University of Reading

Explication, description and enlightenment

Book or Report Section

Accepted Version

Schroeder, S. ORCID: https://orcid.org/0000-0002-4480-6458 and Preston, J. ORCID: https://orcid.org/0000-0002-3740-2308 (2019) Explication, description and enlightenment. In: Analysis and Explication in 20th Century Philosophy. Logical Analysis and History of Philosophy, 22. Brill, pp. 106-120. Available at https://centaur.reading.ac.uk/82707/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

Publisher: Brill

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Explication, Description and Enlightenment

In the first chapter of his book Logical Foundations of Probability, Rudolf Carnap introduced and endorsed a philosophical methodology which he called the method of '*explication'*. P.F. Strawson took issue with this methodology, but it is currently undergoing a revival. In a series of articles, Patrick Maher has recently argued that explication is an appropriate method for 'formal epistemology', has defended it against Strawson's objection, and has himself put it to work in the philosophy of science in further clarification of the very concepts on which Carnap originally used it (degree of confirmation, and probability), as well as some concepts to which Carnap did not apply it (such as justified degree of belief.

We shall outline Carnap's original idea, plus Maher's recent application of such a methodology, and then seek to show that the problem Strawson raised for it has not been dealt with. The method is indeed, we argue, problematic and therefore not obviously superior to the 'descriptive' method associated with Strawson. Our targets will not only be Carnapians, though, for what we shall say also bears negatively on a project that Paul Horwich has pursued under the name 'therapeutic', or 'Wittgensteinian' Bayesianism. Finally, explication, as we shall suggest and as Carnap recognised, is not the only route to philosophical enlightenment.

History of the Concept

Carnap took explication to be in the service of the *clarification*, and if possible *definition*, of concepts such as degree of confirmation and probability, and of the nature of induction (Carnap 1950, p.1). He considered such concepts to be well enough understood for practical purposes, but 'hardly ever precisely explained' (ibid., p.2). The use of the terms in question would therefore be made more precise, by being given an 'explication', which, he said,

> consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second. We call the given concept (or the term used for it) the *explicandum*, and the exact concept proposed to take the place of the first (or the term proposed for it) the *explicatum*. (Carnap 1950, p.3).

Note, in particular, Carnap's reference to 'replacing' the *explicandum* by the *explicatum* here (reiterated in Carnap 1963, p.955) – his view was not merely that the latter is a refinement of the former. Carnap also clearly allows that the *explicandum* may be either a concept of everyday language, or one from a previous stage in the development of 'scientific language'. There will be a good

mutual understanding of its meaning but the term will be '*unsystematic'* and '*inexact'* (ibid., p.4, emphasis added), and our understanding of it will be 'far from perfect theoretically' (ibid., p.5).

The *explicatum*, however, 'must be given by explicit rules for its use, for example, by a definition which incorporates it into a well-constructed system of scientific… concepts' (ibid.). Whereas in ordinary scientific problems both the datum and the solution are already formulated in exact terms, in a problem of explication the *explicandum* is not. According to Carnap, this means that a proposed explication can't be said to be right or wrong, since there can be no clear-cut answer. An explication *can*, though, be said to be 'satisfactory', more satisfactory than another, etc.

Carnap details the following ways in which the new *explicatum* must correspond to the old *explicandum*. The former must be *similar* to the latter 'in such a way that, in most cases in which the *explicandum* has so far been used, the *explicatum* can be used' (ibid., p.7). The former must be *fruitful* (ibid., p.6), that is, more fruitful than the latter. For scientific concepts, at least, this means it can be 'used for the formulation of laws' (ibid., p.6), either *empirical* laws (for nonlogical concepts) or logical *theorems* (for logical concepts) (ibid., p.7). In addition, as we already saw, the new concept must be *exact*, that is, one must give exact rules for its use, e.g., a definition (ibid., p.7). And finally, it must be *simple*, in the way that scientists appreciate, where this means either that the

form of its definition is simple, or that the forms of the laws connecting it with other concepts are simple (ibid., p.7).

P.F. Strawson's Objection to Carnap

P.F. Strawson, in the Carnap volume of the Library of Living Philosophers (1963), compared Carnap's method of 'rational reconstruction' with his own descriptive method, and took issue with the former. He noted, first, that there could hardly be a dispute between the two as long as one understood Carnap's method to be merely 'the right means of getting an idea into shape for use in the formal or empirical sciences' (Strawson 1963, pp.504-5). This is an important concession. Strawson exhibited a healthy scepticism about the extent to which scientists *do* actually use concepts that have been formalized in the way that logicians prefer, but he conceded that there may be no objection to replacing a pre-scientific concept with an exact and fruitful one *for certain purposes*, i.e. for the purposes of science (Strawson 1963, ibid).

Strawson also conceded, rightly, that an explication *can* be helpful in philosophy as 'a model object of linguistic comparison' (ibid., p.513). That is, one might illuminate aspects of an everyday concept by constructing a formalised concept which one then compares with its original. But it's vital to recall here that explications are intended to be *replacements* for everyday

concepts, not merely objects with which to compare them. Carnap's method, as he explicitly conceived it, is not reinstated by this concession.

Strawson's central complaint about Carnap's method was focussed on concepts which have a life outside science. To offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, he complained, 'is to do something utterly irrelevant – a sheer misunderstanding, like offering a textbook on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart' (ibid., p.505). Concepts used in non-scientific discourse could not literally be replaced by scientific concepts that served the *same* purposes, at least partly because scientific uses of language are highly specialized, and language has many *other* kinds of uses. The general lesson he drew was that '[t]he kinds of concept we employ are not independent of the kinds of purpose for which we employ them; even though some concepts can fulfil more than one kind of purpose' (ibid., p.506). An attempt to lay down exact rules of use for concepts that will be fruitful in a particular science, therefore, can be no solution to a typical philosophical problem concerning concepts more typically used in non-scientific or meta-scientific discourse, but *only a changing of the subject*. Even allowing, as Carnap might well insist, that the distinction between scientific and non-scientific language is by no means sharp, Strawson does make an important point.

Carnap's Response to Strawson

Carnap replied to Strawson by complaining, first, that Strawson presupposed a sharp separation between everyday and scientific concepts, where there is in fact only a continuum (p.934). This complaint, though, is ungrounded. Strawson's contrast between 'concepts of daily life' (p.503), 'pre-scientific concepts' (p.504), 'concepts used in non-scientific kinds of discourse' (p.505), and 'non-scientific concepts' (p.505), on the one hand, and 'scientifically exact and fruitful concepts' (p.504), 'scientific concepts' (p.505), on the other, no more presupposed such a sharp separation than Carnap's own way of distinguishing the two. Both parties can allow that the distinction is a matter of degree, but both ought also to agree that some of the concepts that most concern philosophers (knowledge, truth, identity, justice, virtue, etc.) are concepts employed by almost all people, and thus not scientific concepts.

Carnap's second complaint was that Strawson was wrong to believe that philosophical problems are raised by people who know very well how to use the concepts in question. Carnap insisted that although such people *believe* they know how to use everyday concepts, they 'often deceive themselves' in so believing (ibid., p.935), and that the philosopher can help them by showing them that something is wrong with their use of the expressions in question. In describing this possibility, though, Carnap seems to give the game away. For what he says is that 'it may occur that in certain critical contexts the ordinary

usage leads to difficulties, unanswerable questions, even contradictions, demonstrating the surprising fact that people are not completely clear about the correct usage. A misinterpretation of the expressions describing motion led to Zeno's paradoxes…' (ibid., p.935). But in the case of motion it was indeed arguably such a misinterpretation of these expressions, by a philosopher, that generated those problems, not their unreflective use in ordinary contexts. Further, Carnap is too quick to assume that because a concept can lead to a philosophical problem in *some* 'context' (i.e. a 'critical', philosophical context), there is something *wrong* with that concept, and that those who unproblematically use such a concept in everyday contexts only *believe* they know how to use it (Carnap 1963, p.935). One cannot be faulted for not knowing how to use an instrument if there is *some* context, some context into which philosophers put it, in which it gives problems. And its causing a problem in some such 'critical' context doesn't mean there is something wrong with it. Further, if Carnap's aim was really to design concepts which cannot yield problems no matter what philosophical context they might be used in, that may well be Utopian. Carnap supplies no reason for thinking that there could be such concepts and, the development of the kind of philosophy that Carnap had in mind has shown that, as one recent commentator puts it, 'it is well-nigh impossible to devise a formal language that will prevent the formulation of metaphysical questions while permitting us to engage in science, not to mention other forms of discourse' (Glock 2002, p.89). The history of formal and

scientific concepts strongly suggests that they are in no way immune to contracting the kinds of problems that Carnap is concerned to diagnose.

Carnap misses or ignores the distinction (familiar in analytic philosophy from G.E. Moore onwards) between knowing the meaning of a term in the sense of being able to use and understand it (at least in ordinary contexts), and knowing its meaning in the sense of being able to give a correct analysis of that meaning, being able to give an account of what its meaning is. Because one can quite well have the former without the latter, defects in one's grasp of the latter don't impugn one's grasp of the former. The kind of 'connective' philosophical analysis that Strawson later explained (Strawson 1992) is a contribution to a more systematic and higher-level perspective on the meaning of terms and utterances, of whatever degree of scientificity, but it does not impugn the use of such terms and utterances by less reflective speakers.

Maher's Defence against Strawson's Objection

Maher attempts to give a better reply to Strawson on Carnap's behalf. His idea is that the construction of *explicata* can assist in solving a problem in ordinary language since an explication can give us new concepts that can serve the same purposes as the old concepts that created the puzzles (Maher 2007, p.332). He begins by insisting, slightly more strictly than Carnap, on the condition of

sufficient similarity: a good explication must be similar enough to the *explicandum* that it can be used for the same purposes (ibid., p.333). If we are trying to determine whether a particular sentence, *S*, formulated in ordinary language, is true, we can construct *explicata* for the concepts in *S*, formulate a corresponding sentence *S′* using these *explicata*, and then determine whether or not *S′* is true. Although this doesn't *by itself* solve the original problem, Maher argues that it can assist in doing so, in three ways:

> (1) The attempt to formulate *S′* often shows that the original sentence *S* was ambiguous or incomplete and needs to be stated more carefully. (2) If the *explicata* appearing in *S′* are known to correspond well to their *explicanda* in other cases, that is a reason to think the truth-value of *S* will be the same as that of *S′*. (3) We can translate the proof or disproof of *S′* into a parallel argument about the corresponding *explicanda* and see if this seems to be sound; if so, we obtain a direct argument for or against *S*. (Maher 2007, p.334).

There are certain worries about this procedure, though. Let's take each of the three suggestions separately. (1) First, the focus on *sentences* is problematic. It's hardly ever a sentence, an abstract linguistic entity, that provides a philosophical problem, but a sentence put to a certain use, namely to make a statement in a given context. And very often it is exactly the semantically

relevant aspects of the context that stand in the way of any useful formalization. To assume that philosophical problems consist in questions of the truth or falsity of *sentences* already tilts the ground in favour of a formalist approach.

Secondly, Maher is right to think that if an ordinary statement can be shown to be ambiguous, it needs to be stated in an unambiguous way. But such a statement's being incomplete or vague does not necessarily mean that it 'needs to be stated more carefully'. The correct response to a philosophical statement, *p*, framed in such terms may simply be that *p* is neither true nor false, since nothing has been laid down to determine the matter. Consider, for example, some of the well-known philosophical puzzle cases of personal identity: suppose some mechanism turned a person into two indiscernible copies (Derek Parfit's 'bifurcation'), would they both be identical with the original person? We don't know what to say. Our concept of personal identity is not prepared for such science fiction cases, hence there is no correct answer. Does it follow that our concept is defective and needs to be replaced? No. Concepts are shaped according to our needs, and as long as our concept of personal identity is suited to our communicative needs in our current environment, there is nothing wrong with it. The fact that it is not a scientific concept is no objection to it. This is not to say that concepts cannot have defects, but the kind of concepts in which philosophers are often most interested, concepts like knowledge, truth, identity, do not earn their very basic place in our conceptual scheme by virtue of living up to ideals of preciseness which philosophers devise.

A concern also arises with respect to suggestion (2), that if the *explicata* appearing in *S′* are known to correspond well to their *explicanda* in other cases, that is an (inductive) reason to think the truth-value of *S* will be the same as that of *S′*. After all, the very idea of explication is to replace a concept that we find unsatisfactory, one that is resisting our attempts to resolve a philosophical issue, by one that is more accommodating in this respect. Hence, for explication to work, the two concepts must differ in crucial respects — so the truth-conditions of *S* and *S′* are likely to differ. If there had been widespread agreement in the truth-values of *S* and *S′* to begin with, we would not have needed an explication. Typically, though, there will be an indeterminacy in the case of *S* where *S′* yields definite answers, some of which may also be different from the definite answers we can give about *S*. But where a concept is too indeterminate for it to be possible to say whether it applies in a given case, induction to other instances is obviously not applicable, and if in a certain context we do not know how to use '*S*', *S* cannot have the same truth-conditions as *S′*.

What about suggestion (3), that we can translate the proof or disproof of *S′* into a parallel argument about the corresponding *explicanda* and obtain a direct argument for or against *S* if this proof is sound? Again, this is unlikely to work because what allows us to give a proof or disproof of *S′* is probably the neat determinacy of the *explicatum*, the lack of which makes the original concept, and hence *S*, so intractable.

Otherwise, if the difference between S and *S′* isn't relevant to the argument in question, it's hard to see why the detour through explication should be necessary: we might as well go straight for the 'direct argument for or against *S*'. Even if the prior considerations of the *explicatum* and *S′* are heuristically useful (if they help one to think of an argument for *S*), there is no need for them to be published. As long as the ultimate argument is sound and complete on its own, the roundabout thoughts that helped me come up with it can well remain buried in my drafts or notebooks.

No doubt Maher would protest that the detour of explication is not merely of biographical interest, but provides a useful methodology, which deserves to be published for emulation in other cases. But to make this idea good one would have to show how the dilemma sketched above could be overcome. Either the *explicatum* is considerably different from the *explicandum*: then we might indeed find a proof for *S′* which is not available for *S* – so that it doesn't help us with the original problem. Or the difference between the two concepts is so small that it doesn't affect either the truth-conditions of the respective statements or the possible arguments for or against them, in which case it's hard to see how it can still make a heuristic difference to change from one to the other.

A further aspect of Maher's approach becomes clear in his reply to Antony Eagle's accusation that Carnap thought that an ordinary concept, once 'explicated', is of no further use (which Carnap's talk of 'replacement' surely

suggests). Maher replies that Carnap meant only that the *explicatum* replaces the *explicandum in certain contexts*, or *for certain purposes*, not in all contexts or for all purposes. This may be so, but it introduces a very important qualification to Maher's original idea that explications give us new concepts that 'serve the same purposes' as the originals (Maher 2007, p.332). We are now being told that only *some* of these purposes are served. To that extent, of course, the condition of sufficient similarity is going to be weakened.

We feel that Maher is wrong, then, to quote Strawson's remark that an explication can be helpful in philosophy as an object of linguistic comparison, as evidence that he [Strawson] 'seems to concede that explication can be useful in something like the ways [Maher has] indicated' (Maher 2007, p.334). As the context of Strawson's remark makes clear, at least when it comes to concepts that have a life outside science, he did not believe that formal concepts are at all likely to match our original non-formal concepts in the philosophically relevant respects:

> For the types or modes of logical behaviour which ordinary concepts exhibit are extremely diverse. To detect and distinguish them is a task in which one may well be hindered rather than helped by fixing one's eye too firmly on the limited range of types of logical behaviour which the concepts occurring in a formal system can there be shown to display (Strawson 1963, p.513).

Hence, if formal concepts can sometimes be useful as objects of comparison it is rather in virtue of the way they *differ* from our ordinary concepts. Far from matching our ordinary concepts in the way Maher's methodology assumes ((2) & (3)), formal concepts can at best be useful to throw into relief the contrasting peculiarities of our ordinary concepts, which often give rise to philosophical problems (and which, as Strawson emphasises, are bypassed by Carnap's programme of explication).

Moreover, even if we assume that, in a given case, a formal concept is indeed so similar to the concept it is meant to replace that, as Maher envisages, the truth-conditions of *S* and *S′* are likely to be the same (2), and an argument for or against *S′* can be transformed into an equally convincing argument for or against *S* (3);— even then, it might be argued that these procedures do not afford an alternative to ordinary language analysis, but presuppose it. For, as Strawson points out, it all depends on a suitable similarity between the unconstructed and constructed concepts, and hence between *S* and *S′*, which can only be established by first of all 'accurately describing the modes of functioning of the unconstructed concepts':

> But this task is precisely the task of describing the logical behaviour of the linguistic expressions of natural languages; and may by itself achieve the sought-for resolution of the problems and difficulties rooted in the

> elusive, deceptive mode of functioning of unconstructed concepts (Strawson 1963, p.513).

Hence it is problematic to hold, that explication, as reconstructed by Maher, might provide 'an appropriate methodology for doing formal philosophy' (Maher 2007, p.340). That is to say, the use of formal methods cannot amount to a distinct type of philosophy. It can, at best, play a subordinate auxiliary role in a philosophical investigation that must mainly and essentially be an investigation of statements made in natural languages.

Hempel's Paradox and Nicod's Condition

So far we hope to have reinforced Strawson's judgement that the method of explication still has some problems of principle. In this section we would also like to suggest that it may be inferior to descriptive methods in an actual application to a philosophical problem, this problem being the notorious paradox of the ravens (Hempel's paradox).

Hempel's paradox is the conclusion that the generalisation 'All A's have property P', can be confirmed (albeit to some very small extent) by the observation of things that are not A's and do not have property P. This is the problem to which Maher has most strenuously sought to apply the method of

explication (Maher 1999, 2004, 2005), and the problem whose solution Horwich (who accepts the same explication of degree of confirmation as Maher does) advertises as one of the main successes of 'Wittgensteinian Bayesianism' (Horwich 1993, esp. p.616). It is also, however, a problem that one of us has addressed in print, from a more Strawsonian perspective (Schroeder 2009).

The crux of the issue is the idea, known as 'Nicod's condition' (or 'criterion'), that

(N) A generalization is confirmed by any of its instances.

This is one of a famous triad of propositions that together generate the conclusion that observation of a non-black non-raven *confirms* (to some very limited extent) the generalization 'All ravens are black'. Maher's suggested solution of this paradox is as follows: Having put forward a Bayesian explication of the concept of scientific confirmation, he argues, first, that Nicod's condition does not hold when there is no background evidence; and he claims that that already suffices to resolve the paradox (Maher 2004, pp.78f.). Moreover, he employs his *explicatum* of confirmation (i.e., a particular Bayesian account of probabilities) to argue that the counterintuitive conclusion of Hempel's paradox is indeed correct: that the hypothesis 'All ravens are black' can indeed be confirmed by the observation of a non-black non-raven (Maher 2004, p.79).

This latter view is shared by Horwich, who shows no inclination to reject Nicod's condition, but maintains that whereas observation of a black raven confirms the generalization *substantially*, observation of a non-black non-raven confirms it only *negligibly* (Horwich, pp.608, 610). And his diagnosis of the ravens paradox is that its conclusion has a 'paradoxical flavor' because we tend to confuse negligible support with no support at all (pp.608-9). That's to say, really, that the conclusion is *not* paradoxical: observation of non-black non-ravens *does* confirm 'All ravens are black', but it does so only to an insignificant extent.

Let us first consider Maher's diagnosis of the falsity of Nicod's condition, which runs as follows (Maher 2004, pp.77ff.). His explicated version of Nicod's condition, *N'*, is that

(*N'*) C(*A*, *F*a & *G*a, *D*),

which means that the generalisation

(A) All *F* are *G*

is confirmed by any instance of 'This *F* is a *G*', given background evidence *D*.[1]

Maher shows that *N'* doesn't always hold in the case where there's *no* background evidence (which he represents by the situation where *D* is a logical truth) (Maher 2004, p.78). He does so by constructing a case where the generalization (now restricted to a population of just two) is not only *not* confirmed by the evidence (that one of the two objects is *F* and *G*), but is *disconfirmed* by that evidence. His informal explanation of this situation is as follows:

> According to standard logic, "All unicorns are white" is true if there are no unicorns. Given what we know, it is almost certainly true. But now imagine that we discover a white unicorn; this astounding discovery would make it no longer so incredible that a nonwhite unicorn exists and hence would disconfirm "All unicorns are white" (Maher 2004, p.78).

Note that this informal example hinges on the fact that we know *from past experience* that unicorns are very unlikely to exist. So, contrary to what Maher intends, this is not a case involving no background evidence. In addition, its relevance to the ravens scenario is dubious, since there we do know that ravens exist.

[1] It is perhaps worth pointing out that Maher uses the word 'evidence' in an uncommonly broad sense covering what would normally be called 'knowledge'.

Returning, then, to that ravens scenario, note that Maher fixes the initial probability of an object's being a raven at 0.001, while the probability of something's being black is supposed to be 0.1. These are not, as one would expect in probabilistic calculations, relative frequencies. After all, the population is only 2 and its constitution is not known. So how are we to come by these probabilities? — Three points emerge: Again, first, this is not in fact a case that doesn't involve any background knowledge (or 'evidence'). If the given probabilities of sighting a raven or sighting something black are not to be entirely fanciful, they must be based on past experience — what would be called background knowledge.

Secondly, even if our past experience allows us to say, roughly, that seeing a raven is much less likely than seeing something black, any precise quantification of such respective probabilities (as, say, 0.001 vs. 0.1), which is not simply an expression of relative frequencies, must be fairly arbitrary – there is no non-arbitrary reason for having any specific belief about the likely frequency of those occurrences. But this means that the appearance of quantified precision displayed by the results is illusory.

Thirdly, even if the numerical probabilities used in these calculations are not just arbitrary quantifications of vague impressions, but are based on relative frequencies observed in the past, it would still be arbitrary and illicit to apply them to this case. Supposing that on some other occasion a given population of 10,000 objects contained 10 ravens and 1,000 black objects,— why should we

expect this unspecified population, of only 2 objects, to display the same relative frequencies of ravens and black objects? For that to be at all rational we would need some reason to believe, either, that ravens and black things are everywhere evenly distributed (which is not the case), or that the two populations in question are likely to have similar constitutions. Since no such reason has been given, the numeric probabilities suggested must be dismissed as arbitrary.

In any case, Maher is wrong to think that his counterexample to Nicod's Condition suffices to resolve Hempel's Paradox. All his counterexample shows is that in the rather special case where we initially believe the hypothesis 'All *F*s are *G*' to be trivially true due to the non-existence of *F*s, the observation of an *F* that is *G* may not confirm the hypothesis. This does not show that Nicod's Condition does not hold in the more usual (and more interesting) cases where *F*s are known to exist, such as 'All ravens are black'. In other words, Hempel's Paradox requires only a weaker version of Nicod's Condition, in which the general statement 'All Fs are G' is taken to have existential import; and this weaker version of Nicod's Condition is not affected by Maher's examples.[2]

Maher's criticisms of Nicod's Condition do not suffice to resolve the ravens paradox, and indeed in the end he is happy to accept its paradoxical conclusion. He maintains that, in his example and according to his Bayesian

[2] Maher also observes that Nicod's Condition does not hold if our background knowledge includes the proposition 'If this *a* is *F* and *G*, it is not the case that all *F* are *G*' (Maher 2004, p.77). But again, that does not help us with the resolution paradox of the ravens where no such background knowledge seems to be available.

calculations, the observation of a non-black non-raven does indeed confirm the hypothesis (A) that all ravens are black; and continues:

> This is contrary to the intuition of many, but when we understand the situation better it ceases to be unintuitive, as I will now show. … We see that
>
> $$p(Fb.\sim Gb|\sim Fa.\sim Ga) = 0.0006 < 0.0009 = p(Fb.\sim Gb)$$
>
> and so $\sim Fa.\sim Ga$ reduces the probability that *b* is a counterexample to "All ravens are black." This should not be surprising. In addition, $\sim Fa.\sim Ga$ tells us that *a* is not a counterexample to "All ravens are black," which *a priori* it might have been. So, for these two reasons together, it ought not to be surprising that a nonblack nonraven can confirm that all ravens are black. (Maher 2004, p.79)[3]

Now what exactly are Maher's 'two reasons' to accept the counterintuitive conclusion of the Ravens Paradox? The first one appears to be that, according to his Bayesian explication, seeing a non-black non-raven reduces the probability of seeing a non-black raven next. But why should that be a reason to accept his counterintuitive Bayesian explication in the first place?

[3] In this quotation, we have replaced Maher's more complicated abbreviations with simpler equivalents, in order to make the example easier to follow.

(The fact that up to a point a model agrees with our intuitions doesn't make its implausible consequences any less implausible.)

The second reason he offers is that, intuitively, the observation of a non-black non-raven confirms the hypothesis that all ravens are black in so far as it tells us that the object is not a counterexample to it, which *a priori* it might have been. This idea has been considered and criticised in Schroeder's article, to which we now turn.

Schroeder's article gives a treatment of the ravens paradox from the perspective of the method of conceptual analysis. That treatment suggests that Nicod's Condition is false: only *law-like* generalizations are confirmed by their instances, whereas *accidental* generalizations are not (Schroeder 2009, p.250). So, for example, even if the *F* we are considering is the very first *F* we have encountered, no observation of the form 'this *F* is *G*', such as 'This French postman is an alcoholic', would be taken, on its own, to confirm the corresponding generalization 'All French postmen are alcoholics' (ibid.), unless there is some grounded suspicion of there being a law or causal relation in the offing.

Further, even law-like generalizations aren't confirmed by their instances if those generalizations are *absurd* in the light of our causal background knowledge (Schroeder ibid., p.256). Observation of a non-black non-raven, for example, doesn't (*pace* Horwich) confirm the generalization that all non-black things are non-ravens (and thus, *a fortiori*, doesn't confirm to any extent

whatsoever the generalization that all ravens are black), simply because our background knowledge of causal regularities makes it unreasonable to believe that being a non-black thing has any causal connection with being a non-raven.

Schroeder's diagnosis of what has gone wrong in the ravens paradox is that a misleading model has been presupposed. The world is being treated as a single big bag '*out of which we draw one object after another in such a way that any object is equally likely to be drawn at any time*' (Schroeder ibid., p.261). But the world is not like that. It's not the case that the more non-black non-ravens we see, the likelier it becomes that all non-black objects are non-ravens. To treat the paradox, Schroeder suggests replacing this with a more appropriate model:

> Our everyday observations of objects in the world… could more plausibly be likened to drawing objects out of various *different* bags, of which, for the most part, we already know what their likely contents are. One of them contains birds, but most of them contain quite different kinds of things, such as banknotes, carpets or blades of grass. Evidently, a draw from a bag known to contain no birds (e.g. my wardrobe) can neither instantiate [the hypothesis that all ravens are black], nor can it provide a counterexample to it. But if it is known from the outset that such a draw cannot provide a counterexample to the hypothesis, then its not doing so cannot strengthen the hypothesis either (it cannot be regarded as

indirectly relevant as an elimination of a potential counterexample). In short, the contents of a bag known to contain no birds have no bearing on the hypothesis that all ravens are black (Schroeder ibid., p.261).

Applying the method of explication to Hempel's Paradox, Maher and Horwich replace the scientific concept of confirmation (of a scientific hypothesis) with a Bayesian concept of raising a statement's mathematical probability. Having replaced the scientific concept of confirmation with a Bayesian one, Horwich (in agreement with Maher, and other Bayesians)[4] declares that to deny that a white shoe confirms the hypothesis that all ravens are black is a 'confusion of negligible support with no support at all' (Horwich 1993, pp. 608f.). But in fact the confusion is one of two crucially different *concepts* of confirmation:

(i) Scientific confirmation is part of a triad of concepts: An observation can *confirm* a hypothesis, *disconfirm* it, or be *irrelevant* to it (cf. Hempel 1965, p.5). According to the Bayesian concept or explication of the concept, however, there are virtually no neutral or irrelevant observations: *every* observation will either 'confirm' or 'disconfirm' a given statement. Hence, the Bayesian *explicata* have a much larger extension than the scientific concepts they are meant to replace, and so the Bayesian classification will frequently be at odds with scientific practice. Scientists, we suspect, would not agree that, and do not

[4] Other Bayesians offering a similar account include Colin Howson and Peter Urbach (Howson & Urbach 1993, p.127), and Branden Fitelson (Fitelson 2006).

act as if, red herrings confirm generalisations to *any* extent. The concept of confirmation in question here is not an *everyday* concept, nor yet a scientific concept, but a methodological concept used by scientists. It would therefore be inappropriate to assess it by reference to any everyday notion of confirmation, but our argument here is rather that it fails to live up to the way scientists use that concept.

(ii) Scientific confirmation of generalisations is essentially confirmation of law-like hypotheses. That means, for one thing, that they are not limited to a finite number of instances. And for another thing, it means that the scientific confirmation of a hypothesis raises the probability of all its instances. Maher's Bayesian 'confirmation' is crucially different in both respects.

First, as Maher admits, his calculations of probability are usefully applicable only to finite populations.[5] According to his Theorem 9, no evidence can 'confirm' an unlimited universal generalisation (Maher 2004, pp.76f.). Yet unlimited universal generalisations are one of the main concerns of scientific theories. Newton's First Law, for example, is not meant to hold only for a finite number of bodies.

Secondly, the rise in probability that Bayesians call 'confirmation' does not require the hypothesis to be law-like, whereas scientific confirmation of generalisations does. If it is thought to be only accidental that all objects of a

[5] This is not typical of Bayesian accounts, so our objection here will not generalise to other such accounts of confirmation.

given set *F* have the property *G*, then there is no room for scientific confirmation. Learning that one of the *F*s is *G*, or even that most of them are, will not make it any more probable that the remaining ones are. For example, if nine of a series of ten throws of a true die are sixes, there will still be only a 1/6 chance that the tenth throw will be a six. Of course, the general statement 'All ten throws of this series will be sixes' becomes less unlikely with every six you throw. Initially, its probability is only $(1/6)^{10}$; after nine sixes it has increased to 1/6. Bayesians like Maher may be happy to call this 'confirmation', yet scientists are not. As Nelson Goodman astutely observed: 'Confirmation of a hypothesis occurs only when an instance imparts to the hypothesis some credibility that is conveyed to other instances' (Goodman 1983, p.69). When scientists ask whether a given generalisation is confirmed by the evidence they mean: does the evidence make it likely that the generalisation is or reflects a law, or laws, of nature? Unlike Bayesians, they are not interested in generalisations that would only be accidentally true.

(iii) Scientific confirmation is essentially a *reason* to take a hypothesis more seriously. And if this conceptual link between confirmation and rational behaviour is not to be severed, confirmation — like reasons — cannot meaningfully be divided into indefinitely small pieces. We can well distinguish between a reason to accept a hypothesis as true and a reason to consider it as possibly true. According to such rough distinctions there can of course be degrees of confirmation. But where an infinitesimal rise in mathematical

probability (e.g. from 1/999,999 to 1/999,998) does not give us any reason to change our attitude towards a hypothesis — to take it more seriously — this rise in probability cannot usefully be called 'confirmation' (in a scientific context). Hence, to introduce numerical degrees of belief according to Bayesian calculations of probability is not an *idealization* of the concept of confirmation, as Horwich suggests, but — when too finely grained, as in the ravens example, just meaningless.

(iv) Maher's Bayesian explication of confirmation is committed to a lottery model of observation: the unrealistic idea that all possible observations are equiprobable draws. (Not to mention its presupposition that the world divides naturally into a distinct and finite number of objects.) The relevance of such scenarios to confirmation *in science* is tenuous, and they cannot possibly be taken as representative of the scientist's situation. —

In conclusion, Maher's flagship example, of Hempel's Paradox and scientific confirmation, does not cast a favourable light on Maher's revival of Carnap's programme. In fact, the case might well be thought to afford an example of how explication, far from resolving the original problem, just 'changes the subject'. If we replace the concept of scientific confirmation by one of Bayesian probabilities, no paradox ensues. But the original ravens paradox, presented in terms of scientific confirmation, remains unresolved by

this manoeuvre. If the paradox is to be dealt with satisfactorily the descriptive method of conceptual analysis still seems more promising.

Explication and Scientistic Method in Philosophy

Although he endorses the idea of explication (Horwich 1993, p.616), as well as a 'Bayesian' solution to the paradox of the ravens, Horwich is also quite rightly concerned to identify and rebut examples of what he calls 'misplaced scientism' in the philosophy of science (ibid., pp.618ff.). (His tirade against 'hyperformalism' (pp.620-1) is well-taken). Therefore it may be worth bringing out certain ways in which an inappropriately scientistic conception of philosophy lies behind both the method of explication, as Carnap and Maher conceive it, and Horwich's 'solution' to the paradox.

Carnap clearly takes science as his model when thinking about explication. He uses 'the actual procedure of scientists' (Carnap 1950, p.5) as a standard with which to compare the kind of correspondence that should be required between a pre-scientific concept and its explication. The actual procedure in question is to transform an existing but vague concept into a more exact one for scientific purposes (Carnap's main example being the transformation of the concept of a fish into a concept of systematic zoology

(1950, pp.5-6)). But from the fact that this procedure has a legitimate (perhaps even essential) role in science, one cannot straightforwardly conclude that it is appropriate in philosophy.

Consider, now, various aspects of the method of explication. It takes 'ordinary' concepts to be vague or ambiguous, or both, and assumes that vagueness in a concept is a problem, a problem that needs to be dealt with by philosophers (among others). No consideration is given to *why* the original concept might be vague, whether there might not be very good reasons why it might be so, or whether, indeed, a vague concept might not be exactly what is needed.

A good *explicatum*, Carnap and Maher both hold (e.g. Maher 2004, p.69, Maher 2005, p.8), must be similar to the *explicandum*, but it must also be more *precise*, *simple* and *theoretically fruitful*. The imposition of these desiderata shows clearly that an *explicatum*, unlike the original *explicandum* which it is intended to replace, is to be held to certain specifically *scientific* virtues. (Maher requires a good *explicatum* of the ordinary concept of inductive probability, for example, to satisfy the mathematical laws of probability (Maher ibid.)). An *explicatum* is not merely a replacement for its *explicandum*, but a *scientific* replacement for it. The supposedly underlying concept in question is now being held to be evaluable in scientific terms.

Maher, as we saw, conceives the philosopher's task as trying to determine 'whether or not some sentence *S* of ordinary language is true' (Maher 2007,

p.333). We already found it problematic to assume that philosophical problems are raised by sentences (*per se*). But even if one reformulated this conception in terms of statements, rather than sentences, it would also be problematic in modelling philosophy after science, as a quest for further knowledge. Philosophical problems are often of a different form, e.g., 'How is it possible that *p*?' (where there is no doubt about the truth of *p*).

In all these respects one can contrast what is being advertised by the method of explication with the virtue of *clarity*, which might be thought to be a virtue *per se*, or at least an epistemic virtue *per se*, not merely a virtue where scientific purposes are concerned. Philosophers have a professional interest in clarity that is related to their professional goal of *understanding*. But whether philosophy is thereby more closely and constitutively related to science than it is to art, for example, or to law, or to other kinds of human activity, certainly requires arguing. *Some* (more recent, 'naturalistic') conceptions of what it is for a problem to be philosophical will be able to straightforwardly appeal to the prestige that science accrues as our most developed body of *knowledge*. But since Carnap's method and the descriptive method both agree that philosophical problems are *conceptual*, some extra reasoning would be needed to take one from the premise that *understanding* is required for their solution to the scientistic conclusion that their solution should also be under the control of specifically *scientific* virtues.

Two Conceptions of Enlightenment

The method of explication involves working on one's own concepts, transforming one's conceptual scheme by replacing its inexact and unscientific parts with exact and scientific concepts. The clarity thereby attained, by the introduction of brand new concepts, is clarity in the concepts themselves. If philosophical *problems* are conceptual, though, (as both Carnap's method and the descriptive method agree), those problems are left behind, unsolved and untreated by the former method. One has moved away from them towards a conceptual framework which has a reduced tendency to give rise to such problems, presumably. This should certainly count as a benefit, and it is a benefit in which others can share *without* having to re-trace the path of those who introduced the new concepts. That also should count as a merit. On this conception, one might become enlightened, as it were, without ever having been *un*enlightened, or having gone through a transformation in which one subsequently becomes enlightened.

By contrast, the 'descriptive' method involves working on *oneself*, on one's partial and unclear view of our (common) concepts. The clarity thereby attained is clarity and breadth in one's view of what we all already have and use. Philosophical problems being conceptual, they are thereby *dissolved*, they cease to exist *for* the philosopher in question. But this is both *personal*, and an

achievement, one which anyone else would have to re-trace in order to reach the same condition. Enlightenment, conceived of in this way, is not something that someone else could do on one's behalf, and not something that could be effected for a society generally by the offices of a figure who Donald Davidson, in a related connection, memorably dubbed 'the minister of scientific language'. One has to *see* the conceptual problems for oneself, and then see, for oneself, how to dissolve them. This conception is not obviously superior to that inscribed in the method of explication, but it is also not obviously inferior to it. Carnap was quite right to be ecumenical about this (Carnap 1963, pp.939-940).

The Descriptive Method: Philosophy as Conceptual Analysis

To prefer the descriptive method over the method of explication one has to believe that philosophy essentially involves the analysis of concepts, that concepts are present in, and most accessible via, the ways in which people talk about the phenomena in question in non-philosophical contexts, and that philosophers are capable, by analysing such talk, of bringing to light relationships between concepts which function as or reflect norms governing the discourse in question (rather than empirical regularities).

If one or more of these presuppositions is refuted, conceptual analysis makes little sense. But neither Carnap's method of explication nor its followers

have yet shown them to be untenable. That very method of explication, though, is still beset by a serious problem.

Severin Schroeder & John Preston,

Department of Philosophy,

The University of Reading, UK

References

Carnap, R. (1950): Logical Foundations of Probability, (London: Routledge).

Carnap, R.(1963): 'Reply: P.F.Strawson on Linguistic Naturalism', in P.A.Schilpp (ed.), The Philosophy of Rudolph Carnap, (La Salle, IL: Open Court), pp.933-40.

Fitelson, B. 'The Paradox of Confirmation', Philosophy Compass, 1, 2006.

Glock, H-J. (2002): '"Clarity" is not Enough"', in R.Haller & K.Puhl (eds.), Wittgenstein and the Future of Philosophy. A Reassessment After 50 Years, (Vienna: öbv & hpt), pp.81-98.

Goodman, N. (1983): Fact, Fiction and Forecast, 4th edition (Cambridge, MA: Harvard UP).

Hempel, C.G. (1965): Aspects of Scientific Explanation (New York: The Free Press).

Hempel, C.G. (2001): 'Valuation and Objectivity in Science', reprinted in J.H.Fetzer (ed.), The Philosophy of Carl G.Hempel: Studies in Science, Explanation, and Rationality, (Oxford: Oxford University Press), pp.372-95.

Horwich, P. (1993): 'Wittgensteinian Bayesianism', in P.A.French, T.E.Uehling, Jr., & H.K.Wettstein (eds.), Midwest Studies in Philosophy, vol.18 (Notre Dame: University of Notre Dame Press), pp.62-77. (Page references given here are to the reprint in M.Curd & J.A.Cover (eds.), Philosophy of Science: The Central Issues, (New York: W.W.Norton & Co., 1998), pp.607-624).

Howson, C. & Urbach, P. Scientific Reasoning: The Bayesian Approach, 2nd edition (Chicago & LaSalle: Open Court, 1993).

Maher, P. (1999): 'Inductive Logic and the Ravens Paradox', Philosophy of Science, vol.66, pp.50-70.

Maher, P. (2004): 'Probability Captures the Logic of Scientific Confirmation', in C.R.Hithcock (ed.), Contemporary Debates in Philosophy of Science, (Oxford: Blackwell), pp.69-93.

Maher, P. (2005): 'Confirmation Theory', in D.M.Borchert (ed.), The Encyclopedia of Philosophy, 2nd edition (New York: Macmillan), pp.433-42.

Maher, P. (2007): 'Explication Defended', Studia Logica, vol.86, pp.331-341.

Quine, W.v.O. (1951): 'Two Dogmas of Empiricism', as reprinted in his From a Logical Point of View, (Cambridge, MA: Harvard University Press, 1953).

Quine, W.v.O. (1960): Word and Object, (Cambridge, MA: MIT Press).

Schroeder, S.J. (2009): 'Hempel's Paradox, Law-Likeness and Causal Relations', Philosophical Investigations, vol.32, pp.244-263.

Strawson, P.F. (1963): 'Carnap's Views on Constructed Systems versus Natural Languages in Analytic Philosophy', in P.A.Schilpp (ed.), The Philosophy of Rudolph Carnap, (La Salle, IL: Open Court), pp.503-18.

Strawson, P.F. (1992): Analysis and Metaphysics: An Introduction to Philosophy, (Oxford: Oxford University Press).