Logique & Analyse **203** (2008), 263–269

JUSTIFYING INDUCTION MATHEMATICALLY: STRATEGIES AND FUNCTIONS

ALEXANDER PASEAU

1. Think of the philosophical problem of induction thus. Model time as the real number line and encode the total state of the universe at a given time by a real number. Suppose our world is representable in this manner by a function from reals to reals, which we may call $F_@$. Next define $F_{@|t}$ as the function $F_@$ restricted to times less than t; thus $F_@ : \mathbb{R} \to \mathbb{R}$ and $F_{@|t} : \{x \in \mathbb{R} : x < t\} \to \mathbb{R}$. If t is the present instant, $F_{@|t}$ encodes the past history of the universe since for any given past time $t*$ the value of $F_{@|t}$ applied to $t*$, $F_{@|t}(t*)$, represents the state of the universe at $t*$. An inductive rule I is then a rule that, given $F_{@|t}$, outputs a function $I(F_{@|t})$ whose value at t agrees with F for most values of t. I is thus a function from partial real functions to real functions ($I :{}^{\{x\in\mathbb{R}:x<t\}}\mathbb{R} \to {}^{\mathbb{R}}\mathbb{R}$) such that $(I(F_{@|t}))(t) = F_@(t)$ for most t. It is a rule that predicts, mostly correctly, the present on the basis of the past.

Suppose that we could exhibit such a function I, prove that it has this property, and perhaps also prove that it serves as an inductive rule for some other worlds too — so that it is no fluke that it works for $F_@$. We will then have solved a version of the philosophical problem of induction. We will have exhibited a rule that reliably predicts the present on the basis of the past (and does so in nearby worlds as well). If such an I can be found, philosophy will have solved one of its apparently unsolvable problems.

Mirabile dictu, a recent mathematical result establishes the existence of such an inductive function I. Christopher Hardin and Alan Taylor (2008) have shown that a function I with the property that for any F, $(I(F_{|t}))(t) = F(t)$ for almost all values of t exists. ‘Almost all’ can be made precise: the set of times for which $(I(F_{|t}))(t) \neq F(t)$ has measure zero, indeed is countable. It is in this strong sense that the function I is almost always right: I is right for all but a measure-zero, countable subset of the set of time instants (modelled as the reals). Moreover, as the notation indicates, Hardin and Taylor’s result holds for any function F from reals to reals, not just $F_@$. So it is no fluke that I works for $F_@$.

I’s existence follows from basic set theory. Take a well-ordering of the set ${}^{\mathbb{R}}\mathbb{R}$ of all functions from reals to reals. Such a well-ordering exists in virtue of the Axiom of Choice, equivalent in Zermelo-Fraenkel set theory to the Well-Ordering Principle. For each element F of ${}^{\mathbb{R}}\mathbb{R}$ and time-instant t, consider the set of functions that agree with F for all values less than t. Let I be the function that for any F and t picks the least element of this set under the chosen well-ordering of ${}^{\mathbb{R}}\mathbb{R}$. It is then provable that I has the required property.[1] For more details, consult Hardin & Taylor (2008).

The Hardin-Taylor result is beyond question, given standard set theory (ZFC). So we apparently have an unassailable answer to a version of the philosophical problem of induction, understood as the problem of reliably inferring the present state of the world from its past ones. A remarkable and entirely unexpected resolution, coming more than 250 years after Hume first articulated the problem, and at so advanced a stage of the debate. Can it really be true? Is there a catch?

2. The mathematical result is significant; but there is an important catch. Of course one can complain that the result makes essential use of the Axiom of Choice, a suspect principle. Or one can complain that the function I, even if it could be exhibited, would almost certainly be of no practical value since we are unlikely ever to know the total past state of the universe. Or that the encoding of the total state of the universe may be imperspicuous, meaning that we could never decode it. Or that the total state of our universe may not be encodable by a real number — perhaps our universe's complexity is greater than that — so that the Hardin-Taylor result may not even yield a theoretical solution to the problem of induction. Or that this is only one version of the problem of induction, and not nearly as important as, say, predicting the future given the past and present. Indeed, given that we always observe the present, the epistemological problem of reliably predicting the present given the past is perhaps not even correctly described as a version of the problem of induction, which after all is the problem of how to justify inferences from observed cases to unobserved ones.

[1] Sketch proof: Suppose $t < t*$ and neither $I(F_{|t})$ nor $I(F_{|t*})$ get the value of F right at their respective 'presents', i.e. $(I(F_{|t}))(t) \neq F(t)$ and $(I(F_{|t*}))(t*) \neq F(t*)$. $I(F_{|t})$ is no greater than $I(F_{|t*})$ in the well-ordering of ${}^{\mathbb{R}}\mathbb{R}$ because the set of functions that agree with F up to t includes the set of functions that agree with F up to $t*$, hence the least element of the latter cannot be smaller than the least element of the former. Also, $I(F_{|t}) \neq I(F_{|t*})$ since they disagree on t: $(I(F_{|t})$ gives the wrong value — it disagrees with F at t — while $I(F_{|t*})$ gives the right value — it agrees with F at t. Thus if $t < t*$ then $I(F_{|t}) < I(F_{|t*})$ in the well-ordering of ${}^{\mathbb{R}}\mathbb{R}$. It follows that there cannot be an infinite-descending chain of such times, as it would induce an infinite-descending chain of functions in the well-ordered set ${}^{\mathbb{R}}\mathbb{R}$. But any set of reals that has no infinite-descending chain (under $\mathbb{R}$'s natural order) is countable and thus of measure zero.

These objections are answerable to a degree. The use of the Axiom of Choice is indeed essential; but these days the axiom hardly attracts the controversy that it did a century ago when Zermelo introduced it, for good reasons we will not rehearse here. The result does not just apply to functions from $\mathbb{R}$ to $\mathbb{R}$ but generalises a little beyond it. In particular, it does not assume that the range of the function is $\mathbb{R}$: for example the proof generalises to a universe so complex that its states are only encodable by elements of a set of cardinality greater than $\mathbb{R}$, or a universe so simple that its states are binary. The only assumption made about the range is that it has more than one element. However, the proof does make essential use of the fact that well-ordered subsets of the domain $\mathbb{R}$ (under its natural order) are of smaller cardinality than $\mathbb{R}$ itself, a property lacked by many other potential domains, e.g. the set of rationals.[2] And nothing rules out the possibility that, even if we knew I and the past states of the universe, the countably many — presumably finitely many — times at which we humans use I to predict the present are all times at which $I(F_{@|t})$ disagrees with $F_@$.

More positively, as Hardin and Taylor demonstrate, the result can be extended to show that for any $s < t$, there is an $\varepsilon > 0$ such that if we know the state of the universe during the time interval (s, t) then there is a function that predicts the state of the universe correctly for almost all times in the interval $[t, t+\varepsilon)$. In other words, to apply the result we need only know the total past state of the universe for some interval prior to the present; a function then exists that predicts the state of the universe on this basis mostly correctly not just for the present but into the future as well.

Suppose then that one interprets the Hardin-Taylor result modestly, as showing that, given the assumption that the total state of the universe is encodable and that time can be modelled by the reals, there is a solution to one version, perhaps not the most important one, of the purely theoretical problem of induction. Even that interpretation of the Hardin-Taylor result is not modest enough, I claim.

3. To see this, consider what the philosopher Alexander George has read into the result. George points out that there are two things we might have been asking when we first posed the question of induction. The first is "whether we can justify the existence of some robust strategy that, basing itself only on information about the past, accurately predicts the future"; the second is

[2] Suppose that the set of times under the (linear) 'earlier than' relation in our universe is $(T, <_t)$ and that the state of the universe is encodable by an element of the set S. If $(T, <_t)$ has the property that any of its well-ordered subsets (under the order $<_t$) is of smaller cardinality than T then an analogous result follows. If not, all that we can say is that set of times t of T at which $I(F_{@|t})$ disagrees with $F_@$ at t is of cardinality smaller than ${}^{T}S$ (I picks an element of ${}^{T}S$ no greater than $F_@$) but of course $|\,{}^{T}S\,| > |\,T\,|$ given $|\,S\,| \geq 2$.

"whether we can justify *our* inductive extrapolations into the future" (2007, p. 5). Given its non-constructive nature, the Hardin-Taylor proof does not answer the second question. But according to George it answers the first.

George is quite right to distinguish the two questions. What the existence of an inductive function I has to do with our inductive extrapolations, our taking the sightings of only white swans in the past as good evidence that all swans are white, that night regularly follows day and day regularly follows night, etc., is not immediate. But there is an important further distinction which his discussion misses. This distinction is between different types of nonconstructive proof, and it turns out to be crucial to the evaluation of the Hardin-Taylor proof as providing a reliable present-predicting strategy. Once it is appreciated, the Hardin-Taylor proof can no longer plausibly be called a predictive *strategy*. Because it is so nonconstructive, it fails to yield a strategy.

4. Let me explain. To do so, an overview of nonconstructive reasoning in mathematics is required. Certain kinds of existence proofs are nonconstructive in the sense that they do not exhibit an instance of the object in question. An example is the standard proof of the existence of transcendental numbers — i.e. non-algebraic numbers or real numbers that are not the solution of some polynomial equation with rational coefficients (numbers such as π or e, and unlike $1 + \sqrt{3}$, which is the solution to $(x-1)^2 = 3$). Since there are only countably many algebraic numbers, there must be uncountably many transcendental numbers (there are uncountably many reals altogether). But this proof is of no help in finding an instance of a transcendental number.[3] Other proofs are nonconstructive because, though they exhibit a range of possible instances, they do not exhibit a single instance known to satisfy the condition. For example, a customary proof of the fact that an irrational number raised to the power of an irrational can be rational considers $\sqrt{2}^{\sqrt{2}}$. Either this number is rational, in which case it provides the instance in question (since $\sqrt{2}$ is irrational); or it is irrational, in which case $\sqrt{2}^{\sqrt{2}}$ raised to $\sqrt{2}$ is the sought-after instance (since $(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = (\sqrt{2})^2 = 2$). But this proof is nonconstructive, because it does not tell us which of the two pairs of numbers $(\sqrt{2}, \sqrt{2})$ or $(\sqrt{2}^{\sqrt{2}}, \sqrt{2})$ is the relevant instance.

There is an even stronger kind of nonconstructive proof. Proofs of this kind not only fail to exhibit an instance but are such that no instance can be exhibited. And it may even be provable that no instance is exhibitable. This is the situation we find ourselves in with well-orderings of the reals.

[3] Though other, constructive arguments give us particular instances of transcendental numbers such as π or e.

The proof that there is a well-ordering of the reals is nonconstructive in this stronger sense. Using forcing, it can be shown that no well-ordering of the real numbers is definable in ZFC (nor even in ZFC + the General Continuum Hypothesis), assuming its consistency. No formula of the language of set theory expresses a real well-ordering.[4] Since ZFC is currently our strongest accepted mathematical theory, this is equivalent to saying that a real well-ordering cannot be defined.

(Of course, ZFC may not always be our strongest accepted mathematically theory. It may eventually be superseded by a more powerful theory in which a well-ordering of the reals *is* definable. But ZFC currently holds that status, so a proof of indefinability in ZFC is for us a proof of indefinability.)

Applied to our present discussion, what the indefinability of a real well-ordering shows is that the Hardin-Taylor proof is nonconstructive in the strong sense. The proof derives the function I from a well-ordering of the set ${}^{\mathbb{R}}\mathbb{R}$ of all functions from reals to reals. But it would follow straightforwardly from the definability of a well-ordering of ${}^{\mathbb{R}}\mathbb{R}$ that the real numbers also have a definable well-ordering (e.g. real numbers may be identified with constant real-valued functions). Thus the well-ordering of ${}^{\mathbb{R}}\mathbb{R}$ which the proof uses to derive I is not definable, so the proof does not give us a way of defining the function I. It follows that the Hardin-Taylor proof does not give us a 'strategy', 'method' or 'rule' for predicting the present. A strategy for predicting the present given the past that cannot be expressed in words does not merit the label.[5] If we cannot in principle express or exhibit our supposed method, it is no method at all, not even an in principle one. Likewise for the notion of 'rule' which has to be expressible to count as such.

The point that a strategy or method or rule has to be in some way describable and not ineffable should be obvious enough, and can be illustrated with numerous examples. One should suffice here. Suppose that the president of a large, powerful country which has invaded and occupied a smaller one is called before the United Nations. He is asked what his exit strategy is for withdrawing his nation's forces from the occupied country. The president, alas, has no exit strategy. Not to worry: a sly adviser informs him that a strategy can be expressed as a finite sequence of sentences, that all such

[4] I first came across the result in Kunen (1980, p. 245, E4), where it is offered as an exercise. One of the earliest applications of Cohen's forcing method, it was proved in Feferman (1964).

[5] In several areas of mathematics — e.g. game theory or descriptive set theory — the word 'strategy' is used to denote arbitrary functions of a certain kind, be they expressible or not. We should not confuse this usage with the standard non-mathematical one, which is the usage relevant to the problem of induction.

sequences exist (in the platonic realm, at least, he reassures him), and consequently that some exit strategy exists. The president, bereft of any better ideas, stands up and tells the general assembly that, contrary to general suspicion, there is an exit strategy for withdrawing his nation's troops from __. Sadly, however, the president cannot express the strategy nor offer any guidance as how to articulate it. How do the UN delegates react? With outrage: that is no strategy, they protest. The mere existence of some sequence of sentences describing how to withdraw the occupying forces, without any idea of which one it is or how to find it, does not constitute a strategy. They are right, of course: the president does not have a strategy, not even the recipe for a strategy. Similarly, the mere existence of an ineffable function with the required inductive property does not constitute a strategy.

One could respond by augmenting the language of ZFC with a term '$<_{^{\mathbb{R}}\mathbb{R}}$' and adding to the axioms the formula '$<_{^{\mathbb{R}}\mathbb{R}}$ is a well-ordering of $^{\mathbb{R}}\mathbb{R}$', the predicate being spelled out in the standard way. In this new theory (consistent if ZFC is) it would be easy to name a well-ordering of $^{\mathbb{R}}\mathbb{R}$ — its name is simply '$<_{^{\mathbb{R}}\mathbb{R}}$' — and accordingly to name a strategy for picking the least element (under this well-ordering) of a subset of $^{\mathbb{R}}\mathbb{R}$. However, though we would then possess a name for the strategy that name would be uninformative. Paradigmatically, an informative name would be a description purely in the language of ZFC — a language capacious enough for virtually all mathematical purposes. It cannot be a bare stipulated name for something we are assured exists but can get no descriptive handle on. Think back to our analogy: the president cannot respond to his critics by picking some name, 'S' say, stipulating that 'S' names one of the sequences of sentences that describes an exit strategy, and claiming that in virtue of using that name he is now in possession of a strategy. What the delegates want instead is an informative description of how and when the troops will leave the occupied country; a stipulated name (or a description relying on it) does not provide them with that information. Names that constitute strategies must have informational value.

5. Let us take stock. A more nuanced appreciation of nonconstructive proof gives us the means to distinguish *three* questions we might ask given complete information about the past:[6]

(1) Can we justify the existence of a reliable present-predicting *function*?
(2) Can we justify the existence of a reliable present-predicting *strategy*?

[6] Given the refinement mentioned earlier, 'present-predicting' can be strengthened to 'present-and-proximate-future-predicting'.

(3) Can we justify *our* present-predicting strategy?

The Hardin-Taylor proof answers only the first question. It tells us that there is a function I which, applied to any past state of the universe up to the present instant t (encoded by the function $F_{@|t}$) outputs a function $I(F_{@|t})$, which applied to the present time t gives the same value as $F_@$ for most values of t, i.e. $(I(F_{@|t}))(t) = F_@(t)$ for most t. But it does not answer the second question, never mind the third. The function I has not been described. The Hardin-Taylor proof is a pure existence proof — it is nonconstructive in the strong sense — and thus does not provide a strategy, not even a recipe for a strategy.

If the converse of the Hardin-Taylor result could be established, that is, if it could be shown that if a reliable future-predicting function could be defined then so could a well-ordering of the reals, it would follow that the answer to the second question *must* be negative (taking ZFC as our background theory). It would be mathematically demonstrable that no reliable present-predicting strategy is mathematically definable. As it is, the Hardin-Taylor proof gives us no purchase on the question of whether a reliable present-predicting strategy exists. The existence of a function is one thing, the existence of a strategy another.[7]

Wadham College, Oxford OX1 3PN, UK
E-mail: alexander.paseau@philosophy.ox.ac.uk

REFERENCES

Feferman, S. (1964), "Some applications of the notions of forcing and generic sets", *Fundamenta Mathematicae* 56, pp. 325–345.

George, A. (2007), "A Proof of Induction?", *Philosophers' Imprint* 7, pp. 1–5.

Hardin, C. & Taylor, A. (2008), "A Peculiar Connection Between the Axiom of Choice and Predicting the Future", *American Mathematical Monthly* 115, pp. 91–96.

Kunen, K. (1980), *Set Theory: An Introduction to Independence Proofs*, Amsterdam: North-Holland.

[7] Thanks to the *Logique et Analyse* referee for helpful comments.