The Philosophical Quarterly, Vol. 53, No. 212 *July 2003*
ISSN 0031–8094

PROBABILISTIC INDUCTION AND HUME'S PROBLEM: REPLY TO LANGE

By Samir Okasha

Marc Lange has criticized my assertion that relative to a Bayesian conception of inductive reasoning, Hume's argument for inductive scepticism cannot be run. I reply that the way in which Lange suggests one should run the Humean argument in a Bayesian framework ignores the fact that in Bayesian models of learning from experience, the domain of an agent's probability measure is exogenously determined. I also show that Lange is incorrect to equate probability distributions which 'support inductive inferences' with probability distributions which assign probability to contingent propositions/events.

In his famous argument for inductive scepticism, Hume operated with a very simple description of what our inductive behaviour is like. He focused exclusively on inferences such as 'Bread has always nourished me up to now; therefore bread will nourish me today'. It is widely recognized that the actual inductive behaviour of both scientists and laymen is much more complicated than this. This prompts the question whether the force of Hume's sceptical argument depends on the simple description of our inductive behaviour with which he operated, or not. Most philosophers think the answer is 'No': sceptical considerations of Hume's sort will apply whatever the correct description of our inductive behaviour turns out to be, they believe. In a recent paper, I contested this view.[1] I argued that (i) it would be a coincidence if Hume's argument did have this generality, given that he was not just making the 'boring' point that inductive reasoning is fallible; and (ii) in relation to the Bayesian characterization of inductive reasoning, there is no way in which a Humean sceptical argument can be run.

In a recent reply to my paper, Marc Lange contests a central plank in my argument for (ii).[2] In rejecting (ii), Lange is in good company. Most philosophers of science believe that Hume's problem resurfaces in a Bayesian framework as the problem of justifying one's choice of prior probability distribution. Bayesian agents' states of opinion at any time are heavily dependent on their prior distributions, and the probability axioms do not in themselves tell us what those distributions should be, beyond requiring that they must be coherent. So two Bayesian agents who

[1] S. Okasha, 'What Did Hume Really Show about Induction?', *The Philosophical Quarterly*, 51 (2001), pp. 307–27.

[2] M. Lange, 'Okasha on Inductive Scepticism', *The Philosophical Quarterly*, 52 (2002), pp. 226–32.

receive the same evidence may end up in entirely different states of opinion, if they start from different priors; and the Bayesian theory says nothing about which prior distribution is correct. Because of this 'problem of the priors', Bayesianism has usually been regarded, even by its most forceful advocates, as impotent in the face of Humean scepticism. This pessimistic view was articulated clearly by Frank Ramsey, one of the founders of modern Bayesianism, and has recently been defended at length by Colin Howson.[3]

In my opinion this pessimistic view is not entirely right. It is true that if a Bayesian reasoner is asked to justify starting with any actual prior distribution, there is no satisfactory answer to give. So Bayesianism does not *solve* the traditional problem of induction, in that it does not show that any particular set of opinions about the unobserved is the 'right' one to have. But, as I argued in my original paper, this in itself does not vindicate inductive scepticism; it only shows that an element of guesswork is involved in all empirical enquiry. Prior probabilities, in Bayesian models of induction, represent background beliefs about the world. An inductive sceptic who objects to Bayesians' reliance on prior probabilities is accusing them of helping themselves to substantive assumptions about the world without entitlement, hence begging the question. This accusation fails, in my view, simply because there is *no alternative* to adopting some prior distribution that embodies an opinion about how the world is. (When I say that a probability distribution 'embodies' an opinion about the world, I mean that the agent whose doxastic state the probability distribution represents holds that opinion.) This of course does not show that any actual prior distribution is justified (that would amount to a 'justification of induction'). However, it does cast doubt on an inductive sceptic's attempt to show that a Bayesian reasoner behaves in an epistemically *unjustified* way. If there is no alternative, logically, not just psychologically, to operating with a prior distribution that embodies opinions about the world, it is hard to see how the sceptic's charge of question-begging can stick.

Were there a viable principle of indifference, matters would be different. For the principle of indifference was precisely an attempt to generate 'information-free' prior distributions, which correspond to the epistemic state of 'pure ignorance' about which of a set of alternative states of affairs obtains. Had the principle of indifference worked, the inductive sceptic's charge would succeed. For then anyone adopting a prior distribution other than the one dictated by the indifference principle would indeed be making a substantial assumption about the world without entitlement. But the indifference principle does not work: the notion of 'information-free' priors is chimerical. Suppose a Bayesian agent knows nothing about the value of a particular experimental parameter, except that it can be any real number between p and q. The parameter is represented by a continuous random variable x. The indifference principle would recommend agents to distribute their probability uniformly across x – to adopt any other distribution is to go beyond the known data. But a uniform distribution across x must be non-uniform across x^2, and an agent who knows nothing about the value of x obviously knows nothing about the value of

[3] F. Ramsey, 'Truth and Probability', repr. in R. Braithwaite (ed.), *The Foundations of Mathematics* (London: Routledge, 2000), pp. 156–98; C. Howson, *Hume's Problem* (Oxford UP, 2001).

x^2 either. So, on pain of violating the probability calculus, the agent cannot adopt a prior distribution which is totally neutral. Therefore while it is true that in adopting any particular prior distribution, a Bayesian agent is making an assumption about what the world is like, there is no alternative to doing so: *any* prior distribution embodies some opinion about the world. The inductive sceptic can hardly call Bayesians to task for doing what logic constrains them to do.

Lange contests this line of argument. He claims (p. 228) that there *is* an alternative, namely 'to operate from a prior distribution that embodies no opinion whatsoever about the way in which the world happens to be arranged'. This sounds very much as if Lange is a believer in the discredited principle of indifference. However, it turns out that this is not so. By a 'prior distribution that embodies no opinion about the world', Lange in fact means 'a distribution that makes no probability assignment at all to any claim about the world which concerns logically contingent matters of fact'.

A concrete example will illuminate the problem with his suggestion. Suppose a Bayesian B is designing a scientific experiment which has ten possible outcomes, represented by a random variable y whose values are integers from 1 to 10 inclusive. B considers how likely each of the possible outcomes is, i.e., how to distribute his prior probability over these outcomes. He is then approached by an inductive sceptic who tells him that since knowledge of the future is impossible, he is rationally obliged to choose a prior distribution which 'embodies no opinion about how the world happens to be arranged'. What might such a distribution look like? Advocates of the indifference principle would claim that the uniform distribution $P(y=i) = 1/10$ for each i is such a distribution; but as I have said, a distribution that is uniform over one variable must be non-uniform over another. Lange suggests instead a distribution that makes 'no probability assignment to any logically contingent claim about the world'. Since each of the $y=i$ is a logically contingent claim, none would be assigned a probability by Lange's distribution. The only two events to which Lange's distribution would assign a probability would be the sample space itself, $\{1, ..., 10\}$, and the null-set $\varnothing$. In effect, Lange's advice to B is simply to choose a probability measure with a smaller domain than the one he started with.

From the mathematical point of view, Lange's suggestion certainly makes sense. In probability theory, the domain of a probability measure is required to be a sigma-field of subsets of the sample space, and the set $\{\varnothing, \{1, ..., 10\}\}$ is obviously a sigma-field of the sample space $\{1, ..., 10\}$. But in instructing B that the domain of his probability measure should be $\{\varnothing, \{1, ..., 10\}\}$, Lange is in effect saying 'Don't even think about which experimental outcome is most likely'. Even if B were psychologically capable of heeding that advice, it is hard to see why he should be expected to do so.

The key idea of Bayesianism is to provide a probabilistic representation of an agent's states of opinion at any time, and of his changes of opinion from one time to another; the domain of his probability measure is therefore standardly taken to include all the propositions (or events) which the agent 'contemplates', and Boolean combinations thereof. So in Bayesian models of reasoning the domain of an agent's probability measure is exogenously determined – by what the agent is in fact thinking about. If we tell someone that his probability measure should have a smaller domain, we are therefore telling him that he should cease to think about some of the propositions about which he currently is thinking. But this is strange advice indeed. One would expect an inductive sceptic to recommend that we should cease to *believe* that the sun will rise tomorrow, if we wish our state of opinion to be immune from the sceptical critique, not that we should cease to consider the matter entirely.

Lange's suggestion therefore fails to engage with my argument. I argued that if the inductive sceptic wants to claim that a particular prior probability distribution is unjustified because it embodies opinions about empirical matters to which the agent is unentitled, the onus is on the sceptic to produce an alternative distribution of which the same is not true. Lange claims to discharge this burden, on behalf of the inductive sceptic, by recommending a (trivial) probability measure which distributes probability over a *different* set of propositions. But this is simply to change the subject. When I said that there is 'no alternative' to operating with a prior distribution which embodies some opinion about how the world is, I assumed that we were considering probability measures whose domains were identical, and contained contingent propositions. If we relax that assumption, and consider trivial probability measures of the sort Lange suggests, my claim obviously ceases to be true. But this in no way vindicates the inductive sceptic. In the case of the experiment described above, if I adopt the uniform prior distribution $P(y=i) = 1/10$ for each i (for example), and the inductive sceptic tells me that I am unjustified in doing so, the question he must answer is 'How *should* I distribute probability over the possible outcomes of the experiment?'. It is no answer to this question to say that I should cease to think about the possible outcomes altogether, and contract the domain of my probability measure accordingly, which is what Lange's suggestion amounts to.

Of course, one could oppose my argument by simply *denying* that the inductive sceptic faces the onus to tell people what state of opinion they should adopt, if he wants to claim that their actual state of opinion is unjustified. After all, on some understandings of philosophical scepticism, the sceptic is not recommending that we should alter or abandon our everyday beliefs, but only insisting that they do not constitute genuine knowledge.[4] A sceptic of this sort would be unembarrassed by inability to recommend an alternative prior distribution to the Bayesian reasoner (although, as I argued in my original paper, it is unclear whether such a sceptic's argument could really establish *scepticism*, rather than fallibilism). However, this is not Lange's line of reply. Lange *accepts* that the inductive sceptic faces the onus in question, and claims that it can be discharged. However, his suggestion as to how the sceptic should discharge the onus does not work, as I have shown.

There is in fact a second problem with Lange's argument. He maintains (p. 228) that a prior probability distribution that 'distributes subjective probability over some claims that are neither logical truths nor logical falsehoods' is the same as a prior probability distribution 'strong enough to support inductive inferences from our observations'. For convenience, I call a distribution that meets the first condition a 'non-trivial' distribution, and one which meets the second an 'induction-supporting'

[4] I am grateful to an anonymous referee for this observation.

distribution. It is because Lange regards these two conditions as equivalent that he thinks it begs the question against the inductive sceptic to operate with a non-trivial distribution. But in fact the two conditions are not equivalent, at least given a very natural conception of what it means for a probability distribution to 'support inductive inferences'. Their non-equivalence can be illustrated by an old but instructive chapter in the history of inductive logic.

The aim of Carnap's system of inductive logic was to define a unique two-place confirmation function *c*(*h*, *e*) which would tell us the degree to which evidence *e* confirms hypothesis *h*, where *e* and *h* are arbitrary sentences of a first-order language. Carnap took *c* to be a numerical function assigning real numbers in the unit interval to pairs of sentences in the language. His first requirement was that *c* must be a conditional probability measure, but this of course leaves open an infinite number of functions. The function Carnap himself favoured, for most of his life, was called '*c**'. Carnap frequently explained his preference for *c** by contrasting it with another function called '*c*†'.[5] The difference between *c** and *c*† emerges from the following simple example.

A universe contains three objects, a_1 to a_3, each of which either does or does not have a particular property F. The complete state of the universe can be described by saying of each object whether or not it has property F. Clearly there are eight different possible states of the universe, shown in the table below. The question Carnap addressed was 'How should we distribute our prior probability over the eight different states of the universe?'. The function *c*† assigns equal probability to each, i.e., 1/8. The function *c** is slightly more complicated. It divides the eight states into four groups, according to how many objects in the state possess the property F. Each group is assigned equal probability of 1/4, and *within* each group this probability is divided equally among all the states. The resulting probability of each of the eight states is shown in the right-hand column below.[6]

	States of the universe a_1	a_2	a_3	No. of Fs	*c*† prior prob. of state	*c** prior prob. of group	prior prob. of state
1.	F	F	F	3	1/8	1/4	1/4 = 3/12
2.	F	F	–		1/8		1/12
3.	F	–	F	2	1/8	1/4	1/12
4.	–	F	F		1/8		1/12
5.	F	–	–		1/8		1/12
6.	–	F	–	1	1/8	1/4	1/12
7.	–	–	F		1/8		1/12
8.	–	–	–	0	1/8	1/4	3/12

[5] R. Carnap, 'Statistical and Inductive Probability', and 'On Inductive Logic', both repr. in B. Brody and R. Grandy (eds), *Readings in the Philosophy of Science* (Englewood Cliffs: Prentice-Hall, 1989), pp. 279–87, 288–308, contain straightforward accounts of the difference between *c** and *c*†, and of the alleged superiority of the former.

[6] This table is modified from Carnap, 'Statistical and Inductive Probability', p. 285.

How should we choose between c^* and $c^\dagger$? Carnap argued that c^* was preferable because it allowed for the possibility of 'learning from experience', while $c^\dagger$ did not. What he meant was this: suppose you discover that objects a_1 and a_2 both have the property F. How does this affect the probability that a_3 has property F too? For convenience, let '*e*' denote the statement 'a_1 and a_2 have property F', and let '*h*' denote the statement 'a_3 has property F'. So the task is to compare P(*h*), the prior probability of *h*, with P(*h*/*e*), the probability of *h* given evidence *e*. The values of P(*h*) and P(*h*/*e*) depend on which confirmation function we adopt. If we adopt $c^\dagger$, P(*h*) is ½, because *h* is true in four states of the world each with probability ⅛. c^* also gives ½ for P(*h*), because *h* is true in four states of the world whose probabilities sum to ½. But c^* and $c^\dagger$ give different answers for P(*h*/*e*). By definition, $P(h/e) = P(h \& e)/P(e)$. Since *e* is true in worlds 1 and 2 only, P(*e*) is ¼ on $c^\dagger$, and 4/12 on c^*. Since *h* & *e* is true only in world 1, P(*h* & *e*) is ⅛ on $c^\dagger$, and 3/12 on c^*. Therefore P(*h*/*e*) = ½ on $c^\dagger$, while it is ¾ on c^*. So if we use method $c^\dagger$, then learning the truth of *e* will leave the probability of *h* unchanged at ½. Whereas if we use method c^*, then learning the truth of *e* will increase the probability of *h* from ½ to ¾. (This difference between c^* and $c^\dagger$ is quite general; it does not depend on the example chosen.) Since Carnap felt that any adequate inductive method must allow the possibility of learning from experience, he was led to reject $c^\dagger$ in favour of c^*.

The relevance of this example is as follows. $c^\dagger$ is clearly a non-trivial probability distribution: it distributes probability over propositions that are not logical truths or falsehoods – in the example above, propositions about whether objects in the world do or do not possess a certain property. But $c^\dagger$ is *not* 'induction-supporting': this was precisely Carnap's objection to it. If you use $c^\dagger$, then when you learn that certain objects in the universe possess a given property, this will have *absolutely no influence* on the probability you assign to other objects' having that property too. If we adopt the standard characterization of 'induction', which Lange appears to accept, as the practice of assuming that unobserved objects will resemble observed ones, then $c^\dagger$ clearly does not 'support induction'. Therefore Lange's equation of 'non-trivial' probability distributions with 'induction-supporting' distributions is incorrect. It is not possible to distinguish between probability distributions that do and do not 'support induction' merely by considering whether or not the distributions assign probability to contingent propositions/events.[7]

University of York
Universidad Nacional Autonoma de Mexico

[7] Thanks to Marc Lange and two anonymous referees for comments and discussion.