

Abstract. How can formal methods be applied to philosophical problems that involve informal concepts of ordinary language? Carnap answered this question by describing a methodology that he called “explication.” Strawson objected that explication changes the subject and does not address the original philosophical problem; this paper shows that Carnap’s response to that objection was inadequate and offers a better response. More recent criticisms of explication by Boniolo and Eagle are shown to rest on misunderstandings of the nature of explication. It is concluded that explication is an appropriate methodology for formal philosophy.

Keywords: explication, formal methods, formal philosophy, Carnap.

1. Introduction

This issue of *Studia Logica* is devoted to “formal epistemology,” that is, the application of formal methods to problems and issues in epistemology. However, these problems and issues are typically stated using informal concepts of ordinary language, such as knowledge, rational belief, probability, and confirmation; how can formal methods be applied to such problems and issues? More generally, problems and issues in all areas of philosophy are typically stated using informal concepts of ordinary language; how can formal methods be applied to them?

Carnap answered that question by describing a methodology that he called “explication.” Here is his description of that methodology:

The task of *explication* 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*. The explicandum may belong to everyday language or to a previous stage in the development of scientific language. The explicatum must be given by explicit rules for its use, for example, by a definition which incorporates it into a well-constructed system of scientific either logico-mathematical or empirical concepts. [2, p. 3]

Quine [11, §53] endorsed essentially the same methodology. The term “explication” is sometimes used in a different or looser way, but I will use it to mean the methodology described by Carnap.

Strawson [13] claimed that explication is not relevant to the original philosophical problem, it merely changes the question. In Section 2 I consider this objection, and Carnap’s reply to it; I show that Carnap’s response was inadequate and I give a better response.

Recently, Boniolo [1] has argued that informal analysis is better than explication because explication involves giving definitions that may be wrong. In Section 3 I show that explication need not involve giving definitions and, when it does, the definitions cannot be wrong.

Another recent critic of explication is Eagle [5]. He claimed that the methodology of explication is incomplete because there needs in addition to be a “conceptual clarification” of the explicatum. In Section 4 I show that there is no room for further “conceptual clarification” of an explicatum. Eagle also criticized Carnap for suggesting that explication makes the explicandum of “no further importance,” but in Section 5 I show that this was not Carnap’s view.

I conclude that explication is an appropriate methodology for formal philosophy, including formal epistemology.

2. Strawson on relevance

Strawson, in his discussion of Carnap’s method of explication, said:

It seems *prima facie* evident that to offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant—is 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. [13, p. 505]

Carnap replied that explication can solve philosophical problems arising in ordinary language because it gives us improved new concepts that can serve the same purposes as the ordinary concepts that created the puzzles; the problems are solved by using the new language instead of ordinary language in the problematic contexts. Carnap gave the following analogy:

A natural language is like a crude, primitive pocketknife, very useful for a hundred different purposes. But for certain specific purposes, special tools are more efficient, e.g., chisels, cutting machines, and

finally the microtome. If we find that the pocket knife is too crude for a given purpose and creates defective products, we shall try to discover the cause of the failure, and then either use the knife more skillfully, or replace it for this special purpose by a more suitable tool, or even invent a new one. [Strawson's] thesis is like saying that by using a special tool we evade the problem of the correct use of the cruder tool. But would anyone criticize the bacteriologist for using a microtome, and assert that he is evading the problem of correctly using the pocketknife? [4, pp. 938f.]

Of course, nobody would criticize the bacteriologist, but that is because the bacteriologist's problem was not about the pocketknife. However, the relevant analogy for "one who seeks philosophical illumination of essential concepts of non-scientific discourse" is someone who seeks knowledge of proper use of the pocketknife; Carnap has offered nothing to satisfy such a person.

Carnap seems to have thought that we don't need to take problems about ordinary language very seriously because, when such problems arise, we can develop a new more precise language that serves the same purposes and avoids the problems. But in many cases our purpose is to resolve a problem about a concept of ordinary language, and Carnap has not indicated how a new more precise language can serve that purpose.

For example, suppose our purpose is to determine whether some evidence E makes it more probable that a hypothesis H is true. This problem concerns a concept of probability in ordinary language. We could construct a new more precise language, with a mathematically defined function that is intended to be an explicatum for the relevant ordinary concept of probability. But in order for this new more precise language to serve our purposes, it must enable us to determine whether E makes it more probable (in the ordinary sense) that H is true; Carnap has not explained how the new language could do that.

Furthermore, a good explicatum needs to be sufficiently similar to the explicandum that it can be used for the same purposes, and to determine whether this is the case the explicator must understand the explicandum well. Therefore, an explicator cannot dismiss problems about ordinary language.

So Carnap's response to Strawson was insufficient. I will now propose a better response. Suppose our problem is to determine whether or not some sentence S of ordinary language is true. If we apply the method of explication to this problem, we will construct explicata for the concepts in S , formulate a corresponding sentence S' using these explicata, and deter-

mine whether or not S' is true. This does not *by itself* solve the original problem—that is Strawson's point—but it can greatly assist in solving the problem, 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 that they will correspond well in this case too, and hence to think that 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 . In these ways, explication can provide insights and lines of argument that we may not discover if we reason only in terms of the vague explicanda.

Here is an illustration of these points. Nicod [9, p. 189] claimed that a law of the form “All F are G ” is made more probable by evidence that something is both F and G . Suppose our problem is to determine whether this is correct. Following Nicod, let us use the term “confirms” to mean “raises the probability of;” thus our problem becomes whether a law of the form “All F are G ” is confirmed by evidence that something is both F and G . If we attempt to explicate the concept of confirmation we soon realize that whether or not evidence E confirms hypothesis H depends not only on E and H but also on the background evidence, something that Nicod neglected to specify. If we specify that we are interested in the case where there is no background evidence, then Nicod's claim becomes:

N. A law of the form “All F are G ” is confirmed by evidence that something is both F and G , given no background evidence.

Hempel [7] argued that N is true and Good [6] argued that it is false. In [8] I applied the method of explication to N ; I defined an explicatum C for confirmation (p. 71), formulated an analog of N using C —let us call this N' —and proved that N' is false (p. 77). This does not by itself show that N is false. However, I had argued that C corresponds well with the concept of confirmation in other cases, which is a *prima facie* reason to think that there is correspondence here too, and hence that N is also false. Furthermore, I showed (p. 78) that the proof that N' is false makes intuitive sense when translated back into qualitative explicandum terms. Thus the method of explication provides us with a good argument that the ordinary language hypothesis N is false.

Strawson seems to concede that explication can be useful in something like the ways I have indicated. He wrote:

I should not wish to deny that in the discharge of this task [resolving problems in unconstructed concepts], the construction of a model object of linguistic comparison may sometimes be of great help. [13, p. 513]

But if explication can “be of great help” then it is *not* “like offering a textbook on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart.”

3. Boniolo on definitions

Explication has received little attention in the literature since the 1960s, but two authors have recently published criticisms of it. One of these is Boniolo, who believes that an explication proceeds by giving a definition and argues that this is an inappropriate method for philosophers to use. He says:

If a philosopher defined, he would construe the concept with all of its notes *ab initio*. But, in such a way he would bar his own chances to investigate whether the aspects upon which to dwell have been fixed at the beginning. Moreover, the philosopher who wants to ape the mathematician in using definitions instead of [discursive analyses] runs the risk of believing that his definitions are right when they may in fact be wrong. Conversely, the philosopher who [discursively analyzes] is well aware that his [analyses] may be wrong and incomplete and in such a way, during his analysis, he can suitably modify them. [1, p. 297]

Although Boniolo makes other negative remarks about Carnap and explication, I believe the above passage contains his main substantive objection to (Carnapian) explication.

The first thing to say about this is that an explication need *not* involve giving a definition, at least not if “definition” is understood in the ordinary sense that Carnap uses. I have already quoted Carnap saying that “the explicatum must be given by explicit rules for its use, *for example*, by a definition” (emphasis mine). The alternative to defining the explicatum is to give rules for its use that do not allow it to be eliminated in sentences that contain it; in this case the explicatum is treated as a “theoretical concept” [3].

But let us now consider the case in which an explication does involve a definition. It is important to observe that in this case, what is defined is the explicatum, not the explicandum. So “if the explication consists in giving an explicit definition, then both the definiens and the definiendum

in the definition express the explicatum, while the explicandum does not occur" [2, p. 3]. For example, Carnap's explicandum in [2] was a concept of ordinary language that he called "degree of confirmation" and "probability₁" (p. 25); his explicatum was a function that he called c^* (p. 562). Carnap specified c^* by giving a definition that specified its values for all possible arguments; this is a stipulative definition that specifies what is meant by " c^* ." Carnap tried to *clarify* his explicandum but did not try to *define* it.

So when explication is done by giving a definition, the definition is stipulative—it specifies what the explicatum is—and consequently there is no possibility of the definition being wrong. Therefore, the philosopher who explicates by giving a definition runs no "risk of believing that his definitions are right when they may in fact be wrong," contrary to what Boniolo asserts.

Although explications cannot fail in the way Boniolo supposes, they can fail in other ways, of course. A purported explication sometimes fails because the explicator has failed to distinguish different concepts that might be intended as the explicandum. It may also fail because the explicatum differs from the explicandum in ways that prevent the former being used in place of the latter. However, nothing in the method of explication precludes critical consideration of these issues; in fact, there are many critical discussions of just these issues in Carnap's own work. Hence Boniolo is mistaken in thinking that the method of explication is inimical to the recognition of errors.

4. Eagle on conceptual clarification

The other recent critic of explication is Eagle, who writes:

Carnap [1950] has a long discussion of what he calls "explication" of a pre-theoretical concept in terms of a scientifically precise concept. He gives a number of criteria: that the proposed explicatum (i) be sufficiently similar to the original concept to be recognizably an explication of it; (ii) be more exact or precise, and have clear criteria for application; (iii) play a unified and useful role in the scientific economy (so that it is not just gerrymandered and accidental); and (iv) be enmeshed in conceptual schemes simpler than any other putative explication that also meets criteria (i) – (iii). These are good constraints to keep in mind. However, this model is altogether too compressed; for it presumes that we have an independently good analysis of the scientifically precise concept (in effect, it suggests that scientific theories are not in need of conceptual clarification—that the

“clear conditions of application” are sufficient for conceptual understanding). [5, p. 372]

If the term “scientific theories” is being used in its ordinary sense, then it is undeniable that scientific theories are often in need of conceptual clarification, but that is because these theories often contain concepts that are vague and lack explicit rules governing their use. For example, there are biological theories that contain the vague concept of a species. Such vague concepts are suitable targets for Carnap’s methodology of explication and so it is wrong to say that Carnap’s model “suggests that scientific theories are not in need of conceptual clarification.” Carnap himself, in the passage quoted at the beginning of this essay, said that the explicandum “may belong to . . . a previous stage in the development of scientific language.”

So I take the real issue to be this: Eagle thinks a “scientifically precise concept” which has “clear conditions of application” may nevertheless require an “analysis” or “conceptual clarification” before we can have “conceptual understanding” of it; Carnap believes that if a concept is specified by “explicit rules for its use” then it requires no further clarification. Carnap’s position here accords with the widely shared idea that knowing how to use a term is a sufficient condition for knowing what the term means. How can Eagle deny this?

Just before the passage of Eagle’s quoted above, Eagle gave two examples of what he has in mind. The first concerns probability; Eagle writes:

We wish to find an analysis of probability that makes the scientific use an explication of the pre-scientific use; but this project should not be mistaken for the project of discovering a scientific concept of probability [i.e., an explicatum]. The second task had been performed exactly when we identified scientific probabilities with normed additive measures over the event spaces of scientific theories. But to make this formal structure conceptually adequate we need to give an analysis of both the explicandum and the explicatum. [5, p. 372]

To say that a function p is a “normed additive measure over the event spaces of scientific theories” (i.e., to say it satisfies the mathematical laws of probability) is not enough to give the “explicit rules for its use” that Carnap requires of an explicatum. The laws of probability leave the values of p completely indeterminate except for a few special cases (e.g., the probability of a logical truth is one), whereas “explicit rules for its use” must tell us under what conditions a sentence like “ $p(h, e) = r$ ” is true. Thus Carnap’s

specification of the function c^* , which was his explicatum for probability₁, does not say merely that c^* satisfies the mathematical laws of probability; Carnap fixed c^* uniquely by specifying all its values. And it would make no sense to try to give a “conceptual clarification” of c^* ; the function is just what it is defined to be.

Eagle’s second example concerns Kripke semantics for modal logic. Eagle thinks that this semantics provides an explication that requires “philosophical attention.” But Kripke semantics for modal logic also fails to meet Carnap’s criterion of having “explicit rules for its use;” it does not contain rules that determine which claims about possible worlds are true. And if we had such rules, there would be no room for further “conceptual clarification.”

So Eagle’s belief that explicata require “conceptual clarification” rests on a misunderstanding of the concept of an explicatum. When we understand the concept correctly, we can see that there is no room for further “conceptual clarification” of an explicatum.

Eagle presents himself as being more demanding than Carnap, requiring not just that an explicatum be specified but also that it be given a “conceptual clarification” or “philosophical interpretation.” It is unclear to me what Eagle means by the latter phrases, but from his examples I gather that he does not require the formulation of explicit rules for the use of the concept. Carnap, on the other hand, requires that an explicatum be given by stating such rules. So it is really Carnap, not Eagle, who has the higher standard of what philosophical analysis requires.

5. Eagle on elimination

After the passage just discussed, Eagle makes another criticism of explication. He says of this method:

It also suggests that the explicatum replace or eliminate the explicandum; and that satisfying these constraints is enough to show that the initial concept has no further importance. But clearly the relation between the scientific and pre-scientific concepts is not so one-sided; after all, the folk are the ones who accept the scientific theories, and if the theory disagrees too much with their ordinary usage, it simply won’t get accepted. I take this kind of approach to philosophical analysis to be *pragmatist* in some broad sense; it emphasizes the conceptual needs of the users of scientific theories in understanding the aims and content of those theories. (pp. 372f.)

Eagle's assertion that "the folk are the ones who accept the scientific theories" seems obviously false and the "pragmatist" approach that Eagle endorses is consistent with Carnap's views on explication. But I think that the earlier part of this passage does raise a plausible objection to the method of explication.

I would put the objection this way: Carnap [2, p. 3] talked of the explicatum "replacing" the explicandum and Quine [11, p. 260] said "explication is elimination." This suggests that a successful explication renders the explicandum of "no further importance," as Eagle says. Yet in most cases, explications do not have this effect. For example, the ordinary concept of inductive probability (Carnap's probability₁) continues to be important despite the various explications of it, and it is utterly unrealistic to suppose that any future explicatum will make this ordinary concept disappear. It is neither possible nor desirable to replace statements like "John will probably be late" with some precise quantitative explicatum.

But when Carnap said an explicatum "replaces" the explicandum, he did not mean that it does so in all contexts, only that it does so in particular contexts for which the explicatum is designed. This is shown by the following quotations from Carnap's reply to Strawson ([4], emphases mine):

An explication replaces the imprecise explicandum by a more precise explicatum. Therefore, *whenever greater precision in communication is desired*, it will be advisable to use the explicatum instead of the explicandum. (p. 935)

[A scientific explicatum] will frequently be accepted later into the everyday language, such as "at 4:30 P.M.", "temperature", "speed" as a quantitative term. *In other cases, the explicatum is chiefly used in technical, scientific contexts.* (p. 936)

The constructionist [one who explicates concepts] may ... propose to use, *in certain philosophical contexts (not in contexts of everyday life)*, certain words of everyday language according to certain rules (e.g., to use the word "or" only in the non-exclusive sense). (p. 937)

A natural language is like a crude, primitive pocketknife, very useful for a hundred different purposes. But *for certain specific purposes*, special tools are more efficient. (p. 938)

Strawson already understood this point, writing that:

A pre-scientific concept *C* is clarified in [Carnap's] sense if it is *for certain purposes* replaced (or supplanted or succeeded) by a concept

C' which is unlike C in being both *exact* and *fruitful*. ([13], p. 504, emphases in original)

Since an explicatum is only intended to replace the explicandum in certain contexts and for certain purposes, explication does *not* aim to make the explicandum “of no further importance.”

6. Conclusion

In this paper I have examined four objections to the methodology of explication and argued that none of them is correct. In this way I have defended the view that explication is an appropriate methodology for doing formal philosophy. In addition to defending explication, this discussion has also served to clarify what explication is and how it works.

Acknowledgments. Thanks to Giovanni Boniolo, Antony Eagle, and Branden Fitelson for comments.

References

- [1] BONIOLO, GIOVANNI, ‘Kant’s explication and Carnap’s explication: The *Redde Rationem*’, *International Philosophical Quarterly*, 43 (2003), 3, 289–298.
- [2] CARNAP, RUDOLF, *Logical Foundations of Probability*, University of Chicago Press, Chicago, 1950. Second edition 1962.
- [3] CARNAP, RUDOLF, ‘The methodological character of theoretical concepts’, *Minnesota Studies in the Philosophy of Science*, 1 (1956), 38–76.
- [4] CARNAP, RUDOLF, ‘Replies and systematic expositions’, in Schilpp [12], pp. 859–1013.
- [5] EAGLE, ANTONY, ‘Twenty-one arguments against propensity analyses of probability’, *Erkenntnis*, 60 (2004), 371–416.
- [6] GOOD, I. J., ‘The white shoe *qua* herring is pink’, *British Journal for the Philosophy of Science*, 19 (1968), 156–157.
- [7] HEMPEL, CARL G., ‘Studies in the logic of confirmation’, *Mind*, 54 (1945), 1–26 and 97–121. Reprinted with some changes in *Aspects of Scientific Explanation*, Carl G. Hempel. New York: The Free Press, 1965.
- [8] MAHER, PATRICK, ‘Probability captures the logic of scientific confirmation’, in Christopher R. Hitchcock (ed.), *Contemporary Debates in Philosophy of Science*, Blackwell, Oxford, 2004, pp. 69–93.
- [9] NICOD, JEAN, *Le Problème Logique de l’Induction*, Alcan, Paris, 1923. Page references are to the English translation in [10].
- [10] NICOD, JEAN, *Geometry and Induction*, University of California Press, Berkeley and Los Angeles, 1970. English translation of works originally published in French in 1923 and 1924.
- [11] QUINE, WILLARD VAN ORMAN, *Word and Object*, MIT Press, Cambridge, Mass., 1960.

- [12] SCHILPP, PAUL ARTHUR (ed.), *The Philosophy of Rudolf Carnap*, Open Court, La Salle, IL, 1963.
- [13] STRAWSON, P. F., 'Carnap's views on constructed systems versus natural languages in analytic philosophy', in Schilpp [12], pp. 503–518.

PATRICK MAHER
Department of Philosophy
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
patrick@maher1.net