L&PS – Logic and Philosophy of Science
Vol. V, No. 1, 2007, pp. 33-39

Reply to Hudson: "Howson on Novel Confirmation"

Colin Howson
Department of Philosophy, Logic and Scientific Method, LSE
e-mail: c.howson@lse.ac.uk

ABSTRACT. In a recent paper in this journal Robert G. Hudson (2006) criticises a discussion originally by me, later elaborated in a jointly authored work with Allan Franklin, of an example due to Maher (1988) which Maher used to motivate a Bayesian solution to the prediction-versus-accommodation problem. Hudson extends his critique to an explanation by us why we thought that Mendeleev's discovery of the Periodic Table of the chemical elements was not susceptible to the same analysis as that we gave of Maher's example. In what follows I shall rebut his charges and show that they rest on a mixture of inattention to the text and some elementary logico-mathematical errors.

Maher's example contrasts two scenarios. In one, call it A, a subject, call him/her Pat, predicts the outcomes of 100 tosses of a coin. In the other, B, Pat waits to be informed of the outcomes of the first 99 tosses before "predicting" the entire 100. In assessing the confidence one should have in Pat's predictive powers one might want to consider the possibility that (s)he possesses some reliable method of prediction, and we shall in particular consider the possibility that he has some such method whose reliability is perfect. Following the notation in Howson and Franklin (1991), which Hudson adopts together with our own formulas, let us call that hypothesis *m*. Let *h* be Pat's prediction of the sequence of 100 outcomes (so *h* specifies the values of a 100-member sequence of *H*s and *T*s), and let *e* describe the result of the first 99 tosses, which we suppose are as described in *h*.

Intuitively, we are inclined to feel that in case A Pat's having genuinely predicted *e* lends quite strong support to *m* and thereby enhances the probability of Pat's prediction of the 100th outcome. In case B, on the other hand, *e* provides no such support either to *m* or to the prediction of the 100th outcome. Using a rather complex argument, Maher concluded that this intuition can be repre-

Published by L&PS – Logic and Philosophy of Science
http://www.units.it/~episteme/ – ISSN: 1826-1043

sented within the Bayesian theory of prior and posterior probabilities. Franklin and I presented a rather simpler way of doing this,[1] whose elegance is extolled by Hudson simultaneously with his declaring that the way we use it is "confused" (an allegation he repeats several times). In what follows I will show that Hudson's charges are without foundation, and that his own presentation is vitiated by inattention to our text – a text which I think I can fairly say went to extreme lengths in attempting to avert any risk of misunderstanding – combined with elementary errors of logic and understanding.

The formal Bayesian argument proposed in Howson (1988) to support the intuitive discrimination between the two cases, and repeated in more detail in Howson and Franklin (1991), is very simple. Take case A first. Let $h(100)$ be Pat's prediction of the 100^{th} outcome. Given e, $h(100)$ is clearly equivalent to h, and so we can represent $P(h(100)|e)$ simply as $P(h|e)$. Using some simple bits of probability theory and assuming that $P(m\&e)$ and $P(\neg m\&e)$ are nonzero, we now expand $P(h|e)$ as follows:

(0) $P(h|e) = P(h|m\&e)P(m|e) + P(h|\neg m\&e)P(\neg m|e)$.

I now quote from Howson and Franklin (the reason for the explicit quotation will be apparent very shortly), "that the subject predicted *h* is now part of the background information relative to which P [...] is computed [...] [so] *m* entails *h*" (1991, p. 576). Letting K_A describe this background, we therefore have $K_A \Rightarrow (m \rightarrow h)$ (read "$\Rightarrow$" as "entails"). It follows that $P(h|m\&e) = 1$. Hence

(1) $P(h|e) = P(m|e) + P(h|e\&\neg m)P(\neg m|e)$.

Also, we can take $P(e|\neg m)$ to be very small, while $P(e|m)$ is 1. Assuming $P(m)$ is not completely negligible, it follows by a standard Bayes Theorem argument that $P(m|e)$ is close to 1 and $P(\neg m|e)$ is close to 0. Hence $P(h|e)$ is also close to 1.

In case B things are very different. I quote again from Howson and Franklin:

> The background information can now be represented by the statement: "The subject was informed of the outcomes of the first 99 flips of the coin, and asserts the conjunction of these with the prediction that the 100^{th} will be a head". The background information does not specify what the outcomes of the first 99 flips were, and so *m* does not entail *h* or *e* relative to that information (although *e*&*m* entails *h*) (1991, *ibid.*).

[1] Though it replicates that in Howson (1988).

The background information, K_B, is specified in this way because we want to be able to represent the possibility that whatever the outcomes of the first 99 tosses are, Pat will incorporate that data into his/her own information stock as a basis for the prediction of the outcome of the 100th toss. Without any loss of generality, therefore, we can represent Pat abstractly as an input-output device $M(x)$, where x is the data, which, for a specific value, e.g. e, of x determines Pat's prediction as $M(e)$. Since m just says "Pat = $M(x)$ is reliable whatever the value of x", while e records one such value, it is quite reasonable to set $P(e|m) = P(e|\neg m)$, from which it follows that $P(m|e) = P(m)$. Also, since e entails e and m entails that $h(100)$ is a head, we have that $K_B \Rightarrow (m\&e) \rightarrow h$, whence $P(h|e\&m) = 1$. Thus from (0) we infer

(2) $\quad P(h|e) = P(m) + P(h|e\&\neg m)P(\neg m)$,

which is approximately equal to $P(h|\text{chance}\&e)$ if $P(m)$ is small.

It might seem pedantic to spell out again what is being assumed in each of the two cases A and B, and we did so in order that the reader could check from themselves that the derivations of (1) and (2) are in order. Hudson, however, affects to find our account "confused" on the ground that

> in their presentation of SCENARIO (A), *there is no mention of what the background specifies as to the exact outcomes predicted by the subject* – and still m is taken to entail h (and thus e). Indeed, given how Howson and Franklin define and use the symbols m, h and e, it does not matter whether the background information specifies what the outcomes are; m entails h (and so e), in any case, for given that the subject has reliable advance information about the outcomes of the 100 flips she will correctly predict h, whether in SCENARIO (A) or SCENARIO (B), that is, whether she was informed about the outcomes of the first 99 flips or not (Hudson 2006, p. 93; my italics).

Hudson has obviously not read carefully, or not understood, what we said, and said I think very clearly, and as a result every single assertion in this quotation is false. But there is worse to come, for he proceeds next to find fault with the plausible claim that, given the circumstances quoted above from our paper, $P(e|m) = P(e|\neg m)$ and hence $P(m|e) = P(m)$:

> But again, given the meaning assigned to m, $P(e|m)$ is surely much larger than $P(e|\neg m)$. However, it is still the case that $P(m|e) = P(m)$ (*Ibid.*).

The first sentence, we know, is false, while the second shows that Hudson cannot perform elementary computations in the probability calculus which, assuming P(h), P(m) are nonzero, is easily seen to pronounce that *it is impossible for $P(e|m)$ to be larger than $P(e|\neg m)$ and for $P(m|e)$ to be equal to $P(m)$*. The first conjunct states that, considered as indicator variables, e and m are positively correlated,[2] while the second states that they are independent. Hudson's conclusion that he has derived (2) "in more obvious fashion" (*ibid*., p. 94) is thus absurd, and the conclusions he wishes to draw from his "derivation" are all invalidated by it.

His dismissal of our argument that the Mendeleev case is radically dissimilar to the coin-tossing example is one such conclusion. Note that (0) is valid for all m, h and e, assuming the unconditional probabilities are all nonzero, and hence we could let m be Mendeleev's theory of the Periodic Table. What Franklin and I had pointed out in our joint paper was that this substitution destroys the asymmetry present in the two possible scenarios of the Maher example, since now m entails h and we immediately obtain (1) in any event, with the straightforward Bayesian corollary that if $P(e|\neg T)$ is small, as it arguably was, there being no alternative explanation around at the time, we could expect $P(h|e)$ to be considerable just for that reason. Hudson's comment about us that

> a Bayesian analysis of the issue, using their own formalism (understood properly), leads to the opposite conclusion [to theirs] (*ibid*., p. 97; Hudson's parentheses)

is therefore not only offensive but simply wrong.

It will be clear, I hope, that Hudson has badly misrepresented what Franklin and I say. Thus it is highly ironic that Hudson himself brings an explicit charge of misrepresentation against me, claiming that I misrepresent the views of John Worrall, my own colleague at LSE, and that I do so moreover by quoting from one of Worrall's own publications! The quotation in question from Worrall, which Hudson reproduces, is this:

> of the empirically accepted logical consequences of a theory those, and only those, used in the construction of the theory fail to count in its support (Worrall 1978, p. 48).

[2] The familiar Bayesian difference measure of support of h by e, i.e. the difference $P(h|e) - P(h)$, is easily seen to be proportional to the degree of correlation between the indicator variables **h** and **e** (these take the value 1 on states for which the corresponding proposition is true, and 0 on states for which it is false).

Fairly unequivocal, one might think, except for the ambiguous phrase "used in the construction of the theory". What exactly does "used" mean? In general it is difficult to give any clear answer, though there is one situation which occurs commonly in science where the meaning is relatively clear, and that is where a theory has adjustable parameters and the data fix the values of one or more of them. Indeed, in Howson (1990) I discuss an example Worrall uses to support his claim in the paper from which the quotation was taken, the use of Mercury's anomalous perihelion – anomalous for CGT, classical gravitation theory – to fix an appropriate parameter in CGT, like the density of a dust cloud, say, to account for the anomaly. I showed, as a simple exercise in the probability calculus, that if two rival theories h and $h(a)$ both predict e, but e fixes the parameter a in $h(x)$, then, using the familiar Bayesian support measure given by the difference between posterior and prior probabilities, two interesting features are seen: (i) the posterior probability of $h(a)$ is equal to the *prior* probability of $h(x)$, and (ii) the ratio of the support of h to $h(a)$ is equal to $[P(h)/P(h(x))].P(e)^{-1}$. To be precise, it is easy to show, given a mild independence assumption, that if $S(h,e)$ signifies the support of $h(a)$ by e given by the difference between posterior and prior probabilities, then

$$S(h,e) = P(h(x))P(\neg e).$$

This simple decomposition of $S(h,e)$ in the circumstances cited is an important feature of the difference measure (and one as far as I am aware unknown before I exhibited it). It accords very closely with the intuition that a hypothesis whose parameters have been adjusted to the data should be exactly as probable given the data as the prior probability of the unadjusted hypothesis, and that there should be a bonus of support to the genuinely predictive hypothesis if both it and $h(x)$ start out with equal priors. In this case of the rivals above we see that the bonus will be exactly equal to $P(e)^{-1}$, so the more unlikely a priori the prediction is to be true the more support accrues proportionately to the predictive hypothesis.

It follows that Hudson's claim that "[Howson] hasn't adequately explained what evidential value there is in prediction *per se*" (2006, p. 99) is as far from the truth as his other charges. If the hypothesis with free parameters and its genuinely predicting rival start off equal, we see from the result above that the latter gets more support than the former once its parameters have been fixed to allow it to make the same predictions. In other words, there *is* a virtue in prediction per se, but it can be dominated by the prior implausibility of the predicting hypothesis. Indeed, depending on the prior probabilities of h and $h(x)$, the accommodating hypothesis $h(a)$ could turn out to be *more* supported than $P(h)$. Hudson thinks this is wrong and that the predictive hypothesis should never receive less support than the accommodating one, but a little thought should convince anyone that this must be wrong. For an extreme example, suppose that $P(h) = 0$ and $P(h(x)) > 0$. If the prior probability of the predictive hypothesis is actually zero then it should not garner any support, while the accommodating hypothesis might, depending on the circumstances, pick up some. I point out in Howson (1988) that such a case is modelled where there are enough computers outputting different sequences of 1 and 0: one of these machines is certain to predict the 99 coin tosses, though there is a zero prior probability that it has a reliable algorithm for predicting coin tosses in general.

Let us now return to the Worrall quotation above. That it is incorrect is easily seen from the following simple example. I take all the balls, each black or white, out of an urn, and I discover that there are r white and s black. I accordingly evaluate the parameter x in the hypothesis $h(x)$: "The proportion of white balls in the urn is x" as $x = r/(r + s)$. Would anyone seriously deny that the hypothesis $h(r/(r + s))$ is maximally supported by the data (it is actually *entailed* by it)? Yet this is what Worrall's assertion implies. We can, incidentally, gauge the quality of Hudson's scholarship by noting that his claim that I misrepresent Worrall is based on the ground that Worrall might have wanted to qualify it in the context of an example *different* from the one he himself used as evidence for that assertion, namely the alleged fact that "Mercury's perihelion [advance] is not regarded as supporting classical theory", although it is predicted by versions of that theory (Worrall 1978, p. 48). In passing, we can see rather clearly that the conclusion Worrall draws from this observation (that the consequences used in the construction of a theory do not support it) is in fact a *non sequitur*: that the data implied by $h(a)$ do not support $h(x)$ – which is effectively what Worrall notes – does not imply that they do not support $h(a)$ itself (which is what he concludes). As I have shown above, in general the data *will* support $h(a)$.

Hudson's objections against me are both unscholarly and without foundation. I hope, nevertheless, that some of the other conclusions that have emerged from this note will be of more positive interest.

References

Howson, C. (1988): "Accommodation, Prediction and Bayesian Confirmation Theory", in A. Fine, J. Leplin (eds.), *PSA 1988. Proceedings of the 1988 Biennal Meet-*

ing of the Philosophy of Science Association, East Lansing (Mich.), Philosophy of Science Association, vol. II, pp. 381-392.

——— (1990): "Fitting Your Theory to the Facts: Probably not such a Bad Thing after all", in C. Wade Savage (ed.), *Minnesota Studies in the Philosophy of Science*, Minneapolis: University of Minnesota Press, pp. 224-245.

HOWSON, C., FRANKLIN, A. (1991): "Maher, Mendeleev and Bayesianism", *Philosophy of Science*, 58, pp. 574-585.

HUDSON, R. G. (2006): "Howson on Novel Prediction", *L&PS – Logic and Philosophy of Science*, IV, pp. 91-104 (www.units.it/episteme).

MAHER, P. (1988): "Prediction, Accommodation and the Logic of Discovery", in A. Fine, J. Leplin (eds.), *PSA 1988. Proceedings of the 1988 Biennal Meeting of the Philosophy of Science Association*, East Lansing (Mich.): Philosophy of Science Association, vol. I, pp. 273-285.

WORRALL, J. (1978): "The Ways in which the Methodology of Scientific Research Programmes Improves on Popper's Methodology", in G. Radnitzky, G. Anderson (eds.), *Progress and Rationality in Science*, Dordrecht: Reidel.