human Understanding*, Section XII, Part III).

while one cannot identify induction with a principle that functions as the Hardin-Taylor rule does, the latter makes possible a form of prediction; and that consequently, to justify (1) *would* be to justify the correctness of an ampliative inference from past events. Such a justification would indeed show us that there is a rule which issues in by-and-large correct predictions, which does not depend on lucky guesses regarding the world in which we might live, and which takes as given only information about the past. Hence, either we have here a mathematical demonstration of the reliability of some form of extrapolation, or there is something suspect in the proof. Of course, there is nothing suspect in the sense of mathematical error: the proof has passed mathematical muster. But might there be some feature of the proof that compromises its ability to offer the kind of justification we are after?

One aspect of the proof that suggests itself is its reliance on the Axiom of Choice. This axiom is indeed needed to establish the existence of a well-ordering on $\mathcal{R}$: this well-ordering is constructed using the choice function whose existence is guaranteed by the Axiom of Choice. But the Axiom of Choice, as wielded by the classical mathematician, does not describe a particular choice function; it merely says that one exists. And so we are not given a particular well-ordering on the collection of functions; we are merely told that there is one. Consequently, while this proof assures us that there does exist a rule for prediction that is largely correct, it gives us no handle on what specifically that rule is.

We must decide, therefore, what we really wanted to know when we raised the problem of induction. Did we merely (as it were) want to know whether we can justify the existence of some robust strategy that, basing itself only on information about the past, accurately predicts the future, or did we want to know whether we can justify *our* inductive extrapolations into the future? Before this proof, one might never have thought to draw this distinction, for one might never have thought that justification of a rule for extrapolation might take the form of a mathematical proof, and so never have imagined that a justification could be highly non-constructive in nature.[8] Any positive answer to the first question that proceeds via a constructive existence proof would illuminate the second question; for it would provide us with a particular justified strategy of extrapolation against which we could compare our own inductive practices. But an answer to the first question that takes the form of a non-constructive proof of existence leaves us no better off than we were before in judging the reliability of our own extrapolations.

Thus, if our only concern was to establish that *our* inductive inferences are by and large correct, then the present proof will not help. But what if one wanted, in the first instance, to show that the past has some rational bearing on the future (where its having precisely the bearing we think it has is of subsequent concern)? What if, moreover, in a spirit of resurgent rationalism, one wanted to establish the existence of such a bearing through a priori reasoning? That is, what if one wanted to know whether it is conceptually incoherent[9] to imagine the future's being radically unpredictable given information about the past? On one interpretation of this question we now have an answer, and a surprising one.

8. Hume seems unwilling to countenance a non-constructive argument for the existence of reliable inferences from the past to the future. Nothing will do for him short of a justification that some *specified* procedure of inference (preferably our own) is by and large correct. Thus he writes: "There is required a medium, which may enable the mind to draw such an inference, if indeed it be drawn by reasoning and argument. What that medium is, I must confess, passes my comprehension; and it is incumbent on those to produce it, who assert that it really exists, and is the origin of all our conclusions concerning matter of fact." (Section IV, Part II)

9. Where a condition on conceptual coherence is consistency with ZFC, Zermelo-Fraenkel set theory plus the Axiom of Choice, a theory most mathematicians believe to be true and indispensable for the formal development of mathematics.